Araştırma Makalesi
Zeki Sistemler Teori ve Uygulamaları Dergisi 9 (2026) 1915517
DOI: 10.38016/jista.1915517

Maliyet Duyarlı Makine Öğrenmesi ile Kredi Kartı Dolandırıcılığı Tespiti: Açık Erişimli ve Ağır Dengesiz İşlem Verisi Üzerinde Karşılaştırmalı Bir Uygulama

Hamid Yeşilyayla[1]

[1,2] Bilgisayar Teknolojileri Bölümü, Pamukkale Üniversitesi, Denizli, Türkiye

hyesilyayla@pau.edu.tr

Öz

Bu çalışma, açık erişimli kredi kartı işlem verisi üzerinde maliyet duyarlı makine öğrenmesi çerçevesinde dolandırıcılık tespitini incelemektedir. Worldline-ULB grubu tarafından paylaşılan karşılaştırma veri kümesindeki 284.807 işlemden 1.081 yinelenen kayıt çıkarılmış, analizler 283.726 gözlem ve 473 dolandırıcılık işlemi üzerinde yapılmıştır. Lojistik Regresyon (Logistic Regression), Rastgele Orman (Random Forest), Aşırı Gradyan Artırma (Extreme Gradient Boosting, XGBoost) ve Hafif Gradyan Artırma Makinesi (Light Gradient Boosting Machine, LightGBM) modelleri ortak eğitim-doğrulama-test bölünmesi altında karşılaştırılmış; V1–V28, Time ve Amount değişkenlerine ek olarak log_amount ve hour_bucket türetilmiştir. Başarım, PR-AUC, ROC-AUC, duyarlılık, F1 skoru ve dengelenmiş doğruluk ile değerlendirilmiş; karar eşiği doğrulama kümesinde beklenen toplam maliyeti en aza indiren değere göre seçilmiştir. Çalışma; rastgele bölme yanında zaman-temelli (kronolojik) bölme, sınıf dengesizliğine yönelik SMOTE, alt örnekleme, SMOTE-Tomek ve focal loss yaklaşımları, Platt ile isotonic olasılık kalibrasyonu ve farklı LightGBM yapılandırmalarının karşılaştırılması üzerinden genişletilmiş bir karşılaştırma sunmaktadır. Doğrulama kümesinde en yüksek PR-AUC 0,8882 ile XGBoost modelinde, en düşük maliyet ise 1.844,88 ile rastgele orman modelinde gözlenmiştir. Test kümesinde XGBoost 0,8207 PR-AUC ve 0,9710 ROC-AUC ile en güçlü ayrıştırma başarımını üretirken, en düşük toplam maliyet 3.896,58 ile rastgele orman tarafından elde edilmiştir; LightGBM Yapılandırma B altında 3.837,01 ile yakın bir maliyet üretmiş, focal loss ise XGBoost taban modeli ile birlikte 3.421,29 maliyetle ek bir kazanım sağlamıştır. Bulgular, ağır dengesiz finansal verilerde model seçiminin tek başına yeterli olmadığını; karar eşiği, sınıf dengesizliği yaklaşımı ve maliyet varsayımlarının model sıralamasını değiştirebildiğini göstermektedir.

Anahtar kelimeler: Kredi Kartı Dolandırıcılığı, Dengesiz Sınıf Yapısı, Maliyet Duyarlı Öğrenme, Karar Eşiği, XGBoost.

Credit Card Fraud Detection with Cost-Sensitive Machine Learning: A Comparative Application on Open-Access and Severely Imbalanced Transaction Data

Abstract

This study examines credit card fraud detection on an open-access transaction dataset within a cost-sensitive machine learning framework. After removing 1,081 duplicate records from the 284,807 transactions in the benchmark dataset shared by the Worldline-ULB group, the analyses were conducted on 283,726 observations and 473 fraudulent transactions. Logistic Regression, Random Forest, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) models were compared under a common training-validation-test split; in addition to the V1–V28, Time, and Amount variables, log_amount and hour_bucket were derived. Performance was evaluated using PR-AUC, ROC-AUC, recall, F1 score, and balanced accuracy, while the decision threshold was selected according to the value that minimizes the expected total cost on the validation set. The study presents an extended comparison covering, alongside random splitting, time-based (chronological) splitting, the SMOTE, undersampling, SMOTE-Tomek, and focal loss approaches for class imbalance, Platt and isotonic probability calibration, and the comparison of different LightGBM configurations. On the validation set, the highest PR-AUC was observed in the XGBoost model with 0.8882, whereas the lowest cost was observed in the Random Forest model with 1,844.88. On the test set, XGBoost produced the strongest discriminative performance

* Sorumlu yazar.
E-posta adresi: hyesilyayla@pau.edu.tr

Alındı : 24 Mart 2026
Revizyon : 27 Nisan 2026
Kabul : 15 Mayıs 2026
Yayın Tarihi : 15 Eylül 2026

with a PR-AUC of 0.8207 and a ROC-AUC of 0.9710, while the lowest total cost was obtained by Random Forest with 3,896.58; LightGBM under Configuration B produced a comparable cost of 3,837.01, and focal loss, together with the XGBoost base model, provided an additional gain with a cost of 3,421.29. The findings show that, in severely imbalanced financial data, model selection alone is not sufficient; the decision threshold, the class-imbalance approach, and the cost assumptions can change the model ranking.

Keywords: Credit Card Fraud Detection, Imbalanced Classification, Cost-Sensitive Learning, Threshold Optimization, XGBoost.

1. Giriş (Introduction)

Dijital ödeme altyapılarındaki genişleme, işlem hacmi ve işlem hızı ile dolandırıcılık tespitinin karmaşıklığını artırmıştır. Kredi kartı dolandırıcılığı verilerindeki temel güçlük, pozitif sınıfın toplam işlemler içindeki payının son derece düşük olması ve yanlış sınıflandırma maliyetlerinin simetrik olmamasıdır. Bir dolandırıcılık işleminin kaçırılması doğrudan finansal zarara yol açarken, meşru bir işlemin yanlış alarm olarak işaretlenmesi çoğu durumda inceleme maliyetine sebep olmaktadır. Bu asimetri, doğruluk ya da alıcı işletim karakteristiği eğrisi altındaki alan (receiver operating characteristic area under the curve, ROC-AUC) bakımından güçlü görünen bir modelin operasyonel kayıp açısından aynı ölçüde başarılı olmayabileceği anlamına gelmektedir.

Son on yıldaki literatür üç ana doğrultuda yoğunlaşmıştır. İlk doğrultu, klasik ve topluluk temelli sınıflayıcıların dengesiz veri üzerinde karşılaştırılmasıdır (Randhawa vd., 2018; Alarfaj vd., 2022; Afriyie vd., 2023). İkinci doğrultu, öznitelik seçimi, veri dengeleme, temsili öğrenme ve otomatik öznitelik üretimi gibi yaklaşımlarla ayrıştırma başarımını artırmaya yönelmektedir (Bahnsen vd., 2016; Ileberi vd., 2022; Ghosh Dastidar vd., 2022; Fanai ve Abbasimehr, 2023; Chhabra vd., 2024). Üçüncü doğrultu ise sekans, grafik ve dağıtık öğrenme çerçeveleriyle işlem ilişkilerini ya da kurumlar arası veri ayrışmasını modellemektedir (Cherif vd., 2023; Motie ve Raahemi, 2024; Tang ve Liang, 2024; Xie vd., 2024).

Buna karşın, açık karşılaştırma veri kümesi üzerinde aynı deney düzeni içinde duyarlılık-kesinlik eğrisi altındaki alanı (precision-recall area under the curve, PR-AUC), maliyet duyarlı eşik seçimini, yanlış negatif ve yanlış pozitif maliyet bileşenlerini ve açıklanabilirlik çözümlemesini birlikte ele alan çalışmalar sınırlıdır. Dengesiz veri örnekleme ile yapay biçimde dengelendiğinde elde edilen çok yüksek başarı oranları, gerçek sınıf dağılımı altında her zaman ulaşılamamaktadır (Cherif vd., 2023; Baisholan vd., 2025). Bu çalışma, söz konusu boşluğu aynı veri kümesi üzerinde ayrıştırma başarımı ile beklenen maliyetin birlikte değerlendirilmesi üzerinden ele almaktadır.

Bu çerçevede çalışma dört soruya odaklanmaktadır. İlk olarak, ağır dengesiz sınıf yapısına sahip açık kredi kartı işlem verisinde hangi model ayrıştırma başarımı bakımından daha üstündür? İkinci olarak, karar eşiği maliyet duyarlı biçimde seçildiğinde model sıralaması değişmekte midir? Üçüncü olarak, log_amount ve hour_bucket değişkenleri modele anlamlı bir katkı sağlamakta mıdır? Son olarak, PR-AUC, duyarlılık (recall) ve F1 skoru gibi ölçütler, ROC-AUC ve dengelenmiş doğruluğa (balanced accuracy) kıyasla bu problem bağlamında daha ayırt edici bir değerlendirme zemini sunmakta mıdır?

Bu dört soruya karşılık gelen hipotezler şöyle tanımlanmıştır: H1, XGBoost modelinin PR-AUC ve F1 bakımından en güçlü sonucu üreteceğini; H2, maliyet duyarlı eşik seçiminin duyarlılık ve toplam beklenen maliyet üzerinde belirgin bir etki yaratacağını; H3, türetilmiş değişkenlerin performansı iyileştireceğini; H4 ise ağır dengesiz veri bağlamında PR-AUC, duyarlılık ve F1 ölçütlerinin ROC-AUC ile dengelenmiş doğruluktan daha açıklayıcı bir resim sunacağını öne sürmektedir.

2. Literatür (Literature)

2.1. Yöntem bazlı literatür değerlendirmesi (Method-based literature review)

Yöntem eksenindeki çalışmalar incelendiğinde, kredi kartı dolandırıcılığı literatürünün öznitelik mühendisliği, dengesiz sınıf düzenlemesi, temsili öğrenme, açıklanabilirlik ve kurumlar arası işbirlikli modelleme başlıklarında yoğunlaştığı görülmektedir. Bahnsen vd. (2016), işlem geçmişinden türetilen periyodik ve toplulaştırılmış değişkenlerin maliyet duyarlı değerlendirme altında anlamlı katkı ürettiğini göstermiştir. Dal Pozzolo vd. (2018) ile Jurgovsky vd. (2018), kronolojik işlem akışı ile sekans bilgisinin göz ardı edilmemesi gerektiğini ortaya koyarak gerçekçi değerlendirme düzeninin önemini vurgulamıştır. Randhawa vd. (2018) ise AdaBoost ve çoğunluk oylamasını birleştiren hibrit topluluk modelinin (ensemble) hem açık hem de kurumsal veri üzerinde dayanıklılık sağlayabildiği sonucuna ulaşmıştır.

Daha güncel çalışmalar dengesizliği, öznitelik seçimini ve derin temsilleri daha doğrudan hedeflemektedir. Ileberi vd. (2022), genetik algoritma ile öznitelik seçiminin sınıflandırma kalitesini artırdığını; Alarfaj vd. (2022) ile Fanai ve Abbasimehr (2023), XGBoost, derin otoenkoder ve derin sınıflandırıcıların açık veri üzerinde yüksek ayrıştırma gücü üretebildiğini; Ghosh Dastidar vd. (2022) ise öğrenilebilir birleştirme yöntemlerinin, insanların elle yaptığı veri toplulaştırma işlemlerine bir alternatif oluşturduğunu ortaya koymuştur. Chhabra vd. (2024), oylama temelli topluluk modelinin tekil modellere göre F1 ve AUC artışı sağladığını bildirirken, Abdul Salam vd. (2024) dağıtık öğrenme düzeninde dengeleme tekniklerinin katkısını tartışmaktadır.

