

ING Hubs Datathon 2025

import-pandas-as-pd / Final Sunumu

Takım Tanıtımı



Özgür Aslan
Machine Learning Engineer
At Cimri



Uğur Ay
Data Analyst
At Medianova

İçerik

01

Problem Tanıtımı

02

Veri Analizi

12

Özellik Seçimi

14

Modelleme

16

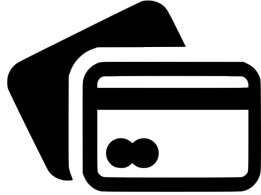
Tavsiyeler

Problem Tanımı ve Senaryo

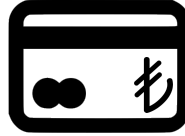
- ING Hubs Datathon 2025 kapsamında, hayali bir **bankanın müşteri verileri incelenmiştir**.
- Veri setinde her müşterinin geçmiş **banka hareketleri** ile **demografik** bilgileri yer almaktadır.
- **Amaç:** Her müşteri için, verilen son tarihten itibaren sonraki 6 ay içinde “churn” olma olasılığını tahmin etmektir.

Veri Analizi

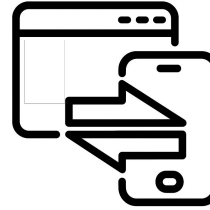
Veri Seti – Müşteri Banka İşlem Verileri



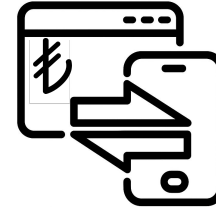
CC İşlem Sayısı



CC İşlem Miktarı



EFT İşlem Sayısı



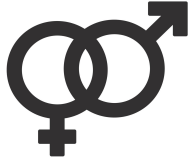
EFT İşlem Miktarı



Aktif Ürün Sayısı

Veri Analizi

Veri Seti – Müşteri Bilgileri



Cinsiyet



Yaş



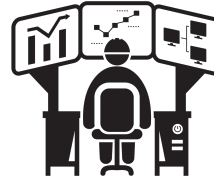
İkamet İli



Din



Çalışma
Şekli



Çalıştığı
Sektör



Müşteri
Olma Süresi

Veri Analizi

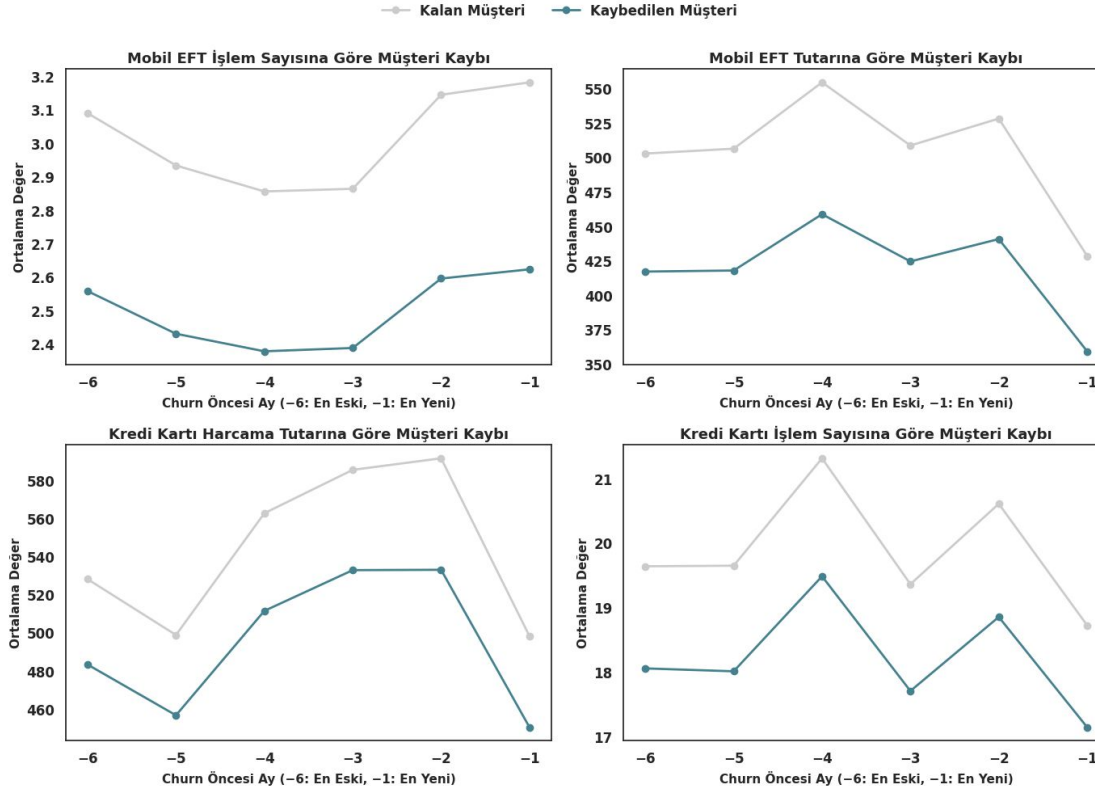
Veriyi Nasıl Analiz Ettik?

