

# Ulusal Faktoring

Kredi Riski Tahmin Modeli

Cek Temerrut Olasılık Modeli (Check Default Prediction)



Optuna ile Optimize Edilmiş 3-Model Ensemble (LightGBM + CatBoost + XGBoost)

Skorlama: AUC | F1 için Esik Deger Optimizasyonu | SHAP Yorumlanabilirlik

Subat 2026

# Problem Tanimi

## Is Hedefi ve Veri Seti Ozeti

### Is Hedefi

- Ceklerin vade tarihinden once temerrut olasiliklarini tahmin etmek
- Proaktif risk yönetimi ve portfoy optimizasyonu saglamak
- Hedef: Zamana dayali test setinde AUC  $\geq 0.75$

### Veri Seti Ozellikleri

**Kayit Sayisi:** ~100.000 cek

**Ozellik Sayisi:** 20+ kolon (sayisal + kategorik)

**Hedef Degisken:** Ikili (temerrut = 1, odenmis = 0)

**Temerrut Orani:** ~%5.87 (yuksek dengesizlik, 16:1)

**Donem:** Ocak 2024 - Ekim 2024

### ONEMLI BULGU

Iki birbirini dislayan segment:

Mikro (kurucu\_yasi dolu)

KOBi (sirket\_yasi dolu)

Eksik degerler MNAR (bilgilendirici):

Odenmemis orani yok = cek gecmisi yok

Kirmizi bayrak yok = temerrut yok

Kredi skoru yok = skorlanmamis

Eksik veri = GURULTU degil, SiNYAL!

# Metodoloji

Uctan Uca Pipeline Tasarimi



## Dogrulama Stratejisi



- StratifiedKFold CV (5-katlı) eğitim verisi üzerinde stabilite analizi
- scale\_pos\_weight ile tutarlı dengesizlik yönetimi (temel + Optuna)
- Sadece eğitim setinden hesaplanan istatistiklerle imputation (veri sizintisi yok)
- PSI (Populasyon Stabilite indeksi) ile ozellik/skor kayması tespiti

# Ozellik Muhendisligi

MNAR-Duyarli + Alan Bilgisi Tabanli Ozellik Uretimi

Eksik Bayraklari (MNAR)	Alan Bilgisi Ozellikleri	Segment'e Ozel	Zamansal & Guncellik
5 ozellik	5 ozellik	5 ozellik	4 ozellik

- |                       |                              |                             |                           |
|-----------------------|------------------------------|-----------------------------|---------------------------|
| - has_credit_score    | - check_maturity_months      | - founder_age_risk (Mikro)  | - term_month, entry_month |
| - has_unpaid_history  | - amount_log, risk_other_log | - young_company (KOBi)      | - redflag_recency         |
| - has_redflag_history | - amount_to_max_balance      | - is_new_customer           | - is_recent_defaulter     |
| - is_micro_segment    | - unpaid_severity            | - is_first_offer            | - branch OHE kukla deg.   |
| - has_late_balance    | - entity_age (birlesik)      | - tenure_amount_interaction |                           |

## Temel Prensip:

- Eksik bayraklar MNAR kaliplarini yakalar. Eksik veri = bilgilendirici sinyal, gurultu degil.
- Tum imputation istatistikleri SADECE egitim verisinden hesaplandi (veri sizintisi yok).

# Modelleme Yaklasimi

Optuna ile AUC Optimizasyonu + Ayri F1 Esik Degeri Ayari

## Skorlama Stratejisi

### Adim 1: AUC ile Model Optimizasyonu

- Optuna 50 deneme x 3 model = 150 toplam
- Her model saf AUC ile optimize edilir
- AUC = siralama kalitesi (threshold bagimsiz)
- En iyi parametreler ile final modeller egitilir

### Adim 2: F1 icin Esik Degeri Ayari

- 500 noktali grid arama (0.01 - 0.99)
- Her model icin ayri threshold belirlenir
- Dogrulama seti üzerinde (test set dokunulmaz)
- 0.5 yerine optimal esik ile F1 dramatik artar

## NEDEN AYIRDIK?

AUC modelin siralama yetenegini olcer (threshold bagimsiz). F1 ise belirli bir esik degerinde ne kadar iyi siniflandirdigini olcer.