Diğer bir çizgi, işlem ilişkileri ile model açıklanabilirliğini birlikte ele almaktadır. Tang ve Liang

(2024), federated graph learning yaklaşımıyla kurumsal veri siloları arasında ilişkisel yapıyı modellemiş; Xie vd. (2024) davranışsal temsil öğrenimiyle işlem örüntülerini zaman boyutunda çözümlemiştir. Ahmed vd. (2025), Albalawi ve Dardouri (2025), Baisholan vd. (2025), Btoush vd. (2025) ve Tayebi ile El Kafhali (2025) ise örnekleme, eşik ayarı, Bayesyen optimizasyon ve SHAP destekli yorumlama gibi yaklaşımları farklı tasarımlar içinde kullanmıştır. Bu birikim, problem alanında model tipi kadar örnekleme tercihi, karar eşiği ve öznitelik yapısının da sonuçları belirlediğini göstermektedir.

Çalışmada kullanılan modellerden güncel literatüre bakıldığında, Tekin ve Yaman (2023), akıllı ev sistemlerine yönelik saldırı tespitinde XGBoost tabanlı bir sınıflandırma düzeni kurmuş; Yıltırak ve Öztürk (2025) ise Nesnelerin İnterneti saldırı tespitinde Rastgele Orman, XGBoost, LightGBM ve Lojistik Regresyonu birlikte karşılaştırmıştır. Her iki çalışma da ağaç tabanlı yöntemlerin tespit problemlerinde güçlü sonuçlar üretebildiğini göstermektedir.

2.2. Veri bazlı literatür değerlendirmesi (Data-driven literature review)

Veri eksenindeki çalışmalar iki ana kümeye ayrılmaktadır. İlk kümede, açık erişimli European cardholders açık karşılaştırma veri kümesini kullanan çalışmalar yer almaktadır. Bu veri kümesi Eylül 2013'te iki günlük işlem akışını, 284.807 gözlemi ve 492 dolandırıcılık kaydını içermekte; V1–V28, Time ve Amount değişkenlerinden oluşmaktadır. Ileberi vd. (2022), Alarfaj vd. (2022), Sinap (2024), Chhabra vd. (2024), Abdul Salam vd. (2024), Ahmed vd. (2025), Baisholan vd. (2025), Albalawi ve Dardouri (2025), Btoush vd. (2025) ile Tayebi ve El Kafhali (2025) aynı veri ailesini kullanarak sonuçların karşılaştırılabilirliğini artırmaktadır. Açık karşılaştırma veri kümesinin başlıca üstünlüğü tekrar üretilebilirliktir; buna karşılık iki günlük ve anonimleştirilmiş yapısı, müşteri geçmişi ve ilişki ağları gibi operasyonel olarak değerli bileşenleri dışarıda bırakmaktadır.

Veri yapısı bakımından Güney ve Selvi (2023) yaptıkları çalışmada 2006–2022 döneminde 8.224 mesken abonesine ait tüketim verisini önce anomali çözümlemesiyle, ardından sınıflandırma modelleriyle incelemiştir. Bu iki aşamalı kurgu, seyrek ve olağandışı örüntülerin işlem verisi içinde ayrıştırılması bakımından mevcut çalışmayla yöntemsel yakınlık taşımaktadır.

İkinci kümede, kurum içi ya da simüle edilmiş veri kümelerine dayanan çalışmalar bulunmaktadır. Bahnsen vd. (2016), Dal Pozzolo vd. (2018), Jurgovsky vd. (2018), Randhawa vd. (2018), Ghosh Dastidar vd. (2022), Afriyie vd. (2023) ve Altan ile Zafer (2024) farklı ölçekte gerçek ya da kuruma özgü işlem akışlarını kullanmıştır. Bu çalışmalar uygulama bağlamına daha yakındır; işlem saati, kart sahibi özellikleri, müşteri davranış geçmişi ve kurumlar arası dağıtık yapı gibi değişkenleri içerebildikleri için operasyonel karar süreçlerine daha fazla yaklaşmaktadır. Buna karşılık, veri gizliliği nedeniyle dış doğrulama sınırlı kalmakta ve doğrudan karşılaştırma güçleşmektedir. Sistematik derlemeler, bu iki veri yapısı arasındaki açıklığın halen kapanmadığını işaret etmektedir (Cherif vd., 2023; Motie ve Raahemi, 2024). Bu çalışma, iki yönelimin kesişim noktasında durarak açık karşılaştırma veri kümesi üzerinde operasyonel maliyet bakışını öne çıkarmaktadır.

3. Yöntem (Method)

3.1. Araştırma modeli, evren ve veri kaynağı (Research model, population, and data source)

Çalışma nicel, karşılaştırmalı ve uygulamalı bir araştırma tasarımına dayanmaktadır. Analiz birimi işlem düzeyindeki kredi kartı kayıtlarıdır. Çalışmanın evreni, açık erişimli kredi kartı işlem veri kümesinde yer alan tüm gözlemlerden oluşmaktadır. Veri kaynağı, OpenML ve Kaggle üzerinden erişilebilen ve Worldline-ULB grubu tarafından paylaşılan karşılaştırma veri kümesidir. Özgün veri kümesi Eylül 2013'teki iki günlük işlem akışını yansıtmaktadır ve 284.807 işlem ile 492 dolandırıcılık kaydı içermektedir. Veri kümesinden 1.081 yinelenen satır çıkarılmış; nihai analiz 283.726 gözlem ve 473 dolandırıcılık işlemi üzerinden yürütülmüştür. Dolandırıcılık oranı yüzde 0,167 düzeyindedir. Bu yapı, çalışmanın temel problem çerçevesini belirleyen ağır dengesiz sınıf düzenini ortaya koymaktadır. Veri kümesindeki V1–V28 değişkenleri, gizlilik nedeniyle temel bileşen analizi (principal component analysis, PCA) ile dönüştürülmüş bileşenlerdir; Time ve Amount ise dönüştürülmemiş özgün değişkenler olarak yer almaktadır.

3.2. Veri toplama süreci ve ön işleme (Data collection and preprocessing)

Veri toplama süreci, kamuya açık veri kümesinin indirilmesi, yinelenen kayıtların temizlenmesi, değişkenlerin dağılım özelliklerinin incelenmesi ve modellemeye hazır hale getirilmesi adımlarından oluşmaktadır. Ön incelemede eksik gözlem bulunmadığı görülmüştür. Amount değişkeni için aykırı değerler çeyrekler arası aralık (interquartile range, IQR) yaklaşımıyla taranmış; ancak işlem tutarının doğal kuyruk yapısı ve temel bileşen analizi (principal component analysis, PCA) ile anonimleştirilmiş değişkenler olmasından dolayı gözlem silme yoluna gidilmemiştir. Sayısal alanlar RobustScaler ile ölçeklendirilmiştir. Özellik türetme aşamasında Amount değişkeninden log_amount, Time değişkeninden ise hour_bucket türetilmiştir. log_amount, Amount üzerinde log1p dönüşümü uygulanarak elde edilmiş; bu dönüşüm işlem tutarının sağa çarpık dağılımını sıkıştırmayı amaçlamıştır. hour_bucket ise Time değişkeninin 3.600'e bölünüp 24'ün modu alınmasıyla türetilmiş, böylece işlem saatinin günlük döngüsü

kodlanmıştır. Böylece hem ham değişken kümesi hem de genişletilmiş değişken kümesi altında karşılaştırma yapılabilmiştir. Veri kümesinin temizleme sonrasındaki temel özellikleri Tablo 1'de özetlenmektedir.

Tablo 1. Veri kümesinin temizleme sonrası temel özellikleri (Core characteristics of the dataset after cleaning)

Gösterge	Değer
Özgün toplam işlem sayısı	284.807
Özgün dolandırıcılık sayısı	492
Silinen yinelenen kayıt	1.081
Temizleme sonrası toplam işlem	283.726
Temizleme sonrası meşru işlem	283.253
Temizleme sonrası dolandırıcılık işlemi	473
Dolandırıcılık oranı	%0,167
Eğitim / doğrulama / test bölünmesi	170.235 / 56.745 / 56.746

3.3. Modelleme düzeni ve veri analiz yöntemi (Modeling approach and data analysis method)

Veri, sınıf oranını koruyan tabakalı (stratified) bölme ile eğitim, doğrulama ve test kümelerine ayrılmıştır. Eğitim kümesi 170.235, doğrulama kümesi 56.745 ve test kümesi 56.746 gözlem içermektedir. Dolandırıcılık oranı üç alt kümede de yaklaşık yüzde 0,167 düzeyinde korunmuştur. Karşılaştırmaya Lojistik Regresyon (Logistic Regression), Rastgele Orman (Random Forest), Aşırı Gradyan Artırma (Extreme Gradient Boosting, XGBoost) ve Hafif Gradyan Artırma Makinesi (Light Gradient Boosting Machine, LightGBM) modelleri alınmıştır. XGBoost ile LightGBM için sınıf ağırlıkları negatif-pozitif oranı üzerinden tanımlanmış; rastgele ormanda balanced_subsample yaklaşımı, lojistik regresyonda ise class_weight=balanced yaklaşımı kullanılmıştır. Başarım değerlendirmesinde PR-AUC, ROC-AUC, duyarlılık (recall), F1 skoru ve dengelenmiş doğruluk (balanced accuracy) birlikte kullanılmıştır. Ağır dengesiz sınıf yapısı nedeniyle model seçiminin ana ölçütü PR-AUC olarak belirlenmiştir. Karar eşiği sabit 0,50 olarak kabul edilmemiş; her model için doğrulama kümesinde 0,01 ile 0,99 arasındaki eşikler taranmıştır. Yanlış negatif maliyeti kaçırılan dolandırıcılık işleminin tutarı, yanlış pozitif maliyeti ise manuel inceleme maliyeti olarak tanımlanmıştır. Temel senaryoda inceleme maliyeti 10 birim kabul edilmiş, ayrıca 5 ve 20 birimlik iki duyarlılık senaryosu da çalıştırılmıştır.

Bu rastgele tabakalı bölme tek değerlendirme düzeni olarak bırakılmamış; aynı modeller zaman-temelli (kronolojik) bölme altında da yeniden eğitilip değerlendirilmiştir. Kronolojik bölmede tüm kayıtlar Time değişkenine göre sıralanmış, ilk yüzde 60'ı eğitim, sonraki yüzde 20'si doğrulama ve son yüzde 20'si test kümesi olarak ayrılmıştır. Bu düzen, Dal Pozzolo vd.'nin (2018) ve Jurgovsky vd.'nin (2018) belirttiği gibi kart işlemlerinin zamansal bağımlılık taşıdığı durumlar için daha gerçekçi bir değerlendirme zemini sunmakta; eğitim ve test kümeleri arasında kronolojik sızıntıyı önlemektedir. İki bölme stratejisi yan yana raporlanarak rastgele bölmeden kaynaklanabilecek iyimserlik yanlılığı görünür kılınmıştır.