Veriyi analiz ederken **RFM** yönteminden esinlenerek, müşterilerin geçmiş banka işlem davranışlarını inceleyip gelecekteki kayıp riskini değerlendirdik.

1. **Recency:** Müşterilerin en son ne zaman işlem yaptıklarını inceledik; özellikle referans tarihinden önceki 6 aylık dönemdeki aktivitelerine odaklandık.
2. **Frequency:** İşlem sıklığını ve zaman içindeki değişimini analiz etmeye başladık, farklı işlem türlerindeki hareketleri de gözlemledik.
3. **Monetary:** İşlem başına düşen harcamaları ve toplam tutarı değerlendirerek, müşteri davranışlarını anlamaya çalıştık.

Veri Analizi

Kredi Kartı ve EFT İşlemleri

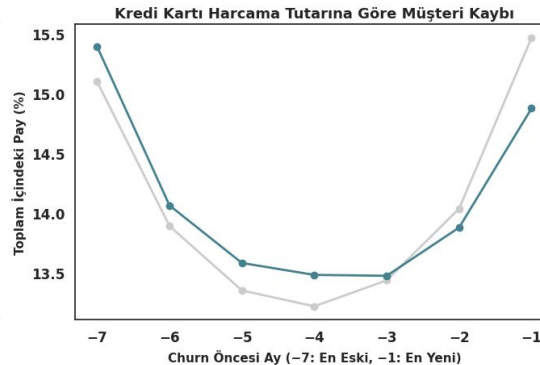
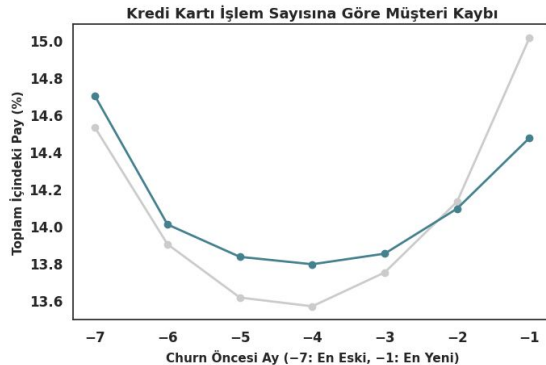
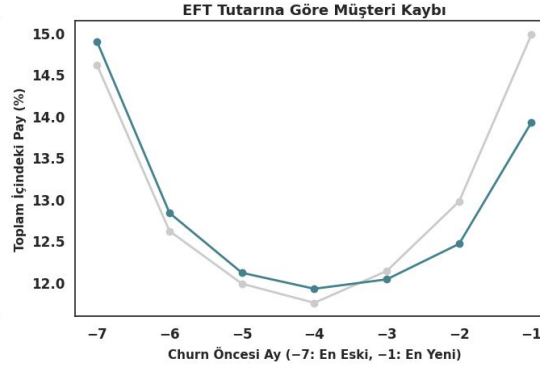
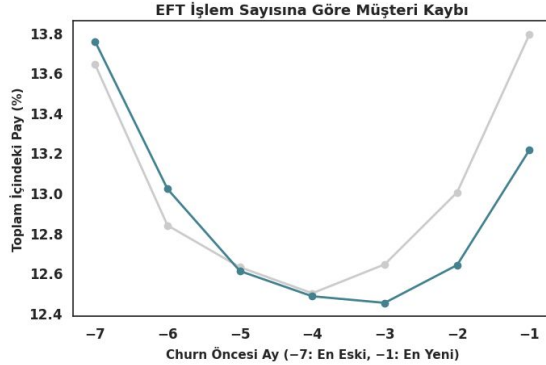


- Değişim oranlarında farklılık var mı?
- Kalan ve kaybedilen müşterilerin değerleri arasında düzenli bir fark var mı?
- Zamanla değişen bir fark var mı?

