



EGE ÜNİVERSİTESİ · FEN FAKÜLTESİ · MATEMATİK BÖLÜMÜ

LİSANS TEZİ SAVUNMASI

Makine Öğrenmesi Yardımıyla Kanal Analitiği Verilerinin Etiketlenmesi

Hazırlayan: **Ahmet Baran BOZKURT** · 11210000045

Danışman: **Doç. Dr. Gülnaz BORUZANLI EKİNCİ**

Bornova - İzmir · 2026



SUNUM PLANI

İçerik



01 Giriş, Problem & Amaç

Bağlam, motivasyon ve çalışmanın hedefi



03 Veri Seti & EDA

Yapı, dengesizlik, kanal ve segment analizi



05 Veri Sızıntısı

Tespit, kanıtlama ve senaryo tabanlı arınma



07 Prescriptive Öneri Sistemi

Aksiyon optimizasyonu ve lift analizi



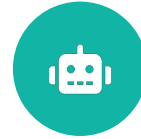
02 Yöntem & Literatür

Pipeline mimarisi ve kuramsal temel



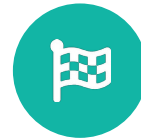
04 Özellik Mühendisliği

17 türetilmiş değişken, beş kategori



06 Modelleme & Performans

LR · Random Forest · XGBoost



08 Sonuç, Kısıtlar & Çıktılar

Değerlendirme ve gelecek çalışmalar



GİRİŞ

Dijital pazarlamada veri-odaklı dönüşüm

Rekabetçi iş dünyasında şirketler, yalnızca ürün kalitesini değil müşteri ilişkilerini de gözetmek zorundadır. İnternetin yaygınlaşmasıyla birlikte çerez ve tıklama verileri, müşteri davranışının çok daha ayrıntılı ve geniş ölçekli analizine olanak tanımıştır.



Müşteri Yaşam Boyu Değeri

CLV, satıcının müşteriyi ilişkisini yönetmede kritik bir kavramdır; kişiye özel kampanyaların temelini oluşturur.



Müşteri Segmentasyonu

Müşteri tabanını ortak özelliklere göre alt gruplara ayırarak her gruba özel strateji geliştirmeyi sağlar.



Kanal Analitiği

Farklı kanallardan toplanan verilerin makine öğrenmesiyle bağlam kazanması ve öngörüye dönüşmesidir.



PROBLEM & AMAÇ

Hangi soruları yanıtlıyoruz?



Dönüşüm tahmini

Hangi müşteri, hangi kanalda dönüşüm sağlar?



Davranış benzerliği

Benzer müşterilere benzer kampanyaların eşleştirilmesi



Aksiyon optimizasyonu

Her müşteri için en uygun kanal-kampanya kararı

ÇALIŞMANIN AMACI

Kanal analitiği verilerini makine öğrenmesiyle anlamlandırıp dönüşüm tahmini yapan uçtan uca bir pipeline ve müşteriye özel bir öneri sistemi tasarlamak.

Prescriptive analitik:

Sadece *ne olacak* değil, *ne yapılmalı* sorusuna yanıt aranmıştır.



YÖNTEM

Uçtan uca makine öğrenmesi pipeline'ı



Çalışmanın özgün katkısı, pipeline'ın dördüncü adımında — veri sızıntısının sistematik biçimde tespit edilip giderilmesinde — yoğunlaşmaktadır.



Kuramsal temel



Gözetimli Öğrenme

- K-En Yakın Komşu (KNN)
- Karar Ağaçları
- Naive Bayes
- Topluluk: Bagging · Boosting · Stacking



Gözetimsiz Öğrenme

- K-Ortalamalar (K-Means)
- Hiyerarşik Kümeleme
- DBSCAN
- Kümeleme & boyut indirgeme



Öneri Sistemleri

- İşbirlikçi filtreleme (kullanıcı/öge)
- İçerik tabanlı filtreleme
- Pearson & kosinüs benzerliği
- Dengesizlik: SMOTE · ADASYN



MATERYAL

Veri seti ve deęişken yapısı

48.000

sentetik müşteri kaydı

20

ham deęişken

%2,04

dönüşüm oranı (980 kayıt)

Veri seti, gerçek dünya pazarlama sistemlerinin davranışsal özelliklerini yansıtacak biçimde tasarlanmıştır. Deęişkenler dört grupta toplanır:



Demografik

Yaş, gelir, sadakat katmanı



Kampanya

Kanal, tür, platform, araç



Dijital Etkileşim

CTR, sayfa/oturum, harcama



Hedef Deęişken

Dönüşüm durumu (ikili)



VERİ YAPISI

Aşırı sınıf dengesizliği

48.000 kayıt içinde yalnızca 980 müşteri dönüşüm gerçekleştirmiştir (%2,04).
Bu yoğun dengesizlik, tüm değerlendirme metodolojisini doğrudan şekillendirmiştir.