Modeller, farklı öğrenme mantıklarını aynı deney düzeni içinde karşılaştırmaya olanak verecek biçimde seçilmiştir. Lojistik regresyon, ikili sonuç olasılığını logit bağlantısı üzerinden tahmin eden parametrik bir taban sınıflayıcıdır ve kredi kartı dolandırıcılığı literatüründe düzenli olarak başvuru modeli olarak kullanılmaktadır (Ileberi vd., 2022; Alarfaj vd., 2022). Rastgele orman, bootstrap örnekleri üzerinde kurulan çok sayıda karar ağacını rastgele değişken altkümeleriyle birleştirerek varyansı düşüren bir topluluk öğrenmesi yaklaşımıdır (Breiman, 2001). XGBoost, Friedman'ın (2001) gradyan artırma çerçevesini ikinci dereceden optimizasyon ve düzenlileştirme ile ölçeklenebilir hale getiren ağaç tabanlı bir yöntemdir (Chen ve Guestrin, 2016). LightGBM ise histogram tabanlı öğrenme, gradyan temelli tek taraflı örnekleme ve özel öznitelik demetleme mekanizmalarıyla hesaplama maliyetini azaltan bir gradyan artırma uygulamasıdır (Ke vd., 2017). Bu nedenle dört model birlikte ele alındığında, doğrusal bir taban çizgi ile torbalama ve artırma temelli topluluk yapılarının dengesiz veri üzerindeki göreli performansı doğrudan gözlenebilmektedir.

LightGBM modelinin yapılandırması, kütüphaneye özgü iki tasarım tercihinin bu çalışma için anlamlı olduğunu göstermektedir. Birincisi, LightGBM'de subsample (bagging_fraction) parametresi, bagging_freq parametresi pozitif bir değere ayarlanmadığı sürece varsayılan olarak devre dışı kalmaktadır; XGBoost'tan ayrıştığı bu davranış, kütüphanenin varsayılan tasarımından kaynaklanmaktadır. İkincisi, negatif/pozitif örneklerin oranından doğrudan türetilen scale_pos_weight değeri, bu veri kümesi için yaklaşık 600 düzeyine ulaşmakta ve LightGBM'in yaprak-temelli büyüme stratejisi altında dengesiz veri için aşırı agresif kalmaktadır. Bu çalışmada LightGBM'in ana karşılaştırma yapılandırması olarak bagging_freq = 5 ve scale_pos_weight = √(negatif/pozitif) ≈ 24 değerleri kullanılmıştır; bu yapılandırma izleyen bölümlerde Yapılandırma B olarak anılmaktadır. Aynı LightGBM modelinin yapılandırma duyarlılığını belgelemek amacıyla iki karşıt yapılandırma daha sınanmıştır: Yapılandırma A, sıkça başvurulan varsayılan örnekleme yapısını ve oran-temelli scale_pos_weight değerini koruyan (subsample = 0,9, bagging_freq = 0, scale_pos_weight = negatif/pozitif ≈ 600) bir alt çizgiye

karşılık gelmektedir. Yapılandırma C, bagging_freq sıfır kalırken scale_pos_weight yerine is_unbalance = True seçeneğini kullanmaktadır. Bölüm 4'te bu üç yapılandırmanın doğrulama ve test başarımları yan yana raporlanmıştır.

$$TC(\tau) = \sum_{i \in FN(\tau)} a_i + c_{FP} \cdot FP(\tau) \quad (1)$$

Burada $TC(\tau)$ toplam maliyeti, $FN(\tau)$ eşik τ altında kaçırılan dolandırıcılık işlemleri kümesini, a_i i'nci kaçırılmış işlemin tutarını, c_{FP} tek bir yanlış pozitif için inceleme maliyetini ve $FP(\tau)$ yanlış pozitif sayısını göstermektedir.

Temel senaryoda inceleme maliyeti $c_{FP} = 10$ birim olarak alınmıştır. Ayrıca $c_{FP} = 5$ ve $c_{FP} = 20$ birimlik iki duyarlılık senaryosu çalıştırılmıştır. H3 hipotezini sınamak üzere ham ve genişletilmiş öznitelik kümeleri ayrıca karşılaştırılmıştır.

Kesinlik (precision), duyarlılık (recall) ve F1 skoru Denklem (2)–(4) ile hesaplanmıştır.

$$Kesinlik = \frac{TP}{TP+FP} \quad (2)$$

$$Duyarlılık = \frac{TP}{TP+FN} \quad (3)$$

$$F1 = 2 \cdot \frac{Kesinlik \cdot Duyarlılık}{Kesinlik + Duyarlılık} \quad (4)$$

Burada TP doğru pozitifleri, FP yanlış pozitifleri ve FN yanlış negatifleri göstermektedir. Denklem (2)–(4) ile hesaplanan ölçütler, dengesiz sınıf yapısında yanlış alarm ile kaçırılan dolandırıcılık işlemi arasındaki ödünleşimi birlikte değerlendirmek için kullanılmıştır.

H3 hipotezini sınamak amacıyla iki özellik kümesi tanımlanmıştır. Ham özellik kümesi V1–V28, Time ve Amount olmak üzere 30 değişkenden; genişletilmiş özellik kümesi ise bunlara ek olarak log_amount ve hour_bucket ile 32 değişkenden oluşmaktadır. En güçlü ayrıştırma başarımını veren model için yerleşik değişken önemleri ve Shapley Additive Explanations (SHAP) özet grafiği üretilmiş, böylece performans bulguları açıklanabilirlik düzeyiyle birlikte yorumlanmıştır.

3.4. Sınıf dengesizliğine yönelik karşılaştırmalı yaklaşımlar (Comparative approaches to class imbalance)

Sınıf ağırlıklarının doğrudan eğitime dahil edilmesi (XGBoost ve LightGBM için scale_pos_weight, lojistik regresyon ve rastgele orman için class_weight) ana karşılaştırmanın taban çizgisini oluşturmaktadır. Bu çizgiye ek olarak literatürde yaygın kullanılan dört yaklaşım daha aynı eğitim, doğrulama ve test bölmeleri üzerinde sınanmıştır. Birincisi, azınlık sınıfı için sentetik örnek üreten Synthetic Minority Over-sampling Technique (SMOTE) yöntemidir (Chawla vd., 2002). İkincisi, çoğunluk sınıfından rastgele alt örnekleme (RandomUnderSampler) yapılarak eğitim setinin sınıf dağılımı dengelenmiştir. Üçüncüsü, SMOTE ile Tomek bağlantılarının temizlenmesini birleştiren SMOTE-Tomek yaklaşımıdır (Batista vd., 2004). Dördüncüsü, kayıp düzeyinde maliyet duyarlı öğrenme uygulayan focal loss yaklaşımıdır (Lin vd., 2017). Bu yaklaşımda kayıp fonksiyonu $(1 - p_y)^\gamma$ çarpanı ile zor sınıflandırılan örneklere ağırlık vermektedir. Çalışmada $\gamma = 2$, $\alpha = 0{,}25$ değerleri kullanılmıştır. Tüm yöntemler XGBoost taban modeli ile eşleştirilerek aynı eşik tarama prosedürü uygulanmış; doğrulama kümesindeki PR-AUC ve toplam beklenen maliyet üzerinden raporlanmıştır.

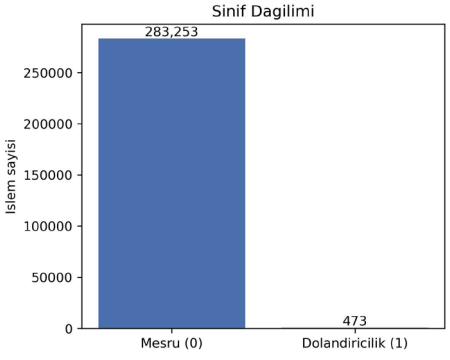
3.5. Olasılık kalibrasyonu (Probability calibration)

Maliyet duyarlı eşik seçimi yalnızca olasılıkların doğru sıralanmasını değil, gerçek olasılığa yakın değerler üretmesini de gerektirir. Bu nedenle birincil model (XGBoost) için iki kalibrasyon prosedürü uygulanmıştır. İlki, doğrulama kümesi üzerinde lojistik bir kalibrasyon eğrisi uyduran Platt sigmoid yaklaşımıdır (Platt, 1999). İkincisi, monoton kısıt altında sıralı izoton regresyonu kullanan isotonic kalibrasyondur (Zadrozny ve Elkan, 2002). Her iki yöntem de eğitilmiş taban modelin doğrulama kümesi olasılıkları üzerinde uydurulmuş, ardından test kümesinde Brier skoru, PR-AUC, ROC-AUC ve maliyet duyarlı eşik seçimi altında hesaplanan toplam maliyet ile değerlendirilmiştir. Brier skoru, kalibrasyon kalitesinin ölçütü olarak doğrudan raporlanmıştır.

4. Bulgular (Findings)

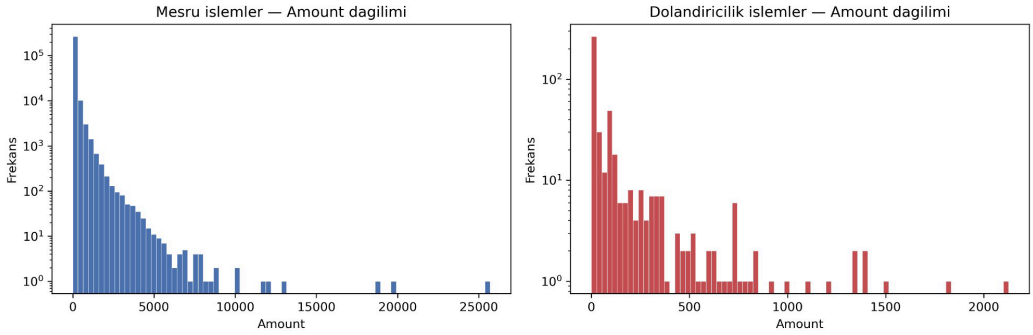
4.1. Veri yapısı ve betimsel bulgular (Data structure and descriptive findings)

Temizleme sonrası veri kümesinde 283.253 meşru işlem ve 473 dolandırıcılık işlemi bulunmaktadır. Dolandırıcılık oranının yüzde 0,167 düzeyinde kalması, modelleme aşamasında sınıf dengesizliğinin temel belirleyici olduğunu göstermektedir. Şekil 1, yinelenen kayıtlar temizlendikten sonra meşru ve dolandırıcılık sınıfları arasındaki keskin dengesizliği açık biçimde göstermektedir.



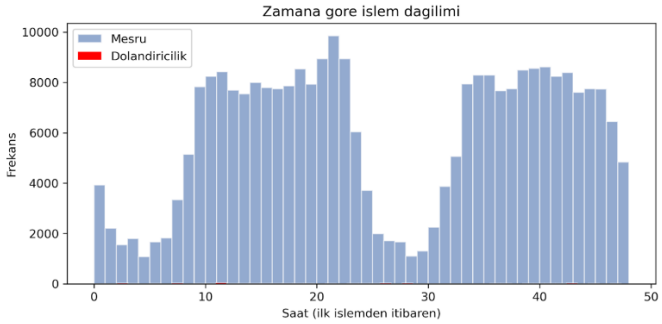
Şekil 1. Yinelenen kayıtlar çıkarıldıktan sonra sınıf dağılımı (Class distribution after duplicate records were removed)

Amount değişkeni sınıfa göre incelendiğinde, meşru işlemlerde ortalama işlem tutarı 88,41, dolandırıcılık işlemlerinde ise 123,87 olarak hesaplanmıştır. Buna karşılık medyan tutarlar sırasıyla 22,00 ve 9,82'dir. Ortalama ile medyan arasındaki ayrışma, her iki sınıfta da sağa çarpık bir dağılıma işaret etmektedir; ancak dolandırıcılık sınıfında kuyruk davranışı daha belirgindir. Şekil 2'de görüldüğü üzere dolandırıcılık işlemleri özellikle düşük tutar bandında yoğunlaşmakta, yüksek tutarlı dolandırıcılık gözlemleri ise seyrekleşmektedir.