Veri Analizi

Kredi Kartı ve EFT İşlemleri

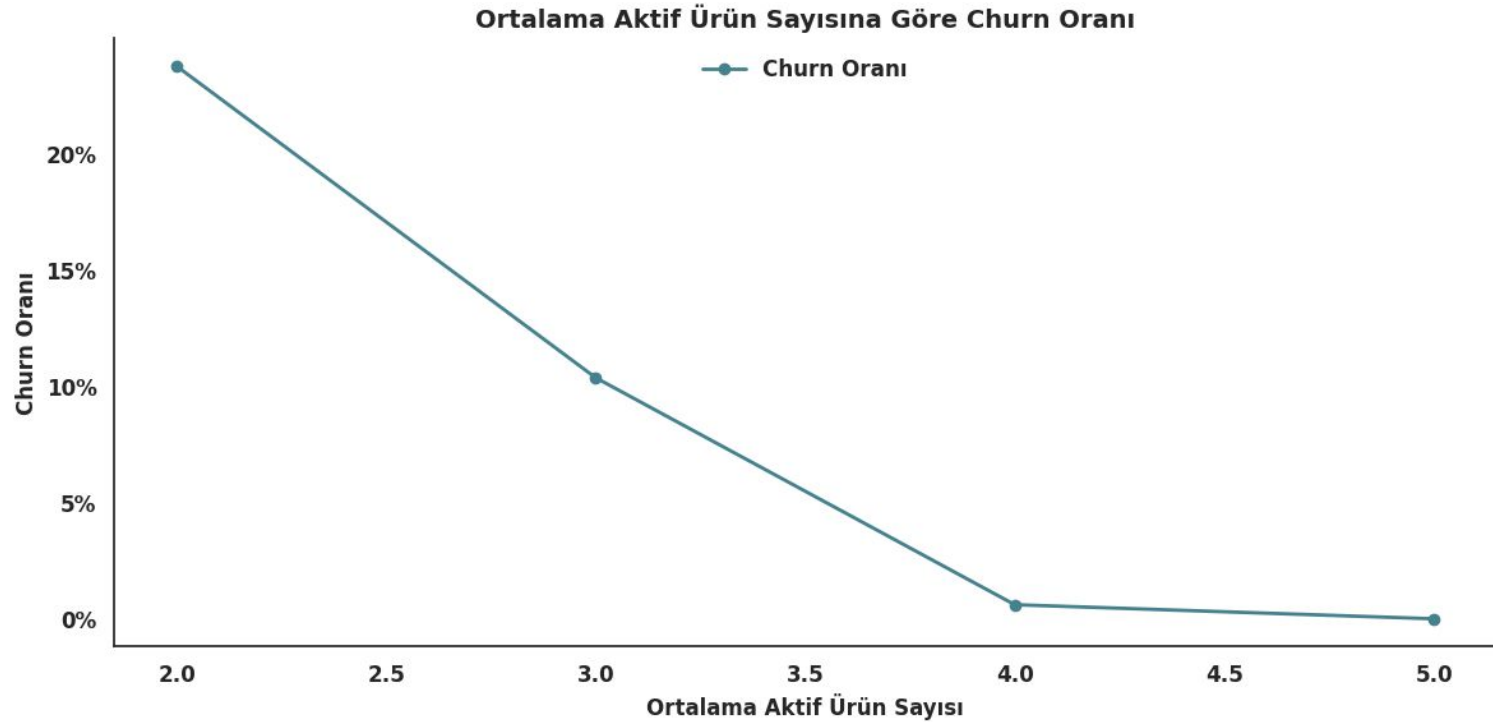
— Kalan Müşteri — Kaybedilen Müşteri



- Toplam bütçe nasıl dağılıyor?
- Kalan ve kaybedilen müşteriler bütçeyi farklı mı paylaşıyor?
- Zamanla bu fark nasıl değişiyor?