Doğruluk (accuracy) yanıltıcıdır

Her şeye “dönüşüm yok” diyen model bile %98 doğruluk alır.



F1 ve ROC-AUC kullanılır

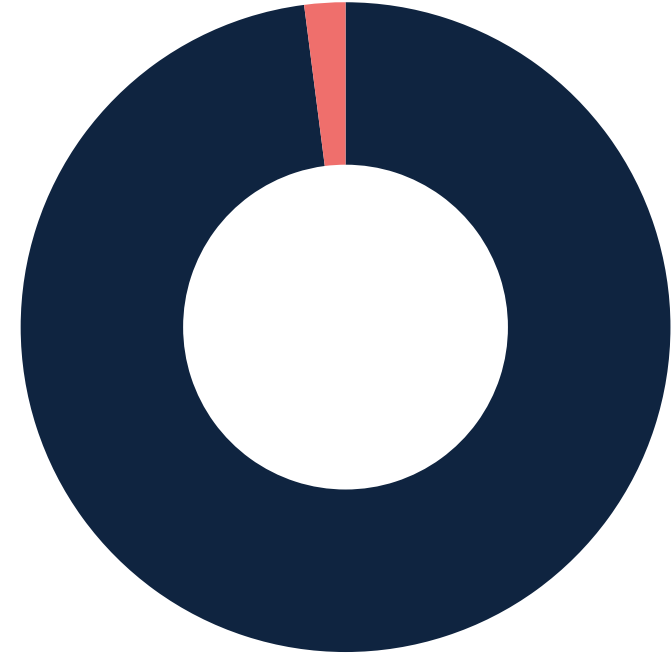
Azınlık sınıfını yakalama başarısı bu metriklerle ölçülür.



Eşik optimizasyonu zorunludur

Varsayılan 0,5 eşiği yerine F1'i maksimize eden eşik seçilir.

SINIF DAĞILIMI



■ Dönüşüm yok · %97,96 ■ Dönüşüm var · %2,04



KEŞİFSEL VERİ ANALIZI

Dört eksen de veri kalitesi

Modelleme öncesi EDA, hem veri kalitesini değerlendirmek hem de özellik mühendisliği kararlarına zemin hazırlamak için dört temel eksen de yürütülmüştür.



Eksik Değer Tespiti



Aykırı Değer Analizi



Çoklu Doğrusallık



Hipotez Testleri

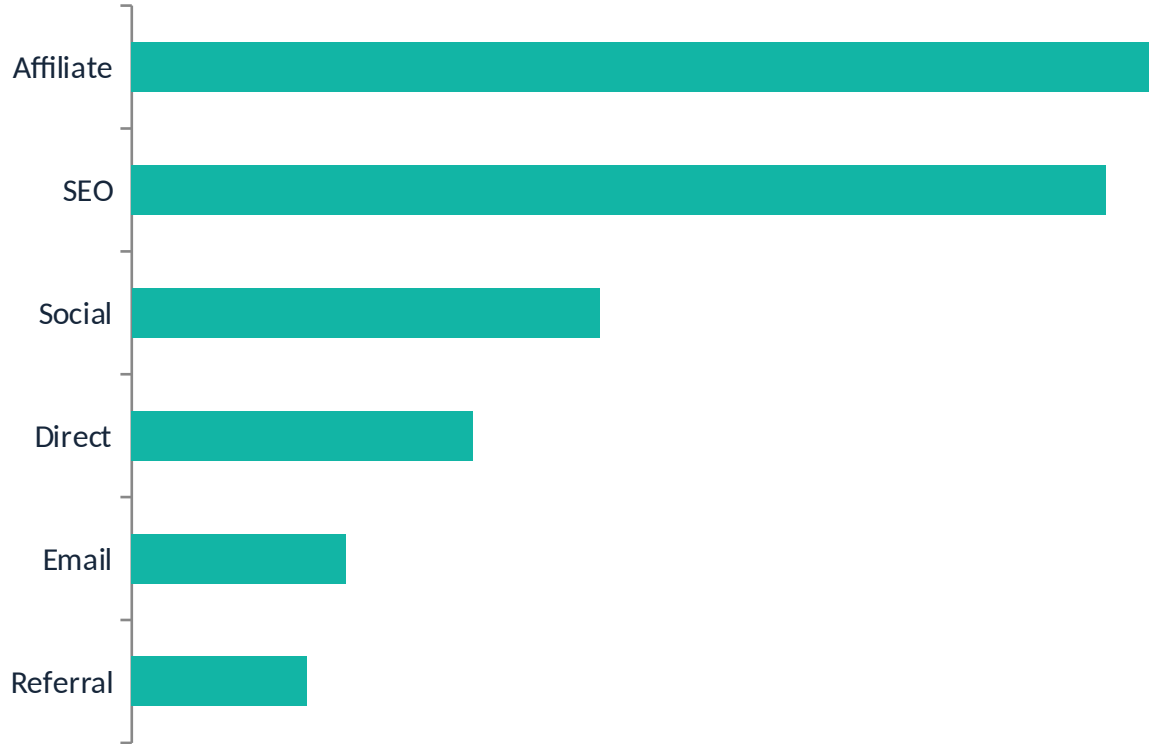


Eksik değerler: ClickThroughRate (2.464) ve PagesPerVisit (2.435) değişkenlerindeki eksikler rastgele kayıp (MCAR) örüntüsüyle uyumlu bulunmuş ve medyan ile doldurulmuştur.