Şekil 2. Meşru ve dolandırıcılık işlemlerinde işlem tutarı dağılımı (Transaction amount distribution for legitimate and fraudulent transactions)

Zaman değişkeni iki günlük işlem döngüsünü izlemektedir. Şekil 3'te görüldüğü gibi meşru işlemler belirli saat bloklarında yoğunlaşırken, dolandırıcılık işlemleri aynı eksende çok daha seyrek dağılmaktadır. Bu görünüm, Time değişkeninin tek başına belirleyici olmadığını, ancak diğer PCA bileşenleri ile ayırt edici katkı sağlayabildiğini düşündürmektedir.



Şekil 3. İşlemlerin zaman eksenindeki dağılımı (Distribution of transactions over time)

4.2. Doğrulama ve test kümesi performansı (Validation and test set performance)

Doğrulama kümesinde en yüksek PR-AUC değeri 0,8882 ile XGBoost modelinde elde edilmiştir. Bununla birlikte en düşük doğrulama maliyeti 1.844,88 ile rastgele orman modelindedir. Başka bir ifadeyle, ayrıştırma gücü ile beklenen maliyet aynı modelde birleşmemiştir. Bu durum, model seçiminin kullanılan ölçüte bağlı olarak değiştiğini göstermektedir. Doğrulama kümesindeki ayrıntılı sonuçlar ve seçilen eşikler Tablo 2'de raporlanmıştır.

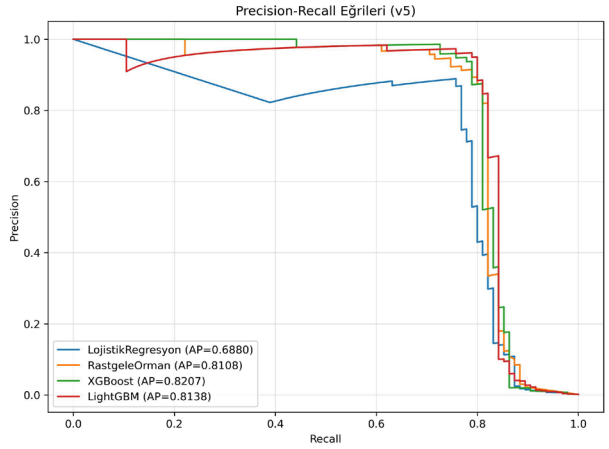
Test kümesinde XGBoost modeli 0,8207 PR-AUC ve 0,9710 ROC-AUC ile en güçlü ayrıştırma başarımını vermiştir. Rastgele orman modeli 0,8108 PR-AUC üretmiş; buna karşılık 3.896,58 toplam maliyet ile operasyonel açıdan daha düşük beklenen kayıp sağlamıştır. Yapılandırma B altındaki LightGBM modeli (Bölüm 3.3 ve Tablo 7) test kümesinde 0,8138 PR-AUC ve 3.837,01 toplam maliyet üretmiş; bu maliyet rastgele ormana yakın bir bant içindedir. Lojistik regresyon ROC-AUC bakımından güçlü görünmesine rağmen F1 ve toplam maliyet bakımından geride kalmıştır. Test kümesindeki ayrıntılı performans değerleri ve maliyet bileşenleri Tablo 3'te yer almaktadır; modellerin duyarlılık-kesinlik ve ROC eğrileri Şekil 4 ile Şekil 5'te sunulmuştur.

Tablo 2. Doğrulama kümesinde model sonuçları ve seçilen eşikler (Model results and selected thresholds in the validation set)

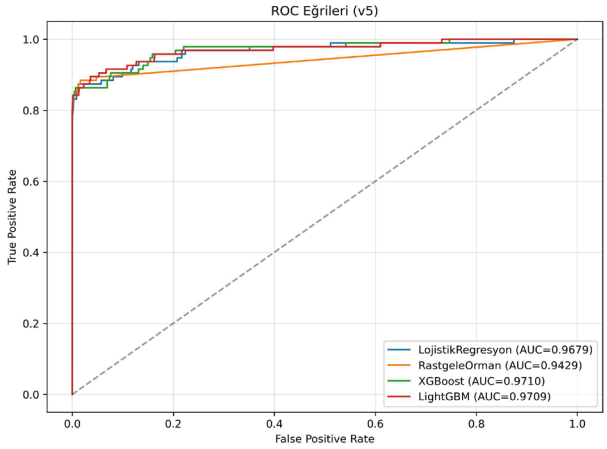
Model	PR-AUC	ROC-AUC	Duyarlılık (0,50)	F1 (0,50)	Seçilen eşik	Doğrulama maliyeti
XGBoost	0,8882	0,9712	0,8723	0,8962	0,63	1.982,14
Rastgele Orman	0,8750	0,9596	0,8191	0,8652	0,13	1.844,88
Lojistik Regresyon	0,7895	0,9827	0,9043	0,1049	0,99	2.384,87
LightGBM	0,8811	0,9751	0,8404	0,8587	0,99	1.865,88

Tablo 3. Test kümesinde model performansı ve maliyet bileşenleri (Model performance and cost components in the test set)

Model	PR-AUC	ROC-AUC	Duyarlılık	F1	Dengelenmiş doğruluk	Yanlış negatif maliyeti	Yanlış pozitif maliyeti	Toplam maliyet
XGBoost	0,8207	0,9710	0,7684	0,8488	0,8842	5.617,49	40,00	5.657,49
Rastgele Orman	0,8108	0,9429	0,8000	0,8352	0,8999	3.786,58	110,00	3.896,58
Lojistik Regresyon	0,6880	0,9679	0,7895	0,6757	0,8943	3.789,34	520,00	4.309,34
LightGBM	0,8138	0,9709	0,8000	0,8588	0,8999	3.777,01	60,00	3.837,01



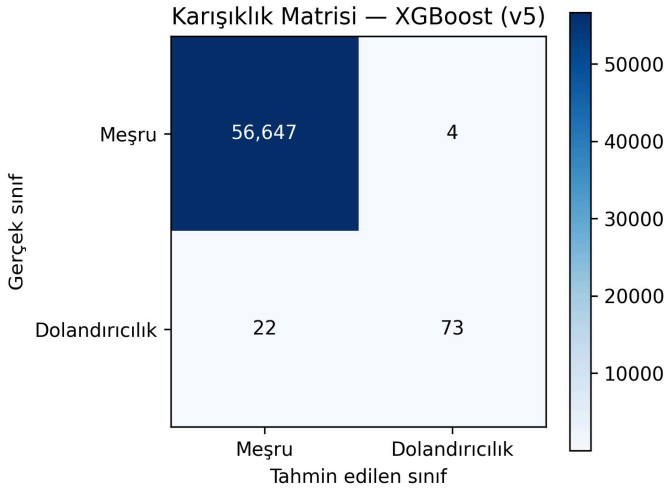
Şekil 4. Modellerin duyarlılık-kesinlik eğrileri (Precision-recall curves of the models)



Şekil 5. Modellerin ROC eğrileri (ROC curves of the models)

4.3. Birincil modelin hata yapısı (Error structure of primary model)

Doğrulama kümesindeki PR-AUC ölçütüne göre seçilen XGBoost modeli, test kümesinde 56.651 meşru işlemin 56.647'sini doğru sınıflandırmış, yalnızca 4 yanlış alarm üretmiştir. Aynı model 95 dolandırıcılık işleminin 73'ünü yakalamış, 22 işlemi kaçırmıştır. Pozitif sınıf için kesinlik 0,9481, duyarlılık 0,7684 ve F1 skoru 0,8488'dir. Sınıflandırma davranışının bütünsel görünümü Şekil 6'daki karışıklık matrisi üzerinden okunabilir.



Şekil 6. XGBoost modeli için karışıklık matrisi (Confusion matrix for the XGBoost model)

4.4. Özellik kümesi karşılaştırması ve maliyet duyarlılığı (Feature set comparison and cost sensitivity)

H3 karşılaştırması, türetilmiş değişkenlerin katkısının sınırlı olduğunu göstermiştir. Genişletilmiş veri kümesi 0,8207 PR-AUC üretirken ham veri kümesi 0,8203 değerinde kalmıştır. Buna karşılık ham veri kümesinin ROC-AUC, duyarlılık, F1 skoru ve toplam maliyet bakımından biraz daha iyi sonuç verdiği görülmüştür. İki özellik kümesinin XGBoost altında karşılaştırmalı sonuçları Tablo 4'te sunulmuştur.

Tablo 4. XGBoost altında ham ve genişletilmiş özellik setlerinin karşılaştırılması (Comparison of raw and extended feature sets under XGBoost)

Özellik Kümesi	PR-AUC	ROC-AUC	Duyarlılık	F1	Dengelenmiş Doğruluk	Toplam Maliyet
Genişletilmiş veri kümesi	0,8207	0,9710	0,7684	0,8488	0,8842	5.657,49
Ham veri kümesi	0,8203	0,9718	0,7789	0,8555	0,8894	3.847,81

Tablo 5. XGBoost modeli için inceleme maliyeti duyarlılık analizi (Review-cost sensitivity analysis for the XGBoost model)

İnceleme Maliyeti(c_{FP})	Optimum Eşik	Duyarlılık	F1	Yanlış Negatif Maliyeti	Yanlış Pozitif Maliyeti	Toplam Maliyet
5	0,06	0,8105	0,7264	3.782,79	200,00	3.982,79
10	0,63	0,7684	0,8488	5.617,49	40,00	5.657,49
20	0,67	0,7684	0,8488	5.617,49	80,00	5.697,49

Tablo 6. Rastgele orman modeli için inceleme maliyeti duyarlılık analizi (Review-cost sensitivity analysis for the Random Forest model)

İnceleme Maliyeti (c_{FP})	Optimum Eşik	Duyarlılık	F1	Yanlış Negatif Maliyeti	Yanlış Pozitif Maliyeti	Toplam Maliyet
5	0,12	0,8105	0,8235	3.782,79	75,00	3.857,79
10	0,13	0,8000	0,8352	3.786,58	110,00	3.896,58
20	0,12	0,8105	0,8235	3.782,79	300,00	4.082,79

Rastgele orman modelinde inceleme maliyetinin 5, 10 ve 20 birim olarak değiştirildiği üç senaryoda optimum eşik 0,12 ile 0,13 arasında dar bir bantta kalmış; toplam maliyet 3.857,79 ile 4.082,79 arasında değişmiştir. XGBoost için aynı senaryolarda optimum eşiğin 0,06 ile 0,67 arasında değişmesi (Tablo 5) dikkate alındığında, rastgele orman modelinin olasılık dağılımının inceleme maliyeti varyasyonuna karşı belirgin biçimde daha dayanıklı olduğu görülmektedir. Bu özellik, inceleme maliyetinin tam olarak bilinmediği veya zamanla değiştiği operasyonel ortamlarda model tercihini etkileyebilecek bir dayanıklılık göstergesi olarak değerlendirilebilir. Tablo 5 ile Tablo 6 birlikte değerlendirildiğinde, model seçiminin yalnızca PR-AUC üzerinden değil, operasyonel inceleme maliyeti varsayımı altında ve maliyet parametresine duyarlılık açısından da incelenmesi gerektiği görülmektedir.

