

Ulusal Faktoring

Kredi Riski Tahmin Modeli

Cek Temerrut Olasililik Modeli (Check Default Prediction)



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

Skorlama: AUC | F1 icin 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 yonetimi ve portfoy optimizasyonu saglamak
- Hedef: Zamana dayali test setinde AUC ≥ 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

Ikisi birbirini dislayan segment:

Mikro (kurucu_yasi dolu)

KOBi (sirket_yasi dolu)

Eksik degerler MNAR (bilgilendirici):

Odenmemis orani yok = cek gecmisi yok

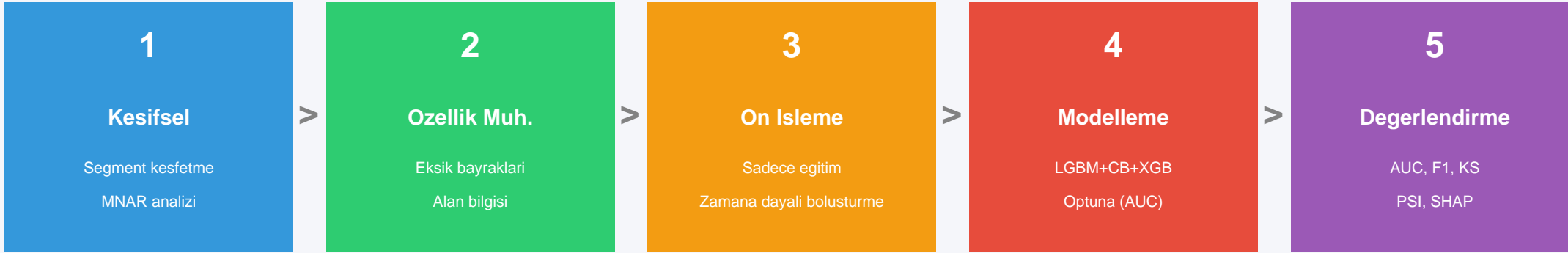
Kirmizi bayrak yok = temerrut yok

Kredi skoru yok = skorlanmamis

Eksik veri = GURULTU degil, SiNYAL!

Metodoloji

Uctan Uca Pipeline Tasarimi



Dogrulama Stratejisi



- StratifiedKFold CV (5-katli) egitim verisi uzerinde stabilite analizi
- scale_pos_weight ile tutarli dengesizlik yonetimi (temel + Optuna)
- Sadece egitim setinden hesaplanan istatistiklerle imputation (veri sizintisi yok)
- PSI (Populasyon Stabilite indeksi) ile ozellik/skor kaymasi tespiti

Ozellik Muhendisligi

MNAR-Duyarli + Alan Bilgisi Tabanlı Özellik Üretimi

Eksik Bayrakları (MNAR)	Alan Bilgisi Özellikleri	Segment'e Özel	Zamansal & Güncellik
5 özellik	5 özellik	5 özellik	4 özellik
<ul style="list-style-type: none">- has_credit_score- has_unpaid_history- has_redflag_history- is_micro_segment- has_late_balance	<ul style="list-style-type: none">- check_maturity_months- amount_log, risk_other_log- amount_to_max_balance- unpaid_severity- entity_age (birlesik)	<ul style="list-style-type: none">- founder_age_risk (Mikro)- young_company (KOBİ)- is_new_customer- is_first_offer- tenure_amount_interaction	<ul style="list-style-type: none">- term_month, entry_month- redflag_recency- is_recent_defaulter- branch OHE kukla deg.

Temel Prensipler:

- Eksik bayraklar MNAR kalıplarını yakalar. Eksik veri = bilgilendirici sinyal, gürültü değil.
- Tüm imputation istatistikleri SADECE eğitim verisinden hesaplandı (veri sızıntısı 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 uzerinde (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.

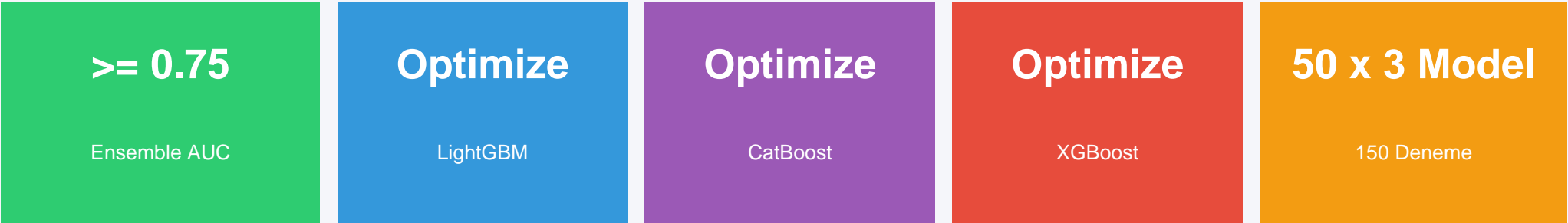
Oncelikle en iyi siralama yapan modeli bul (AUC), sonra en iyi kesme noktasini ayarla (F1). Boylece iki hedef birbirini engellemez.