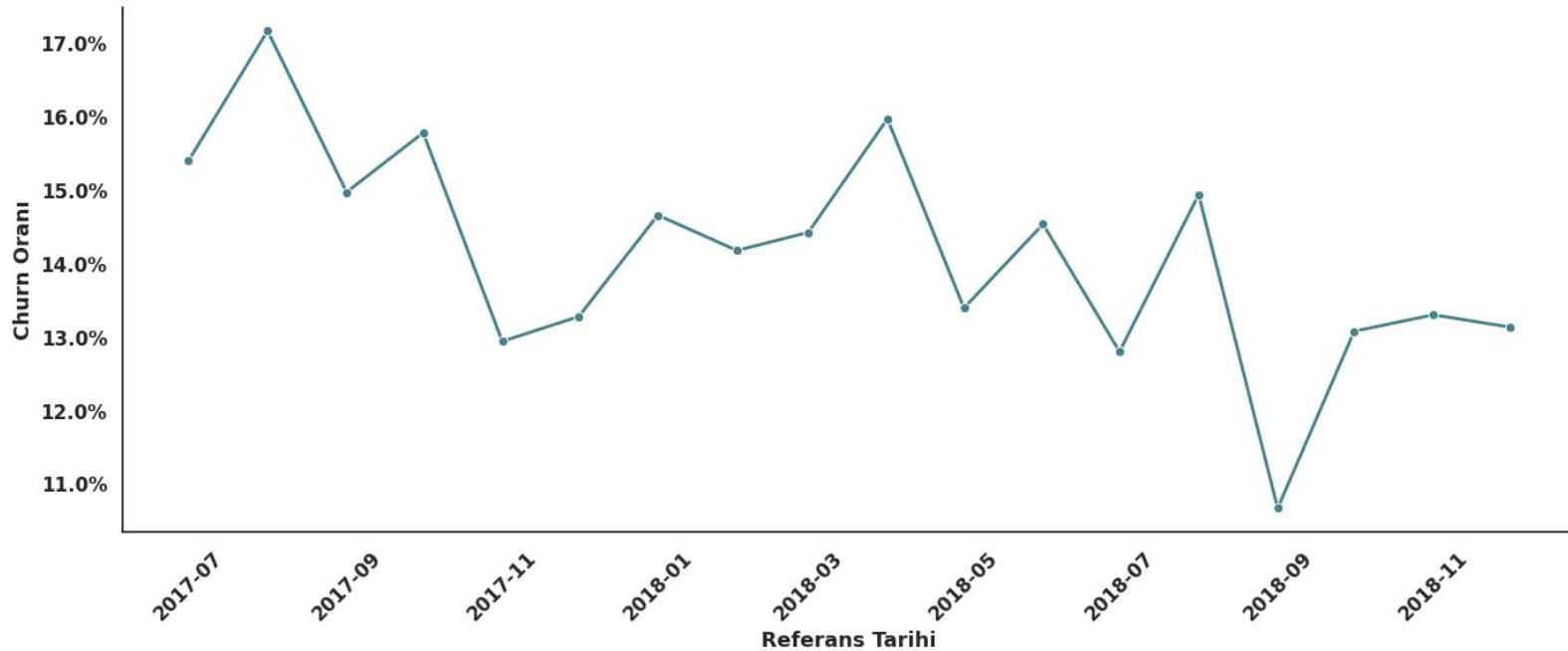
Veri Analizi

Aktif Ürün Sayısının Churn ile İlişkisi



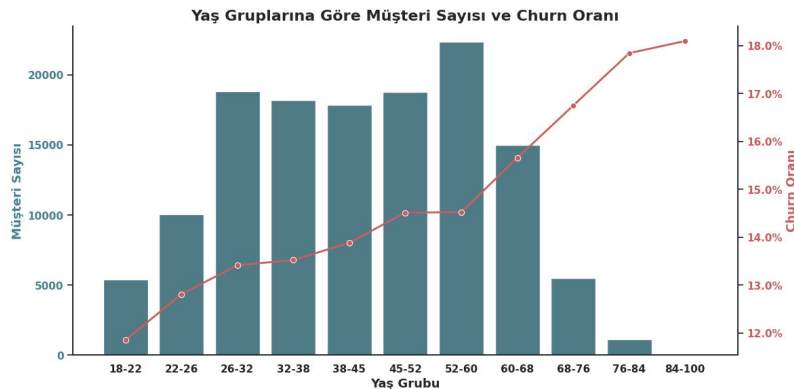
Veri Analizi

Referans Tarihlerine Göre Churn Oranları



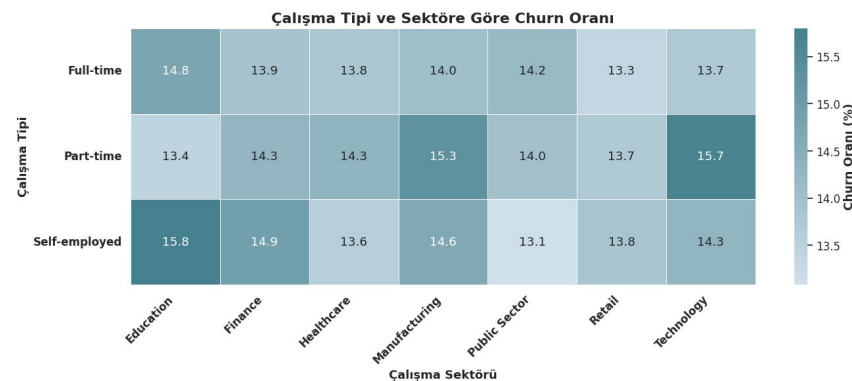
Veri Analizi

Müşteri Bilgileri



Yaş arttıkça churn olasılığı artıyor mu, azalıyor mu?

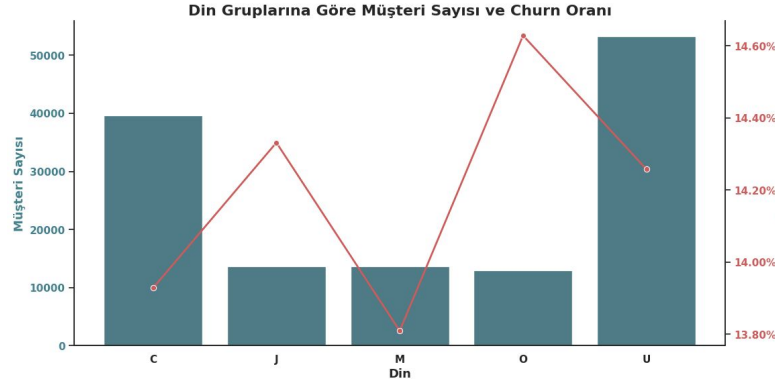
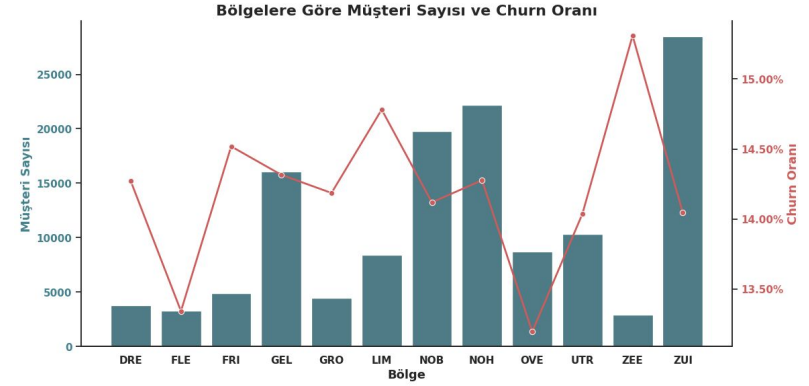
Çalışma tipi ve sektörü churn oranlarını etkiliyor mu?



Veri Analizi

Müşteri Bilgileri

Bölgeler arasında nasıl bir fark var?



Din grupları arasında anlamlı bir fark var mı?

Veri Analizi

Farklı Yaklaşımlar ve Testler (*Son Çözümde Kullanılmadı*)



Harcama miktarları dolara endeksledi.




TÜFE oranı veri setine eklendi.



Churn olan müşterilerin geçmiş verisi ile veri artırımı (data augmentation) yapıldı.