Tablo 7. LightGBM yapılandırma karşılaştırması (Comparison of LightGBM configurations)

Yapılandırma	Doğr. PR-AUC	Doğr. ROC-AUC	Test PR-AUC	Test ROC-AUC	Test Brier	Optimum Eşik	Test Duyarlılığı	Test F1	Test Toplam Maliyeti
Yapılandırma A (subsample = 0,9; bagging_freq = 0)	0,2719	0,8907	0,2769	0,8717	0,0057	0,99	0,8105	0,3860	6.065,01
Yapılandırma B (bagging_freq = 5; scale_pos_weight ≈ 24)	0,8811	0,9751	0,8138	0,9709	0,0004	0,20	0,8000	0,8588	3.837,01
Yapılandırma C (is_unbalance = True; bagging_freq = 0)	0,3186	0,9130	0,3075	0,8399	0,0043	0,99	0,7684	0,4183	5.186,11

Tablo 7'de görüldüğü üzere LightGBM modelinin Yapılandırma A altındaki düşük PR-AUC değeri (0,2769), bagging_freq parametresinin sıfır kalmasından ve aşırı yüksek scale_pos_weight değerinden kaynaklanmaktadır. Yapılandırma B'de bagging_freq = 5 eklenmesi ve scale_pos_weight değerinin √(negatif/pozitif) ile yaklaşık 24 düzeyine indirilmesi, doğrulama PR-AUC değerini 0,8811'e, test PR-AUC değerini ise 0,8138'e taşımış ve modeli XGBoost ile karşılaştırılabilir bir bant içine getirmiştir. Buna karşılık Yapılandırma C'de tek başına is_unbalance = True ayarı (bagging_freq sıfırken) sorunu çözmemekte; doğrulama PR-AUC 0,3186 düzeyinde kalmaktadır. Bu sonuç, sınıf ağırlığı ayarının tek başına yeterli olmadığını göstermektedir.

Tablo 8. Sınıf dengesizliği yöntemleri karşılaştırması (XGBoost taban model) (Comparison of class imbalance methods with XGBoost as the base model)

Yöntem	Doğr. PR-AUC	Test PR-AUC	Test Brier	Optimum Eşik	Test Duyarlılığı	Test F1	FN Maliyeti	FP Maliyeti	Toplam Maliyet
Sınıf ağırlığı (taban çizgi)	0,8882	0,8207	0,0004	0,63	0,7684	0,8488	5.617,49	40,00	5.657,49
SMOTE	0,8792	0,8130	0,0009	0,71	0,8000	0,8128	3.301,73	160,00	3.461,73
RandomUnderSampler	0,8765	0,8154	0,0017	0,96	0,7789	0,8222	3.790,34	110,00	3.900,34
SMOTE-Tomek	0,8792	0,8130	0,0009	0,71	0,8000	0,8128	3.301,73	160,00	3.461,73
Focal loss (γ=2; α=0,25)	0,8847	0,8232	0,0007	0,27	0,7895	0,8427	3.341,29	80,00	3.421,29

SMOTE, alt örnekleme, SMOTE-Tomek ve focal loss yaklaşımlarının taban çizgisi olan sınıf ağırlıklı XGBoost ile karşılaştırması Tablo 8'de sunulmuştur. Beş yöntemin tümünde test PR-AUC değerleri 0,8128 ile 0,8232 arasında dar bir bant içinde kalmış; ayrıştırma kapasitesi yöntemler arasında belirgin bir farklılık göstermemiştir. Buna karşılık toplam maliyet bakımından dikkat çekici bir farklılaşma vardır: focal loss ($\gamma = 2, \alpha = 0{,}25$) 3.421,29 birim ile en düşük maliyeti üretmiş, sınıf ağırlığı taban çizgisini (5.657,49) yaklaşık yüzde 40 oranında geride bırakmıştır. SMOTE ve SMOTE-Tomek yaklaşımları neredeyse özdeş sonuçlar (3.461,73) üretmiş; RandomUnderSampler 3.900,34 düzeyinde kalmıştır. Elde edilen sonuç, sentetik örnekleme tabanlı yöntemlerin sıralama kapasitesinde sınırlı katkı sağladığını, buna karşılık kayıp düzeyinde maliyet duyarlı yeniden ağırlıklandırmanın eşik tarama ile birleştiğinde anlamlı bir kazanım üretebildiğini göstermektedir. Bu görünüm, sentetik dengelemenin başarımı her koşulda artırmadığına ilişkin literatürdeki gözlemlerle uyumludur (Cherif vd., 2023; Baisholan vd., 2025). Gerçek sınıf dağılımı korunduğunda sıralama kazanımları küçülmektedir.

Tablo 9. Olasılık kalibrasyonu sonuçları (XGBoost taban model) (Probability calibration results with XGBoost as the base model)

Kalibrasyon	Doğr. Brier	Test Brier	Test PR-AUC	Test ROC-AUC	Optimum Eşik	Test Duyarlılığı	Test F1	Toplam Maliyet
Kalibrasyonsuz	0,0003	0,0004	0,8207	0,9710	0,63	0,7684	0,8488	5.657,49
Platt sigmoid	0,0003	0,0004	0,8207	0,9710	0,17	0,7684	0,8488	5.657,49
Isotonic	0,0003	0,0004	0,7888	0,9445	0,50	0,7684	0,8488	5.657,49

Tablo 9'da raporlanan kalibrasyon sonuçları, XGBoost taban modelinin halihazırda iyi kalibre olduğunu göstermektedir. Üç yaklaşımda da (kalibrasyonsuz, Platt sigmoid ve isotonic) Brier skoru 0,0003 (doğrulama) ve 0,0004 (test) düzeyinde değişmemiştir. Platt sigmoid kalibrasyonu, optimum eşik değerini 0,63'ten 0,17'ye taşımakla birlikte test maliyetinde değişiklik üretmemiştir. Bu durum, sigmoid dönüşümünün olasılıkları monoton biçimde yeniden ölçeklendirip ayrıştırma sıralamasını koruduğunu yansıtmaktadır. Isotonic kalibrasyon ise PR-AUC değerini 0,8207'den 0,7888'e, ROC-AUC değerini ise 0,9710'dan 0,9445'e düşürmüştür. Bu bulgu, izoton regresyonunun seyrek pozitif sınıf altında parça parça sabit fonksiyon uydururken sıralama bilgisinin bir kısmını yitirebildiğine dair bilinen gözlemle uyumludur. Bu sonuç, mevcut veri ve model yapılandırması altında ek bir kalibrasyon adımının operasyonel kazanım sağlamadığını, ancak farklı veri kümeleri veya kalibrasyon kalitesi başlangıçta zayıf olan modeller için kalibrasyon prosedürünün rutin bir denetim adımı olarak değerlendirilmesi gerektiğini göstermektedir (Niculescu-Mizil ve Caruana, 2005).

Tablo 10. Kronolojik (zaman-temelli) bölme altında test sonuçları (Test results under chronological time-based split)

Model	Test PR-AUC	Test ROC-AUC	Optimum Eşik	Duyarlılık	F1	Dengelenmiş Doğruluk	FN Maliyeti	FP Maliyeti	Toplam Maliyet
Rastgele Orman	0,8109	0,9460	0,27	0,7568	0,8296	0,8783	2.638,27	50,00	2.688,27
XGBoost	0,8019	0,9721	0,59	0,7162	0,8030	0,8581	3.675,20	50,00	3.725,20
Lojistik Regresyon	0,7575	0,9840	0,99	0,7973	0,6310	0,8982	2.397,79	540,00	2.937,79
LightGBM	0,7345	0,9556	0,88	0,6892	0,7846	0,8446	3.721,71	50,00	3.771,71

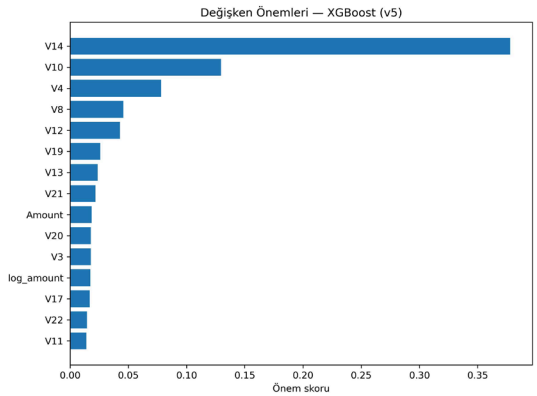
Kronolojik bölme altında elde edilen sonuçlar Tablo 10'da raporlanmıştır. Tablo 10 ile Tablo 3 yan yana değerlendirildiğinde, modellerin göreli sıralamasının iki bölme stratejisi arasında belirgin biçimde değiştiği görülmektedir. Rastgele orman maliyet bakımından her iki bölme stratejisinde de en iyi konumunu korumakla birlikte; XGBoost ile LightGBM arasında rastgele bölmede gözlenen yakın bant kronolojik bölmede genişlemekte (XGBoost 0,8019, LightGBM 0,7345) ve LightGBM'in zaman kayması altında daha az dayanıklı kaldığı görülmektedir. Bu farklılaşma, Dal Pozzolo vd.'nin (2018) ve Jurgovsky vd.'nin (2018) kronolojik

akış altında işlem örüntülerinin değişebileceğine ilişkin tespitlerini somut olarak desteklemektedir. Rastgele bölmenin ürettiği iyimser değerlendirmeyi düzeltmek için kronolojik bölmenin ek bir değerlendirme katmanı olarak kullanılması, modellerin operasyonel ortamda nasıl davranacağı hakkında daha gerçekçi bir resim sunmaktadır.

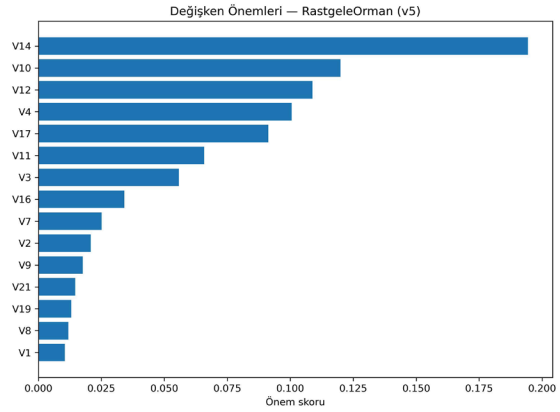
4.5. Değişken önemleri ve Shapley temelli çözümleme (Variable importance and Shapley-based analysis)

Yerleşik değişken önemleri, XGBoost modelinde V14 değişkeninin açık ara en yüksek katkıyı verdiğini göstermektedir (Şekil 7). Bunu V10, V4, V8 ve V12 izlemektedir. Amount ve log_amount değişkenleri ilk on beş değişken içinde yer almakla birlikte ağırlıkları sınırlıdır. Bu bulgu, basit tutar ve zaman türetmelerinin modele katkısının bulunduğunu, ancak belirleyici olmadığını göstermektedir.

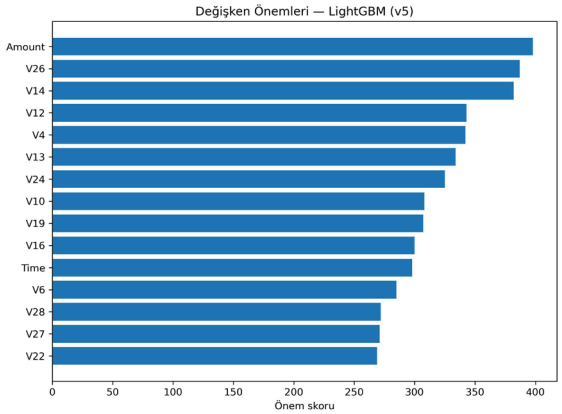
Modeller arası karşılaştırmaya bakıldığında üç farklı kullanım örüntüsü görünmektedir (Şekil 8 ve Şekil 9).