Kanal bazlı maliyet (CPA) analizi

Dönüşüm başına maliyet — kanal karşılaştırması (görelî ölçek)



VERİMLİ KANALLAR

Referral ve Email, düşük maliyetle yüksek dönüşüm sağlar — sınırlı bütçenin öncelikle yönlendirileceği kanallar.

VERİMSİZ KANALLAR

SEO ve Affiliate kanallarında CPA **100.000 \$'ı aşar** — geniş kitleye ulaşır ancak dönüşümde son derece verimsizdir.



EDA · BULGU

Müşteri segmentasyon bulguları



YAŞ GRUBU

MiddleAge (36-50)

En yüksek dönüşüm olasılığı; harcama kapasitesi ve marka sadakati yüksek segment.



SADAKAT KATMANI

Gold > Silver > Bronze

Gold tier müşteriler dönüşüm oranında belirgin biçimde önde; şirkete bağlılık dönüşümü artırır.



KATEGORİK DAĞILIM

**Social Media & Email
ağırlıklı**

AdvertisingPlatform ve AdvertisingTool neredeyse eşit dağılır → ayrıştırıcılık potansiyeli düşük.



ÖZELLİK MÜHENDİSLİĞİ

Ham veriden anlamlı deęişkenlere

Toplam 17 türetilmiş özellik beş kategoride üretildi; her biri modele ayrıştırıcılık katacak biçimde tasarlandı.

	ROI & Maliyet	CPA_Proxy · ROI_Proxy · Spend_Efficiency	3
	Etkileşim Metrikleri	Site_Engagement · Email_Click_Rate · Social_Virality	5
	Müşteri Segmentasyonu	Age_Group · Income_Tier · Customer_Value_Score	4
	Etkileşim Özellikleri	AdSpend×CTR · Income×Loyalty · Age×Purchases	3
	Kanal Performansı	Channel_Performance · Channel_Rank	2



VERİ SIZINTISI

Kavram: hedef sızdırma (target leakage)

Veri sızıntısı, modelin eğitim sırasında gerçek hayatta erişemeyeceği bilgiye ulaşmasıdır. Hedef sızdırma ise bir özelliğin doğrudan hedef değişkenden türetilmiş olmasıdır (Kaufman ve diğerleri, 2012).

BU ÇALIŞMADAKİ ÖRNEK

ConversionRate değişkeni doğrudan hedef değişkenden türetilmişti. Model “öğrenme” yapmadı; yalnızca cevap anahtarını okudu.



Neden tehlikeli?

Sızıntı, eğitim ve test ölçümlerinde kusursuz skorlar üretir; ancak model gerçek dünyada tamamen başarısız olur. Tespit edilmezse projenin tüm sonuçları geçersizdir.



ÇALIŞMANIN EN KRITİK BULGUSU

Tespit: mükemmel skorun tuzağı

“Mükemmel sonuç elde edildiğinde ilk yapılması gereken kutlama değil, şüphedir.”

İlk model **ROC-AUC = 1,0000** ve **F1 = 1,0000** üretti — gerçek bir tahmin yeteneği değil, hedeften türetilmiş özelliklerin yarattığı yapay bir başarıydı.





Senaryo tabanlı arınma süreci

Sızıntı kaynakları aşamalı biçimde ayıklandı. CPA_Proxy çıkarıldığında ROC-AUC'nin aniden düşmesi, sızıntının kaynağını kanıtladı.

Senaryo	Çıkarılan Değişkenler	ROC-AUC	F1	Recall	Durum
A	ConversionRate, CTR_to_Conv.	1,0000	1,0000	1,0000	Sızdırıyor
B	A + ROI_Proxy	1,0000	1,0000	1,0000	Sızdırıyor
C	A + CPA_Proxy	0,7071	0,0755	0,6306	Dürüst
D *	Hepsi birden	0,7072	0,0760	0,6347	Dürüst
E	D + ROI_v2, CPA_v2	0,7063	0,0756	0,6316	Dürüst
F	E + SMOTE	0,6963	0,0740	0,6235	Dürüst

* **Seçilen senaryo (D).** Sızıntısız set içinde en yüksek F1; nihai üretim modeli bu temelden eğitilmiştir. Ek bulgu: SMOTE (F) tüm metrikleri kötüleştirmiştir.