Özellik Seçimi

İstatistiksel Yaklaşım - Welch's T-test

$$t' = \frac{\mu_1 - \mu_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$


Welch's T-test, bir nümerik özelliğin anlamlılığını, churn olan ve olmayan şeklinde iki veri grubuna bölerek ortalama farkını, varyans farklılıklarını da dikkate alarak ölçer; yüksek t skoru, özelliğin churn'u ayırt etme gücünün yüksek olduğunu gösterir.

Öznitelik	t skoru	Ortalama Farkı (%)	Standart Sapma Farkı (%)
avg_products	85,3	10,0%	9,2%
last_3months_products	81,7	11,0%	9,3%
last_6months_products	80,5	10,5%	8,3%
active_product_category_nbr_lag0	78,5	11,3%	8,2%
active_product_category_nbr_lag1	75,5	11,0%	7,7%
active_product_category_nbr_lag2	73,5	10,7%	6,4%
active_product_category_nbr_lag3	70,5	10,3%	5,1%
active_product_category_nbr_lag4	68,3	9,9%	4,1%
active_product_category_nbr_lag5	67,7	9,8%	3,1%
active_product_category_nbr_lag6	67,3	9,7%	3,2%
mobile_eft_all_cnt_lag0	27,8	24,2%	26,4%
last_3months_eft_count	26,5	22,7%	25,8%
mobile_eft_all_cnt_lag1	25,4	22,5%	24,9%
last_6months_eft_count	24,8	21,2%	24,1%
mobile_eft_all_cnt_lag2	23,9	21,3%	23,6%
total_eft_transactions	22,7	18,9%	19,6%
mobile_eft_all_cnt_lag3	22,6	20,6%	22,1%
mobile_eft_all_cnt_lag4	21,1	19,5%	20,7%
mobile_eft_all_cnt_lag5	20,8	19,3%	20,3%
mobile_eft_all_amt_lag0	20,4	25,6%	22,1%

Özellik Seçimi

Sektörel Yaklaşım - Leave One Feature Out (LOFO)

- Herbir özellik, veri setinden tek tek çıkarılarak model tekrar eğitilir ve performans değişimi gözlemlenir
- Eğer model başarısı artıyorsa, o özellik gereksiz olabilir, tabii bir özellik grubuna ait değilse*