Ensemble Yaklasimi

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

Model Sonuclari

Optuna ile Optimize Edilmis LightGBM + CatBoost + XGBoost Ensemble



Model Karsilastirmasi (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 calistirildiktan sonra ulasilir. Ensemble = AUC agirlikli LightGBM + CatBoost + XGBoost ortalamasi.

F1 Esik Degeri Optimizasyonu

Dogrulama Setinde Hassas Threshold Belirleme

Neden Esik Degeri Onemli?

- Varsayilan 0.5 esik degeri, %5.87 temerrut oranli dengesiz veride cok kotusonuclar verir
- Model genellikle "odenmis" tahmin eder cunku %94 veri oyle -> F1 cok dusuk gelir
- Cozum: Dogrulama seti uzerinde F1 skorunu maximize eden esik degerini bulmak
- Bu esik degeri genellikle 0.05-0.15 arasinda cikar (dengesiz veride bu normaldir)

Optimizasyon Sureci

- Adim 1: 500 esik degeri arasinda (0.01 - 0.99) taramali arama yapilir
- Adim 2: Her model icin (LightGBM, CatBoost, XGBoost, Ensemble) ayri ayri denenenir
- Adim 3: Her esik icin Precision, Recall ve F1 hesaplanir
- Adim 4: F1 skoru en yuksek olan esik degeri secilir
- Adim 5: Bu esik degeri test setine uygulanarak gercek performans olculur

Temel Sonuc:

- Threshold 0.5 -> dusuk F1 (model neredeyse hic "temerrut" demiyor)
- Optimize edilmiş threshold -> dramatik F1 artisi (recall yukselir, precision kontrol altinda kalir)

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)

$0.10 \leq \text{PSI} < 0.25$: 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

Kapsamli Degerlendirme

Coklu Metrik ve Gorsellestirmeler

Hesaplanan Metrikler

- ROC-AUC (birincil siralama metrigi)
- PR-AUC / Ortalama Hassasiyet
- KS Istatistigi (ayirma gucu)
- Gini Katsayisi
- Brier Skoru (kalibrasyon kalitesi)
- F1 / Precision / Recall (optimize esik ile)

Is Senaryolari

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

Dengeli: En iyi F1, dogrulama setinde optimize edilmis 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 Grafiği: Ortalama |SHAP| ile sıralanmış genel özellik önemliliği

Arma Sursu Grafiği: Her özeliğin etkisinin yonu ve büyüklüğü

Bagimlilik Grafikleri: En önemli 4 özellik - dogrusal olmayan iliskiler

Kuvvet Grafikleri: Bireysel aciklamalar - BU cek neden isaretlendi?

Beklenen Temel Risk Faktorleri (SHAP Tabanlı):

- drawer_creditbureauscore - Dusuk skorlar temerrudi guclu tahmin eder
- has_redflag_history - Gecmis temerrut kaydi en guclu sinyal
- drawer_unpaid_ratio - Yuksek odenmemis oran = yuksek risk
- check_maturity_months - Uzun vade = daha yuksek belirsizlik
- amount_log - Buyuk cek tutarlari daha fazla risk tasir

Is Onerileri

Model Sonuclarindan Uygulanabilir Cikarimlar

1. Risk Puanlama Sistemi

Ensemble modeli gercek zamanli cek risk puanlayici olarak kullanin.

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

2. Segmente Ozel Stratejiler

Mikro segment (kurucu tabanli): Kurucu yas riski ve geciken bakiyelere odaklanin.

KOBi segmenti (sirket tabanli): Sirket yasi ve kredi buro skoru kapsamina odaklanin.

3. Veri Toplama Oncelikleri

Kredi buro skoru kapsamini artirin - skorlanmamis cirantalarin farkli risk profili var.

Eksik veri bayraklari en onemli tahmin ozelliklerinden; veri eksikligi bilgilendiricidir.

4. Izleme ve Bakim

PSI ile ozellik ve skor kaymasini aylık izleyin. Modeli ucaylikta yeniden egitin.

Temel ozelliklerde $PSI > 0.10$ icin uyarı sistemi kurun.

5. Portfoy Optimizasyonu

Tahmin edilen olasiliklar ile riske dayali fiyatlandırma yapin.

Portfoyu risk katmanlarına (Dusuk/Orta/Yuksek/Kritik) ayirarak sermaye tahsisi yapin.

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

deneme.ipynb dosyasindan uretilmistir - Subat 2026