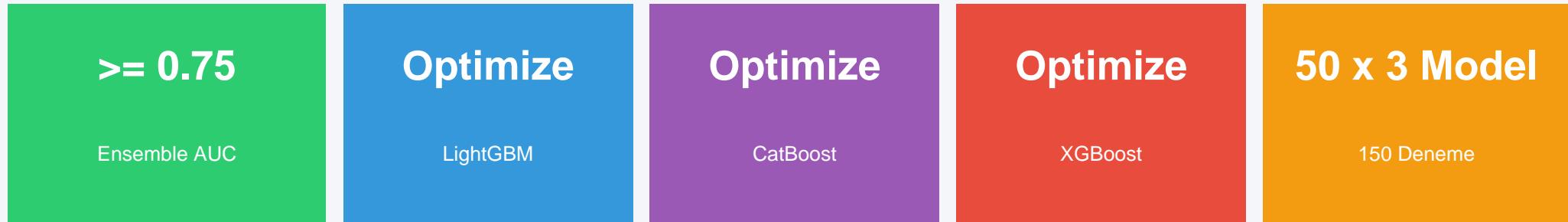
Oncelikle en iyi siralama yapan modeli bul (AUC), sonra en iyi kesme noktasini ayarla (F1). Boylece iki hedef birbirini engellemez.

## Ensemble Yaklasimi

- LightGBM + CatBoost + XGBoost agirlikli ortalama
- Model agirliklari dogrulama seti AUC degerlerine gore belirlenir

# Model Sonucları

Optuna ile Optimize Edilmiş LightGBM + CatBoost + XGBoost Ensemble



## Model Karşılaştırması (AUC)

Model	Dogr. AUC	Test AUC	CV Ort.	CV Std
Temel LGBM	(calistir)	(calistir)	-	-
Optimize LightGBM	(calistir)	(calistir)	(calistir)	(calistir)
Optimize CatBoost	(calistir)	(calistir)	(calistir)	(calistir)
Optimize XGBoost	(calistir)	(calistir)	(calistir)	(calistir)
Ensemble (3 Model)	(calistir)	(calistir)	-	-

\* Degerlere notebook calistirildikten sonra ulasılır. Ensemble = AUC agirlikli LightGBM + CatBoost + XGBoost ortalaması.

# F1 Esik Degeri Optimizasyonu

Dogrulama Setinde Hassas Threshold Belirleme

## Neden Esik Degeri Onemli?

- Varsayılan 0.5 esik değeri, %5.87 temerrut oranlı dengesiz veride çok kotusunuclar verir
- Model genellikle "odenmis" tahmin eder cunku %94 veri oyle -> F1 çok düşük gelir
- Cozum: Dogrulama seti üzerinde F1 skorunu maximize eden esik değerini bulmak
- Bu esik değeri genellikle 0.05-0.15 arasında çıkar (dengesiz veride bu normaldir)

## Optimizasyon Sureci

- Adım 1: 500 esik değeri arasında (0.01 - 0.99) taramalı arama yapılır
- Adım 2: Her model için (LightGBM, CatBoost, XGBoost, Ensemble) ayrı ayrı denenenir
- Adım 3: Her esik için Precision, Recall ve F1 hesaplanır
- Adım 4: F1 skoru en yüksek olan esik değeri seçilir
- Adım 5: Bu esik değeri test setine uygulanarak gerçek performans ölçülür

### Temel Sonuç:

- Threshold 0.5 -> düşük F1 (model neredeyse hiç "temerrut" demiyor)
- Optimize edilmiş threshold -> dramatik F1 artışı (recall yükselsir, precision kontrol altında kalır)

# Stabilite ve Kayma Analizi

StratifiedKFold CV + PSI Ozellik Kaymasi

## Capraz Dogrulama Stabilitesi

- Egitim verisi uzerinde 5 katli StratifiedKFold
- LightGBM, CatBoost ve XGBoost her katta degerlendirilir
- Dusuk varyans (std) = istikrarli tahminler
- Asiri ogrenim kontrolu: Egitim AUC - Test AUC farki < 0.05

## Populasyon Stabilite Indeksi (PSI)

**PSI < 0.10:** Ozellik dagilimi stabil (yesil)

**0.10 <= PSI < 0.25:** Orta derecede kayma, izlenmeli (sari)

**PSI >= 0.25:** Onemli kayma, arastirilmali (kirmizi)

- Skor PSI hesaplanarak modelin dagitim hazirligini dogrulanir
- Ozellik bazli PSI, egitim ve test arasi kayan ozellikleri tespit eder