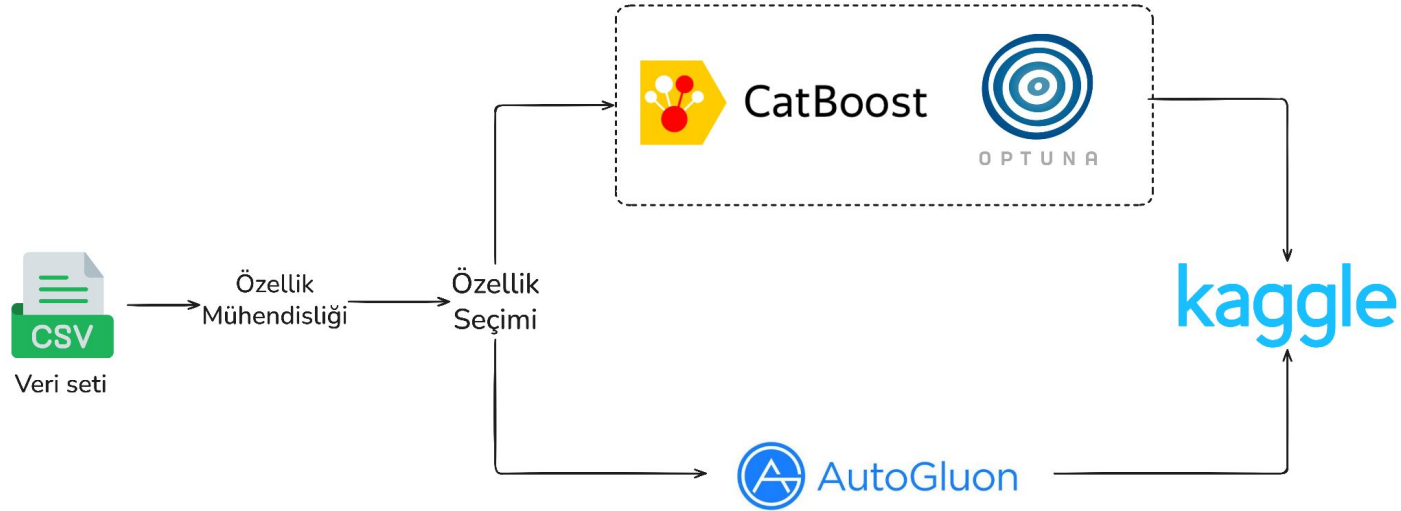


Öznitelik	LOFO Skoru
cc_transaction_all_cnt_lag2	1.1929
total_eft_amount	1.1903
mobile_eft_all_amt_lag0	1.1885
cc_transaction_all_cnt_lag3	1.1880
active_product_category_nbr_lag6	1.1876
Varsayılan	1.1875
mobile_eft_all_amt_lag6	1.1868
last_3months_products	1.1859
work_sector	1.1858
cc_transaction_amt_lag1_perc	1.1856
mobile_eft_amt_lag4_perc	1.1850
last_6months_eft_amount	1.1843
mobile_eft_cnt_lag0_perc	1.1841
mobile_eft_cnt_lag1_perc	1.1840
cc_transaction_amt_lag5_perc	1.1836
cc_transaction_amt_lag3_perc	1.1834
total_months	1.1832
religion	1.1821
mobile_eft_all_cnt_lag2	1.1813
last_6months_cc_count	1.1809
mobile_eft_all_amt_lag3	1.1797

*Lag gibi beraberken anlam oluşturan özelliklerde bir veya birkaç tanesi çıkarıldığında genel anlam bozulabilir ve dolayısıyla modelin gerçek dünyadaki başarısını negatif etkileyebilir