Şekil 7. XGBoost modeli için değişken önemleri (Variable importances for the XGBoost model)



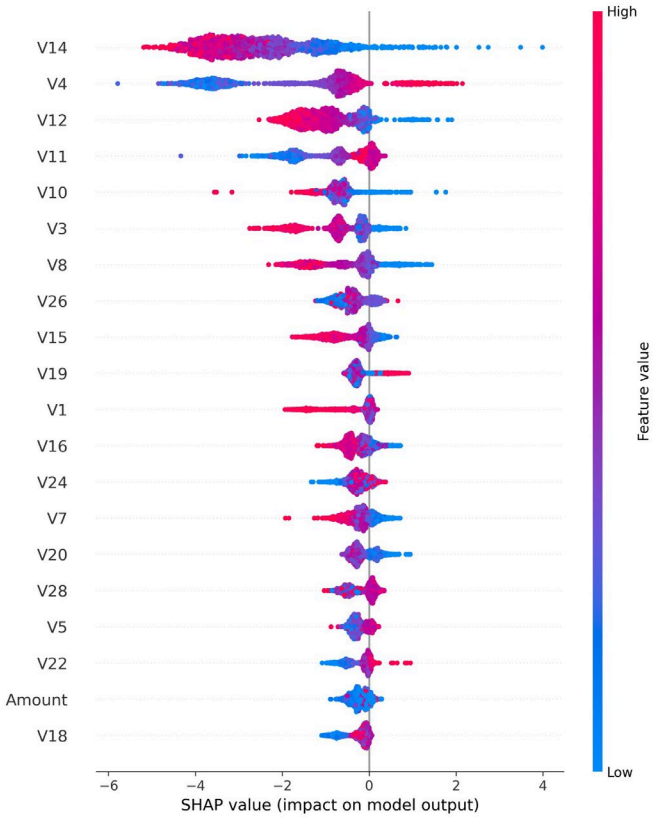
Şekil 8. Rastgele orman modeli için değişken önemleri (Variable importances for the random forest model)



Şekil 9. LightGBM modeli için değişken önemleri (Variable importances for the LightGBM model)

SHAP özet grafiği (Şekil 10) de benzer bir sıralama sunmaktadır. Özellikle V14 için düşük değerler tahmini dolandırıcılık yönüne iterken, V4 ve V12 için yüksek ve düşük gözlemlerin farklı yönlerde katkı verdiği görülmektedir. Bu görünüm, modelin tek bir değişkene dayanmadığını; karar yapısının birkaç baskın bileşen etrafında yoğunlaştığını göstermektedir.

Rastgele orman modelinde V14 yine en baskın değişken olmakla birlikte; log_amount ve hour_bucket değişkenleri en yüksek önemli on beş değişken içinde yer almamaktadır. Bu durum, rastgele orman modelinin türetilmiş değişkenlerden anlamlı ölçüde yararlanmadığını ve ayrımın büyük ölçüde PCA bileşenlerine dayandığını göstermektedir. LightGBM modelinde ise farklı bir örüntü gözlenmektedir: Amount değişkeni en yüksek önem skoruna sahip değişken olarak öne çıkmakta, V26, V14, V12 ve V4 onu izlemektedir. log_amount yine ilk on beş içinde yer almamaktadır. LightGBM'in histogram tabanlı bölme stratejisi Amount değişkenini sürekli sayısal özniteliklerden farklı biçimde değerlendirebilmekte; XGBoost ise aynı değişkeni görece düşük bir sırada tutmaktadır. Bu farklılaşma, üç ağaç tabanlı modelin aynı veri kümesinde bile öznitelik kullanım örüntülerinin uygulamaya özgü kalabildiğini ve birden fazla modelin değişken önemleri üzerinden okunmasının daha güvenilir bir açıklanabilirlik resmi sunduğunu göstermektedir.



Şekil 10. XGBoost modeli için SHAP özet grafiği (SHAP summary plot for the XGBoost model)

5. Tartışma (Discussion)

Bulgular, kredi kartı dolandırıcılığı tespitinde tek bir başarı ölçütü üzerinden karar vermenin yetersiz olduğunu göstermektedir. XGBoost modeli hem doğrulama hem de test aşamasında PR-AUC bakımından ilk sırada yer almıştır. Bu sonuç, boosting tabanlı yapıların kredi kartı dolandırıcılığı veri kümelerinde güçlü bir ayrıştırma kapasitesi sunduğunu bildiren güncel çalışmalarla uyumludur (Ahmed vd., 2025; Altan ve Zafer, 2024; Tayebi ve El Kafhali, 2025). Bununla birlikte, toplam beklenen maliyet bakımından en iyi model rastgele ormandır. Afriyie vd. (2023) ile Albalawi ve Dardouri'nin (2025) rastgele orman lehine bulguları, operasyonel maliyet ve yanlış alarm dengesi birlikte ele alındığında daha anlamlı hale gelmektedir.

XGBoost ve LightGBM iki gradyan artırma uygulaması olarak ilkesel düzeyde birbirine yakındır; her ikisi de eklemeli ağaç toplulukları üzerinden kayıp fonksiyonunu en aza indirmekte ve sınıf dengesizliğini scale_pos_weight benzeri parametrelerle hesaba katmaktadır. Aralarındaki başlıca farklar, bölme stratejisi (XGBoost düzey-temelli, LightGBM yaprak-temelli büyüme), histogram tabanlı hızlandırma ve örnekleme kontrolüdür. Bu çalışmada Yapılandırma A altında gözlenen büyük PR-AUC farkı (XGBoost 0,8207, LightGBM 0,2769) modelin yapısal kapasitesinden değil, iki uygulamanın parametre semantiğindeki farklılıktan kaynaklanmaktadır. LightGBM'de subsample (bagging_fraction) parametresi varsayılan bagging_freq = 0 değerinde sessizce devre dışı kalmaktadır; aynı parametre adına sahip XGBoost ise her iterasyonda aktif biçimde örnekleme yapmaktadır. Bu durum LightGBM'in eğitim sırasında varyans azaltıcı örnekleme yapmamasına yol açmıştır. Buna ek olarak negatif/pozitif örnek oranından doğrudan türetilen scale_pos_weight değeri yaklaşık 600 düzeyine ulaşmakta ve LightGBM'in yaprak-temelli büyüme stratejisi altında dengesiz veri için aşırı agresif kalmaktadır. Yapılandırma B'de bagging_freq parametresinin 5'e ayarlanması ve scale_pos_weight değerinin √(negatif/pozitif) ile yaklaşık 24'e indirilmesi LightGBM'in test PR-AUC değerini 0,8138 düzeyine, toplam maliyetini ise 3.837,01 düzeyine taşımış ve modeli XGBoost ile karşılaştırılabilir bir bant içine getirmiştir. Buna karşılık Yapılandırma C'de tek başına is_unbalance = True alternatifi (bagging_freq sıfır kalırken) sorunu çözmemekte; test PR-AUC 0,3075 düzeyinde kalmaktadır. Yapılandırma A'dan B'ye geçişte bagging_freq ve scale_pos_weight birlikte değiştirildiği için, gözlenen iyileşme bu iki müdahalenin ortak etkisini yansıtmaktadır; yalnızca is_unbalance seçeneğinin etkinleştirildiği Yapılandırma C'nin düşük başarımı ise varyans azaltıcı örneklemenin bu iyileşmede belirgin bir rol oynadığını düşündürmektedir. Daha genel düzeyde bu gözlem, gradyan artırma kütüphaneleri arasında parametre adlarının aynı olması durumunda dahi semantiğin farklılaşabildiğini ve karşılaştırmalı çalışmalarda

kütüphaneye özgü varsayılan davranışların açıkça denetlenmesi gerektiğini göstermektedir.

Çalışmanın en belirgin katkısı, model sıralamasının seçilen performans ölçütüne ve karar eşiğine bağlı olarak değiştiğini aynı deney düzeni içinde göstermesidir. PR-AUC bakımından üstün olan XGBoost, maliyet bileşenleri dikkate alındığında rastgele ormanın gerisine düşmüştür. Bu ayrım, açık karşılaştırma veri kümesi çalışmalarında sıklıkla ikinci planda kalan operasyonel karar problemini görünür kılmaktadır. C-Rella vd. (2024), iki boyutlu maliyet duyarlı eşik seçiminin tek boyutlu yaklaşımlara göre daha düşük beklenen maliyet ürettiğini göstermiştir. Bu bulgu, maliyet odaklı değerlendirmenin model sıralamasını değiştirebildiğine dair tespitimizi desteklemektedir.

Eşik seçiminin model davranışının ayrılmaz bir parçası olduğu açık biçimde görülmüştür. XGBoost için inceleme maliyeti 5, 10 ve 20 birim olduğunda optimum eşik 0,06 ile 0,67 arasında değişmiştir. Eşik düştüğünde duyarlılık artmış, buna karşılık F1 skoru gerilemiştir. Bu kayıp-kazanç ilişkisi, Bahnsen vd. (2016) ile Baisholan vd.'nin (2025) vurguladığı maliyet duyarlı karar mantığıyla uyumludur. Finansal kurumlar açısından model çıktısının işletmeye alınmasında yalnızca olasılık tahmini değil, bu tahminin hangi inceleme maliyeti ve hangi kayıp fonksiyonu altında kullanılacağı da belirleyicidir.

Sınıf dengesizliğine yönelik beş yaklaşımın aynı eşik tarama prosedürü altında karşılaştırılması (Tablo 8), test ayrıştırma kapasitesinin (PR-AUC) yöntemler arasında dar bir bant içinde kaldığını, buna karşılık toplam beklenen maliyetin yöntem seçimine duyarlı olduğunu göstermektedir. Sınıf ağırlığı taban çizgisinin ürettiği 5.657,49 birimlik maliyet, focal loss ($\gamma = 2$, $\alpha = 0{,}25$) ile 3.421,29 birime düşmüştür. Bu fark, ayrıştırma sıralamasının yakın kaldığı durumlarda dahi olasılıkların kayıp fonksiyonu içinde nasıl ağırlıklandırıldığının operasyonel sonuçları belirleyebildiğini göstermektedir. Sentetik örnekleme tabanlı yöntemler (SMOTE, SMOTE-Tomek, RandomUnderSampler) ise ne ayrıştırma sıralamasında ne de toplam maliyette taban çizgiye göre net bir üstünlük sağlamıştır. Bu sonuç, sentetik dengelemenin katkısının veri yapısına ve değerlendirme ölçütüne bağlı kaldığına ilişkin literatürdeki değerlendirmelerle uyumludur (Cherif vd., 2023; Baisholan vd., 2025). Söz konusu bulgu, mevcut taban çizginin göreli üstünlüğü tartışmasını sayısal olarak somutlaştırmakta ve kayıp düzeyinde maliyet duyarlı yeniden ağırlıklandırmanın bu problem ailesinde çalışılmaya değer bir yön olduğunu işaret etmektedir.