# Kapsamlı Degerlendirme

Coklu Metrik ve Gorsellestirmeler

## Hesaplanan Metrikler

- ROC-AUC (birincil sıralama metriği)
- PR-AUC / Ortalama Hassasiyet
- KS İstatistiği (ayırma gücü)
- Gini Katsayısı
- Brier Skoru (kalibrasyon kalitesi)
- F1 / Precision / Recall (optimize esik ile)

## İş Senaryoları

**Tutucu:** Yuksek Precision (>50%), az yanlis alarm

**Dengeli:** En iyi F1, doğrulama setinde optimize edilmiş esik

**Agresif:** Yuksek Recall (>70%), cogu temerrutu yakala

## Gorsellestirmeler

- ROC Egrisi (4 model ustuste)
- Precision-Recall Egrisi
- KS Grafigi (TPR vs FPR)
- Kalibrasyon Egrisi
- Skor Dagilimi (odenmis vs temerrut)
- F1/Precision/Recall vs Threshold egrisi
- Karisiklik Matrisi (karsilastirmali)

# Model Yorumlanabilirliği

SHAP Analizi - Yerel ve Genel Açıklamalar

## SHAP Analiz Bileşenleri

**Ozet Cubuk Grafigi:** Ortalama |SHAP| ile sıralanmış genel özellik önemliliği

**Arma Sırtı Grafigi:** Her特色的in etkisinin yönü ve büyüklüğü

**Bağımlılık Grafikleri:** En önemli 4 özellik - doğrusal olmayan ilişkiler

**Kuvvet Grafikleri:** Bireysel açıklamalar - BU cek neden işaretlendi?

### Beklenen Temel Risk Faktörleri (SHAP Tabanlı):

- drawer\_creditbureauscore - Düşük skorlar temerrüdi güçlü tahmin eder
- has\_redflag\_history - Geçmiş temerrüt kaydı en güçlü sinyal
- drawer\_unpaid\_ratio - Yüksek ödenmemiş oran = yüksek risk
- check\_maturity\_months - Uzun vadeli = daha yüksek belirsizlik
- amount\_log - Büyük çek tutarları daha fazla risk taşırlı

# İş Önerileri

Model Sonuçlarından Uygulanabilir Çıkarımlar

## 1. Risk Puanlama Sistemi

Ensemble modeli gerçek zamanlı çek risk puanlayıcı olarak kullanın.

Tutucu esik ile otomatik onay/red, dengeli esik ile manuel inceleme işaretleme.

## 2. Segmente Özel Stratejiler

Mikro segment (kurucu tabanlı): Kurucu yaşı riski ve geciken bakiyelere odaklanın.

KOBİ segmenti (sirket tabanlı): Sirket yası ve kredi buro skoru kapsamına odaklanın.

## 3. Veri Toplama Öncelikleri

Kredi buro skoru kapsamını artırın - skorlanmış cirantaların farklı risk profili var.

Eksik veri bayrakları en önemli tahmin özelliklerinden; veri eksikliği bilgilendiricidir.

## 4. İzleme ve Bakım

PSI ile özellik ve skor kaymasını aylık izleyin. Modeli ucaylıkta yeniden eğitin.

Temel özelliklerde  $\text{PSI} > 0.10$  için uyarı sistemi kurun.

## 5. Portföy Optimizasyonu

Tahmin edilen olasılıklar ile riske dayalı fiyatlandırma yapın.

Portföyu risk katmanlarına (Düşük/Orta/Yüksek/Kritik) ayırarak sermaye tahsisini yapın.

# Tesekkurler

---

Ulusal Faktoring - Kredi Riski Tahmin Modeli

Hedef: Zamana dayali test setinde AUC  $\geq 0.75$

Yaklasim: MNAR-duyarli FE + Optuna ile optimize LightGBM/CatBoost/XGBoost Ensemble

Skorlama: Saf AUC optimizasyonu + Dogrulama setinde hassas F1 esik degeri ayari

Dogrulama: StratifiedKFold CV (3 model) + PSI kayma izleme

Yorumlanabilirlik: Genel ve yerel duzeyde tam SHAP analizi