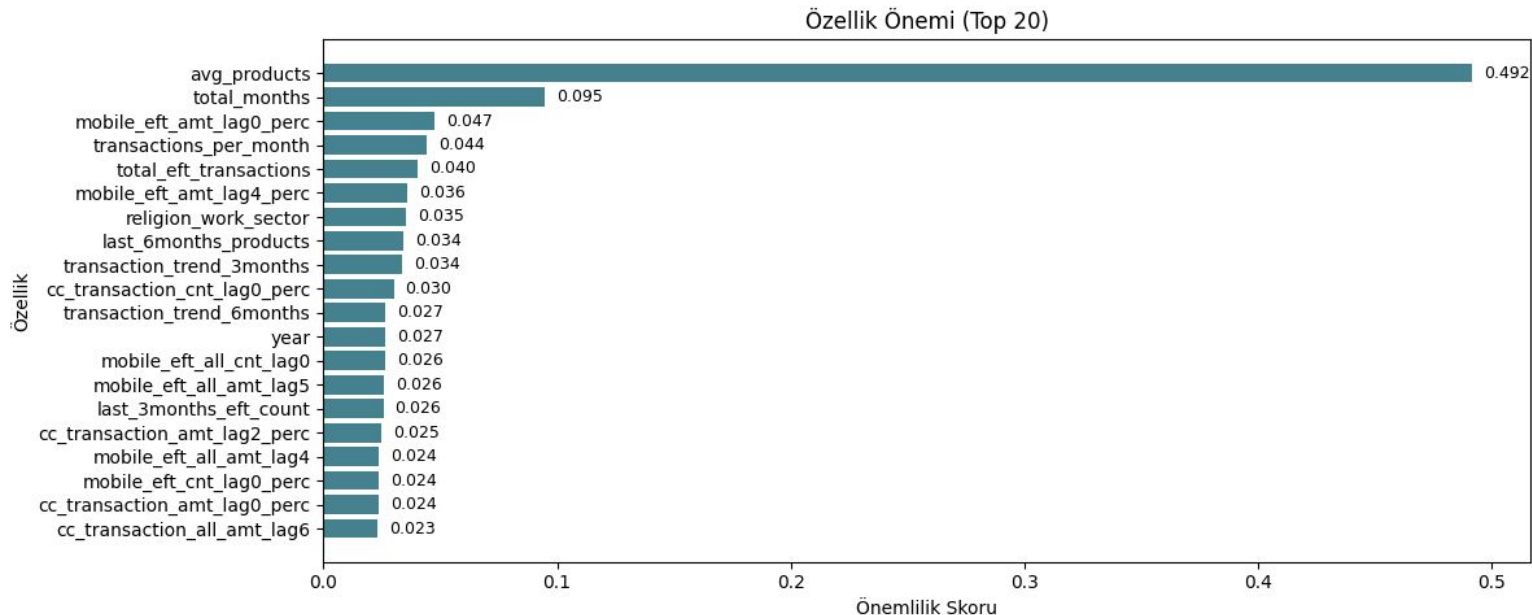
Modelleme

Modelleme Mimarisi



Modelleme

Özellik Önemi



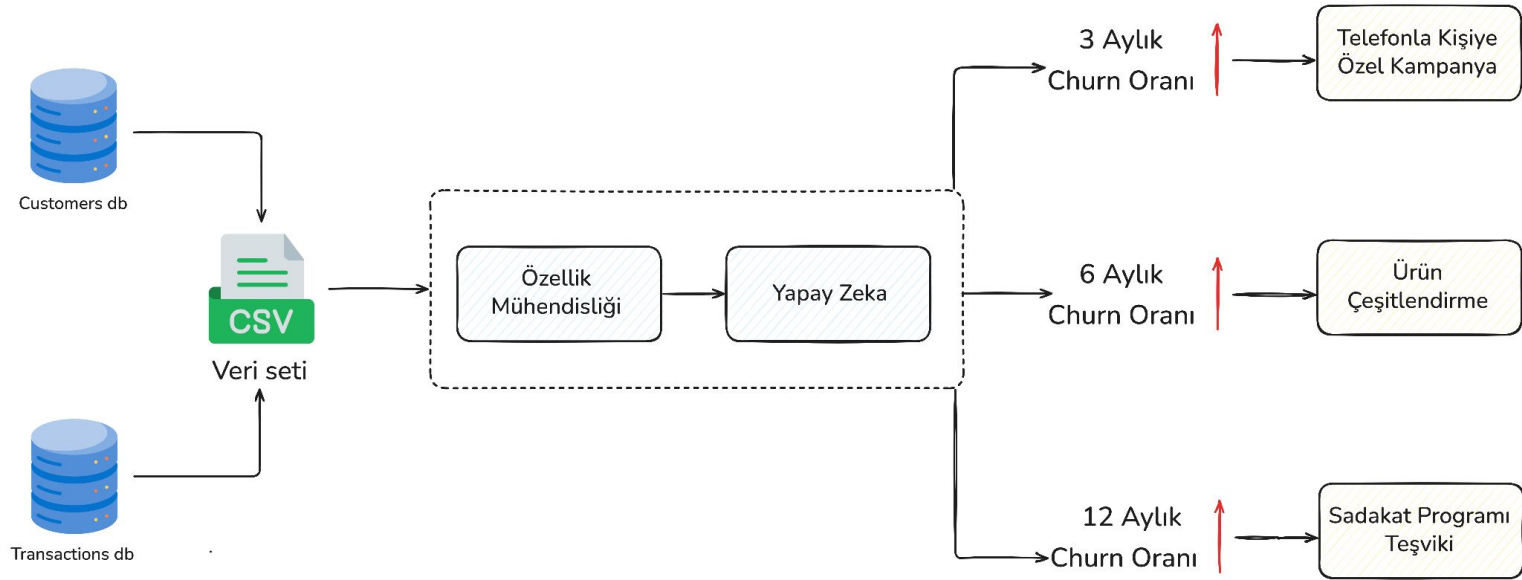
Tavsiyeler

Veri setine eklenebilecek ekstra bilgiler

- **Ekstra Müşteri Bilgileri**
 - ❖ Müşteri segment bilgisi (Bireysel, Ticari, vs.)
 - ❖ Bireysel müşterilerin il/ilçe ve medeni durum bilgisi
 - ❖ Mobil uygulama giriş sıklığı ve ekran süresi
 - ❖ Findeks Puanı
- **Detaylı Ürün Bilgileri**
 - ❖ Müşterinin ürünlerinin kategorileri (Kredi, Kredi kartı, vs.)
 - ❖ Kredi kartı varsa limit doluluk oranı
 - ❖ Mevduat hesabı varsa bakiye tutarı ve trendi

Tavsiyeler

Gerçek Hayat Kullanımı



Teşekkürler