Focal loss yaklaşımının bu veri kümesinde gözlenen göreli üstünlüğü mekanik olarak şu üç noktaya dayanmaktadır. İlk olarak, sınıf ağırlığı taban çizgisi her bir pozitif örneğe sabit bir ağırlık atfetmekte; pozitif sınıf içindeki kolay ve zor örnekler arasında bir ayrım gözetmemektedir. Buna karşılık focal loss, çapraz entropi terimini $(1 - p_y)^{\gamma}$ çarpanı ile ölçeklendirerek halihazırda yüksek doğrulukla sınıflandırılan örneklerin gradyan katkısını üstel biçimde azaltmakta; modelin kapasitesini ayrımı zor olan azınlık örneklerine yönlendirmektedir (Lin vd., 2017). $\gamma = 2$ yapılandırmasında, $p_y = 0{,}9$ düzeyinde sınıflandırılan kolay bir örneğin kayıp terimi sınıf ağırlığı düzenine göre yüz kat azalmaktadır. Bu sönümleme, eğitim sırasında modelin azınlık sınıfının sınır bölgesine odaklanmasını sağlamaktadır. İkinci olarak, $\alpha = 0{,}25$ dengeleme parametresi sınıflar arası ağırlık dengesini ayarlayarak, ham sınıf oranından (yaklaşık 1/600) doğrudan türetilen scale_pos_weight yapısının yarattığı aşırı duyarlılığı engellemektedir. Üçüncü olarak, Worldline-ULB veri kümesindeki pozitif örneklerin PCA bileşenlerinde dağınık ve heterojen bir görünüm sergilemesi, sıkı bir karar yüzeyinin yalnızca zor örnekler üzerinden öğrenilmesini gerektirmektedir; focal loss'un odaklanma davranışı bu yapıya uygun düşmektedir. Bu mekanizmaların bir araya gelmesi, ayrıştırma sıralaması taban çizgisine yakın kalsa da olasılık düzeyindeki dağılımın eşik tarama altında daha düşük toplam maliyete dönüşmesini açıklamaktadır.

H3 hipotezi desteklenmemiştir. log_amount ve hour_bucket değişkenleri doğrulama aşamasında çok küçük bir PR-AUC artışı üretmiş; ancak test kümesinde F1 skoru ve toplam maliyet bakımından ham özellik veri kümesinin önüne geçememiştir. Bu bulgu, Bahnsen vd.'nin (2016) işlem geçmişinden türetilen toplulaştırılmış ve periyodik değişkenlerinin, bu çalışmada kullanılan tek adımlı türetmelerden farklı bir bilgi kaynağına dayandığını düşündürmektedir. Ghosh Dastidar vd. (2022) ise bu tür toplulaştırmaların elle üretilmesi yerine öğrenilebilir bir yapıyla elde edilebileceğini göstermiştir. Başka bir ifadeyle, her öznitelik mühendisliği müdahalesi aynı ölçüde değer üretmemektedir; katkının veri bağlamı içinde deneysel olarak sınanması gerekmektedir.

H3'ün başarısızlığı yapısal bir nedene dayanmaktadır. Worldline-ULB karşılaştırma veri kümesinde V1–V28 değişkenleri PCA ile anonimleştirilmiş bileşenlerdir; Time ve Amount ise PCA ile dönüştürülmeyen iki ayrı değişkendir. Bu nedenle log_amount ve hour_bucket değişkenlerinin sınırlı katkısı, bu bilgilerin PCA bilşenlerinde önceden kodlandığı varsayımıyla açıklanamaz. Bu çalışmada türetilen log_amount değişkeni Amount'un determinist log1p dönüşümü, hour_bucket değişkeni ise Time'ın 3.600'e bölünüp 24'ün modu alınmasıyla elde edilen kapalı formlu bir türevidir. Eğitim kümesi üzerinde yapılan tanı analizinde log_amount ile Amount arasında Spearman korelasyonu 1,00 (determinist beklendiği üzere), log_amount ile V2 arasında ise −0,50 düzeyinde bulunmuştur. hour_bucket ile Time arasındaki Spearman korelasyonu yaklaşık 0,47, hour_bucket ile V12 arasında ise 0,23 düzeyindedir. log_amount ve hour_bucket ile bazı PCA bileşenleri arasında orta düzeyde korelasyon gözlenmesi, bu türetmelerin ürettiği ayrımın bir kısmının mevcut değişken uzayıyla örtüşebileceğini düşündürmektedir. Bulgular, bu iki

deterministik türetmenin kullanılan model ve deney düzeninde ek ve kararlı bir ayırt edici bilgi sağlamadığını; sağladıkları bilginin büyük ölçüde mevcut değişken uzayıyla örtüştüğünü göstermektedir. Bu durum, Bahnsen vd.'nin (2016) kart sahibi işlem geçmişine dayanan toplulaştırılmış değişkenleriyle mevcut çalışmadaki tek adımlı türetmeler arasındaki farkı ortaya koymaktadır: Worldline-ULB veri kümesi müşteri kimliği, satıcı ve kart bazlı geçmiş bilgisi içermediğinden, bu türden toplulaştırmaların üretilmesi mümkün değildir. Kart sahibi geçmişi ve satıcı bilgisi içermeyen, PCA ile anonimleştirilmiş veri yapılarında benzer öznitelik mühendisliği müdahalelerinin sınırlı kalması beklenebilir.

Xie vd. (2024), Jurgovsky vd. (2018) ve Tang ile Liang'ın (2024) çalışmaları, işlem ilişkileri ve zaman bağımlılığını doğrudan modelleyen mimarilerin ek kazanımlar sağlayabildiğini göstermektedir. Bu çalışmada kullanılan veri kümesi, müşteri kimliği, satıcı ağı ya da kurumsal ilişki grafikleri gibi bilgileri içermediği için sekans ve grafik yaklaşımlarının avantajı sınanamamıştır. Buna karşın SHAP bulguları, model kararının belirli PCA bileşenleri üzerinde yoğunlaştığını ve özellikle V14 ile V12 bileşenlerinin ayırt edici rol oynadığını göstermektedir. Baisholan vd. (2025) de açıklanabilirlik çözümlemesinde benzer değişkenlerin öne çıktığını bildirmektedir.

Çalışmanın başlıca sınırlılıkları veri kümesinin iki günlük bir dönemi kapsaması, değişkenlerin PCA ile anonimleştirilmiş olması ve maliyet fonksiyonunun sabit inceleme maliyeti ile işlem tutarı üzerinden kurulmasıdır. Gerçek kurumsal uygulamalarda ters ibraz (chargeback) süreci, müşteri kaybı, operasyon ekip kapasitesi, itiraz maliyeti ve düzenleyici risk gibi bileşenler bu maliyet yapısını değiştirebilir. Buna karşın çalışma, aynı açık veri üzerinde ayrıştırma başarımı, maliyet ve açıklanabilirlik boyutlarını birlikte ele alarak yöntem ve uygulama tartışmasına somut bir katkı sunmaktadır.

6. Sonuçlar (Conclusions)

Bu çalışma, açık erişimli ve ağır dengesiz kredi kartı işlem verisi üzerinde dört sınıflayıcıyı maliyet duyarlı karar çerçevesinde karşılaştırmıştır. Doğrulama kümesinde en yüksek PR-AUC değeri XGBoost modelinde, en düşük doğrulama maliyeti ise rastgele orman modelinde elde edilmiştir. Test kümesinde XGBoost ayrıştırma başarımı bakımından ilk sırada yer alırken, toplam maliyet bakımından rastgele orman daha iyi sonuç vermiştir. Bu bulgu, model seçiminin başarı tanımına bağlı olduğunu göstermektedir.

Hipotezler birlikte değerlendirildiğinde H1 desteklenmiş, H2 kısmen desteklenmiş, H3 desteklenmemiş ve H4 desteklenmiştir. Bu hipotez çerçevesinin ötesinde, genişletilmiş karşılaştırma üç ek bulgu ortaya koymuştur. Birincisi, sınıf dengesizliği yaklaşımının seçimi taban çizginin önüne geçen bir karar değişkeni olarak öne çıkmaktadır: focal loss yaklaşımı, ayrıştırma sıralaması yakın kalsa da toplam maliyeti sınıf ağırlığı taban çizgisine göre yüzde 40 oranında düşürmüştür (5.657,49 birimden 3.421,29 birime). Bu sonuç, yöntem ailesi düzeyinde kayıp fonksiyonunun yeniden tanımlanmasının, model değişikliğinden daha büyük bir operasyonel kazanım üretebildiğini göstermektedir. İkincisi, LightGBM'in Yapılandırma A altındaki düşük başarımı modelin yapısal kapasitesinden değil, kütüphaneye özgü parametre semantiğinden kaynaklanmıştır; Yapılandırma B altında LightGBM, XGBoost ile karşılaştırılabilir bir ayrıştırma bandına ve rastgele ormana yakın bir maliyet düzeyine taşınmıştır. Bu durum, gradyan artırma kütüphaneleri arasındaki parametre farklılıklarının karşılaştırmalı çalışmalarda açıkça denetlenmesi gerektiği yönünde uyarı niteliğindedir. Üçüncüsü, kronolojik bölme altında modellerin göreli sıralaması korunmamakta; özellikle LightGBM zaman kayması altında daha az dayanıklı kalmaktadır. Bu farklılaşma, rastgele bölmenin tek başına kullanılmasının iyimserlik yanlılığı yaratabileceğini somut biçimde göstermektedir. Uygulama düzeyinde değerlendirme şu sıralamayı önermektedir: kayıp düzeyinde maliyet duyarlı yeniden ağırlıklandırma (focal loss) ile XGBoost, en düşük toplam maliyeti üreten yapı olarak ilk tercih konumundadır. Sınıf ağırlığı taban çizgisi içinde kalınması durumunda ise Yapılandırma B altındaki LightGBM ile rastgele orman yakın bir bant içinde kalmaktadır; seçim, eşik dayanıklılığı ve operasyonel inceleme maliyetinin bilinirliğine göre yapılabilir.

Çalışmanın literatüre katkısı dört boyutta özetlenebilir. İlk olarak, açık karşılaştırma veri kümesi üzerinde maliyet duyarlı eşik seçimi ile ayrıştırma başarımını birlikte değerlendirmektedir. İkinci olarak, yanlış negatif ve yanlış pozitif maliyetlerini ayrıştırarak model tercihini operasyonel zemine taşımaktadır. Üçüncü olarak, sınıf dengesizliğine yönelik beş yaklaşımı (sınıf ağırlığı, SMOTE, alt örnekleme, SMOTE-Tomek, focal loss) aynı eşik tarama prosedürü altında karşılaştırarak, kayıp düzeyinde maliyet duyarlı yeniden ağırlıklandırmanın taban çizgi üzerinde anlamlı bir kazanım üretebildiğini somut sayılarla göstermektedir. Dördüncü olarak, SHAP ve değişken önemleri üzerinden model kararının hangi bileşenlere dayandığını görünür kılmakta; üç ağaç tabanlı modelin aynı veri üzerinde farklı öznitelik kullanım örüntüleri sergileyebildiğini ortaya koymaktadır. Gelecek çalışmaların aynı maliyet çerçevesini kurumsal veri kümeleri, daha uzun zaman ufukları, ilişkisel grafik yapıları ve çevrim içi öğrenme düzenleri üzerinde sınaması, bulguların dış geçerliliğini güçlendirecektir. Ayrıca, maliyet fonksiyonunun ters ibraz (chargeback) süreçleri, müşteri yaşam boyu değeri ve düzenleyici cezalar gibi gerçek bileşenlerle genişletilmesi, operasyonel uygulamaya daha yakın sonuçlar verecektir.

Kaynaklar (References)

Abdul Salam, M., Fouad, K. M., Elbably, D. L. ve Elsayed, S. M. 2024. Federated learning model for credit card fraud detection with data balancing techniques. Neural Computing and Applications, 36(11), 6231–6256.

Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A. ve Eshun, J. 2023. A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decision Analytics Journal, 6, 100163.

Ahmed, K. H., Axelsson, S., Li, Y. ve Sagheer, A. M. 2025. A credit card fraud detection approach based on ensemble machine learning classifier with hybrid data sampling. Machine Learning with Applications, 20, 100675.

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M. ve Ahmed, M. 2022. Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. IEEE Access, 10, 39700–39715.

Albalawi, T. ve Dardouri, S. 2025. Enhancing credit card fraud detection using traditional and deep learning models with class imbalance mitigation. Frontiers in Artificial Intelligence, 8, 1643292.