MODELLEME

Aday modeller ve eşik optimizasyonu

Logistic Regression

Yorumlanabilir katsayılar

✓ ÜRETİM MODELİ

Random Forest

Doğrusal olmayan örüntüler

XGBoost

Güçlü gradyan artırma

EŞİK OPTİMİZASYONU

~~0,5~~ → ~0,80

Dengesiz veride varsayılan 0,5 eşiği işe yaramaz; F1 maksimizasyonu ile optimal eşik ~0,80'e taşınmıştır.

MODEL SEÇİM GEREKÇESİ

Lojistik Regresyon, en yüksek ROC-AUC (0,713) ile XGBoost'a yakın performans gösterirken sunduğu **katsayı yorumlanabilirliği** prescriptive sistem için belirleyici olmuştur. Üç modelin de optimal eşiği 0,80 civarında yoğunlaşmıştır.



BULGULAR

Model performans karşılaştırması

Model	ROC-AUC	F1	Precision	Recall	Optimal Eşik
Logistic Regression *	0,713	0,15	0,134	0,179	0,797
XGBoost	0,705	0,14	0,128	0,155	0,803
Random Forest	0,698	0,13	0,121	0,142	0,812



Üretim modeli — Lojistik Regresyon. ROC-AUC 0,713 ve F1 \approx 0,15 değerleri, %2'nin altındaki dönüşüm oranı bağlamında anlamlı bir sınıflandırma kapasitesidir.



Tahminden aksiyona: sistem mimarisi



Sistem, 9.600 müşterilik test setinde uçtan uca çalıştırılarak her müşteri için beklenen ek dönüşüm potansiyeli (lift) ölçülmüştür.



PRESCRIPTIVE · BULGULAR

Lift analizi ve pazarlama önerileri

ORTALAMA BEKLENEN LİFT

+0,0227

Sistem tüm müşterilere aynı değeri sunmaz; en yüksek lift potansiyeline sahip segmentlere odaklanır.



1

Sadakat odağı

Gold tier müşteriler ve sadakat programları önceliklendirilmeli — dönüşüm artırma potansiyeli en yüksek segment.



2

Doğru kampanya türü

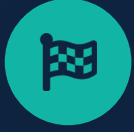
Dönüşüm odaklı (Conversion) kampanyalar diğer türlere göre belirgin biçimde daha etkilidir.



3

Bütçe optimizasyonu

Referral ve Email kanallarına yatırım, dönüşüm başına maliyeti ciddi ölçüde düşürür.



SONUÇ

Genel değerlendirme

48.000 sentetik kayıt üzerinde uçtan uca bir **ML pipeline'ı** ve **prescriptive öneri sistemi** tasarlandı. Sürecin en kritik metodolojik katkısı, **veri sızıntısının sistematik tespiti ve giderilmesi** oldu.

CPA_Proxy'nin tek başına tüm yapay başarıyı sürdürebildiğinin kanıtlanması, hedef sızdırma probleminin somut bir örneğini oluşturur. Performanstaki düşüş bir başarısızlık değil, modelin artık **dürüst bir tahmin** yaptığının göstergesidir.



Şüpheli yaklaşım

"Mükemmel skor" ilk uyarı işaretidir



Dürüst metrik

Accuracy değil F1 / ROC-AUC



Aksiyona dönük

Tahmini öneriye çeviren sistem



DEĞERLENDİRME

Kısıtlar ve gelecek çalışmalar

KISITLAR

- ! Veri sentetiktir; gerçek dünya gürültüsünü tam yansıtmaz.
- ! Aşırı dengesizlik nedeniyle F1 skoru düşük seviyede kalmıştır.
- ! Bulgular tek bir pazarlama bağlamına dayanmaktadır.

GELECEK ÇALIŞMALAR

- Gerçek veri ile validasyon (sanayi danışmanının TEYDEB projesi)
- Derin öğrenme temelli hibrit öneri sistemi
- ADASYN ve gelişmiş dengeleme teknikleri
- Daha zengin etkileşim özellikleri

PROJE ÇIKTILARI

Bilimsel paylaşım & teşekkür

Bu çalışma TÜBİTAK 2209-B programı kapsamında desteklenmiş ve proje çıktıları Ege Üniversitesi 70. Yıl Bilim Şenliği (28.05.2025) kapsamında poster olarak sunulmuştur.



Danışmanım Doç. Dr. Gülnaz BORUZANLI EKİNCİ ve sanayi danışmanı Mine Şener Girgin'e değerli katkıları için teşekkür ederim.



Ek-1 - 70. Yıl Bilim Şenliği Katılım Belgesi

Teşekkürler - Sorularınız?