Altan, G. ve Zafer, M. R. 2024. Denetimli makine öğrenmesi yöntemleri ile kredi kartı sahteciliğini tahmin etme: karşılaştırmalı analiz. İktisat Politikası Araştırmaları Dergisi, 11(2), 242–262.

Bahnsen, A. C., Aouada, D., Stojanovic, A. ve Ottersten, B. 2016. Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 51, 134–142.

Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T. ve Baisholanova, K. 2025. FraudX AI: An interpretable machine learning framework for credit card fraud detection on imbalanced datasets. Computers, 14(4), 120.

Batista, G. E. A. P. A., Prati, R. C. ve Monard, M. C. 2004. A study of the behavior of several methods for balancing machine learning training data. ACM SIGKDD Explorations Newsletter, 6(1), 20–29.

Breiman, L. 2001. Random forests. Machine Learning, 45(1), 5–32.

Btoush, E., Zhou, X., Gururajan, R., Chan, K. C. ve Alsodi, O. 2025. Achieving excellence in cyber fraud detection: A hybrid ML+DL ensemble approach for credit cards. Applied Sciences, 15(3), 1081.

C-Rella, J., Cao, R. ve Vilar, J. M. 2024. Cost-sensitive thresholding over a two-dimensional decision region for fraud detection. Information Sciences, 657, 119956.

Chawla, N. V., Bowyer, K. W., Hall, L. O. ve Kegelmeyer, W. P. 2002. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357.

Chen, T. ve Guestrin, C. 2016. XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M. ve Imine, A. 2023. Credit card fraud detection in the era of disruptive technologies: A systematic review. Journal of King Saud University – Computer and Information Sciences, 35(1), 145–174.

Chhabra, R., Goswami, S. ve Ranjan, R. K. 2024. A voting ensemble machine learning based credit card fraud detection using highly imbalance data. Multimedia Tools and Applications, 83(18), 54729–54753.

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C. ve Bontempi, G. 2018. Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), 3784–3797.

Fanai, H. ve Abbasimehr, H. 2023. A novel combined approach based on deep Autoencoder and deep classifiers for credit card fraud detection. Expert Systems with Applications, 217, 119562.

Friedman, J. H. 2001. Greedy function approximation: A gradient boosting machine. The Annals of Statistics, 29(5), 1189–1232.

Ghosh Dastidar, K., Jurgovsky, J., Siblini, W. ve Granitzer, M. 2022. NAG: Neural feature aggregation framework for credit card fraud detection. Knowledge and Information Systems, 64(3), 831–858.

Güney, İ. ve Selvi, İ. H. 2023. Makine Öğrenmesi Yöntemleriyle Anormal İçme Suyu Tüketimlerinin Tespit Edilmesi ve Tahmin Modellerinin Geliştirilmesi. Journal of Intelligent Systems: Theory and Applications, 6(2), 159-173.

Ileberi, E., Sun, Y. ve Wang, Z. 2022. A machine learning based credit card fraud detection using the GA algorithm for feature selection. Journal of Big Data, 9, 24.

Jurgovsky, J., Granitzer, M., Ziegler, K., Calabretto, S., Portier, P.-E., He-Guelton, L. ve Caelen, O. 2018. Sequence classification for credit-card fraud detection. Expert Systems with Applications, 100, 234–245.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. ve Liu, T.-Y. 2017. LightGBM: A highly efficient gradient boosting decision tree. Advances in Neural Information Processing Systems, 30, 3146–3154.

Lin, T.-Y., Goyal, P., Girshick, R., He, K. ve Dollár, P. 2017. Focal loss for dense object detection. Proceedings of the IEEE International Conference on Computer Vision, 2999–3007.

Motie, S. ve Raahemi, B. 2024. Financial fraud detection using graph neural networks: A systematic review. Expert Systems with Applications, 240, 122156.

Niculescu-Mizil, A. ve Caruana, R. 2005. Predicting good probabilities with supervised learning. Proceedings of the 22nd International Conference on Machine Learning, 625–632.

Platt, J. C. 1999. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. In A. J. Smola, P. Bartlett, B. Schölkopf ve D. Schuurmans (Eds.), Advances in Large Margin Classifiers, MIT Press, 61–74.

Randhawa, K., Loo, C. K., Seera, M., Lim, C. P. ve Nandi, A. K. 2018. Credit card fraud detection using AdaBoost and majority voting. IEEE Access, 6, 14277–14284.

Sinap, V. 2024. Comparative analysis of machine learning techniques for credit card fraud detection: Dealing with imbalanced datasets. Turkish Journal of Engineering, 8(2), 196–208.

Tang, Y. ve Liang, Y. 2024. Credit card fraud detection based on federated graph learning. Expert Systems with Applications, 256, 124979.

Tayebi, M. ve El Kafhali, S. 2025. A novel approach based on XGBoost classifier and Bayesian optimization for credit card fraud detection. Cyber Security and Applications, 3, 100093.

Tekin, R. ve Yaman, O. 2023. Akıllı Ev Sistemleri için XGBoost Tabanlı Saldırı Tespit Yöntemi. Journal of Intelligent Systems: Theory and Applications, 6(2), 152-158.

Xie, Y., Liu, G., Yan, C., Jiang, C., Zhou, M. C. ve Li, M. 2024. Learning transactional behavioral representations for credit card fraud detection. IEEE Transactions on Neural Networks and Learning Systems, 35(4), 5735–5748.

Yıltırak, İ. ve Öztürk, A. 2025. Nesnelerin İnterneti Saldırılarının Tespitinde Makine Öğrenmesi Algoritmalarının Performanslarının Karşılaştırılması. Journal of Intelligent Systems: Theory and Applications, 8(2), 130-140.

Zadrozny, B. ve Elkan, C. 2002. Transforming classifier scores into accurate multiclass probability estimates. Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 694–699.

Ek (Appendix)

Ek Tablo 1. Yöntem ve veri bazlı literatürün tekilleştirilmiş bütünleşik görünümü (Deduplicated integrated view of the method- and data-based literature)

Yazar	Kullanılan Veri	Değişkenler	Yöntem	Sonuç
Zadrozny ve Elkan (2002)	UCI ML Repository ve KDD-98 karşılaştırma veri kümeleri	Çoklu sınıflandırma görevlerinin sayısal ve kategorik değişkenleri	Naive Bayes ve karar ağacı çıktıları için sigmoid (Platt) ile isotonic regresyon kalibrasyonu karşılaştırması	Isotonic regresyon parça parça sabit ve esnek bir kalibrasyon sağlamış; veri yeterli olduğunda sigmoid yaklaşımına göre daha iyi sonuç vermiştir.
Batista vd. (2004)	UCI ML Repository; 13 dengesiz sınıflandırma veri kümesi	Çoklu sınıflandırma görevlerinin sayısal ve kategorik değişkenleri	On dengeleme yönteminin (rastgele örnekleme, Tomek bağlantıları, ENN, SMOTE, SMOTE-Tomek, SMOTE-ENN vb.) C4.5 ile sistematik karşılaştırması	Sentetik örnekleme ve örnek temizlemenin birleştirildiği hibrit yöntemler (SMOTE-Tomek, SMOTE-ENN) tek başına kullanılan yöntemlerden daha tutarlı sonuçlar üretmiştir.
Niculescu-Mizil ve Caruana (2005)	UCI ML Repository ve diğer kaynaklardan derlenen 8 karşılaştırma veri kümesi	Çoklu sınıflandırma görevlerinin sayısal ve kategorik değişkenleri	On sınıflandırma algoritması (boosted trees, RF, SVM, NN, NB, LR vb.) için Platt sigmoid ve isotonic kalibrasyon karşılaştırması	Boosted trees ve SVM doğal olarak iyi kalibre değildir; Platt ile isotonic kalibrasyon olasılık tahminlerini iyileştirmiştir; isotonic kalibrasyon yeterli veri olduğunda daha esnektir.
Bahnsen vd. (2016)	Özel Avrupa işlem verisi; dönem raporlanmamış; işlem düzeyi	Ham nitelikler, toplulaştırılmış ve periyodik değişkenler	Öznitelik mühendisliği, maliyet duyarlı değerlendirme	Periyodik ve toplulaştırılmış özellikler maliyet temelli performansı iyileştirmiştir.
Dal Pozzolo vd. (2018)	Gerçek kart işlemleri; kronolojik akış	Zaman akışı, işlem sıraları, dengesiz sınıf	Gerçekçi zamansal değerlendirme	Gerçekçi zaman bölmesi altında performans farklılaşmıştır.
Jurgovsky vd. (2018)	Gerçek işlem dizileri; sekans düzeyi	İşlem geçmişi, kart bazlı sıralı kayıtlar	Sekans sınıflandırması, LSTM	Sekans öğreniciler statik modellere göre güçlü ayrıştırma sağlamıştır.
Randhawa vd. (2018)	Açık veri + kurumsal veri	İşlem değişkenleri, gürültülü senaryolar	AdaBoost, çoğunluk oylaması, hibrit topluluk	Hibrit topluluk modeli dayanıklılık açısından üstün bulunmuştur.
Alarfaj vd. (2022)	Kaggle açık verisi; işlem düzeyi	V1-V28, Time, Amount	LR, RF, KNN, XGBoost, DNN, LSTM	XGBoost ve RF yüksek AUC üretmiş; DNN karşılaştırılabilir sonuç vermiştir.
Ghosh Dastidar vd. (2022)	Milyonlarca gerçek işlem	İşlem geçmişi özetleri, öğrenilebilir toplulaştırmalar	Neural Aggregate Generator, LSTM, CNN	Öğrenilebilir özellikler manuel agregatlara tercih edilmiştir.
Ileberi vd. (2022)	Eylül 2013; 2 gün; açık veri	V1-V28, Time, Amount, Class	GA öznitelik seçimi; DT, RF, LR, ANN, NB	GA tabanlı seçim sınıflandırma başarımını artırmıştır.
Afriyie vd. (2023)	2020; Batı ABD; simüle veri	İşlem kayıtları, yaş, saat	LR, RF, karar ağacı	RF en yüksek AUC ve doğruluğa ulaşmıştır.
Cherif vd. (2023)	Sistematik derleme; 2015-2022	Çoklu veri kaynakları	ML, DL, hibrit yöntemlerin karşılaştırması	Topluluk ve derin öğrenme yaklaşımlarının artan üstünlüğü vurgulanmıştır.
Fanai ve Abbasimehr (2023)	Açık veri (Eylül 2013) + German Credit	Anonim değişkenler, dönüştürülmüş temsiller	Derin otoenkoder, DNN, RNN, CNN-RNN	Temsil öğrenimi sınıflandırmayı güçlendirmiştir.

Abdul Salam vd. (2024)	Eylül 2013; 2 gün; açık veri	İşlem değişkenleri, dengeleme türevleri	Federated learning, veri dengeleme, CNN	Dengeleme teknikleri federated çerçevede performansı güçlendirmiştir.
Altan ve Zafer (2024)	Ocak 2023; kamu bankası verisi	13.050 gözlem, banka nitelikleri	RF, LR, KNN, DT, gradyan artırma	Gradyan artırma en başarılı model olmuştur.
C-Rella vd. (2024)	Gerçek işlem verisi; işlem düzeyi	İşlem tutarı, olasılık skoru	İki boyutlu maliyet duyarlı eşik seçimi	Maliyet duyarlı eşik seçimi tek boyutlu yaklaşıma göre daha düşük beklenen maliyet üretmiştir.
Chhabra vd. (2024)	Açık veri; 284.807 işlem	V1-V28, Time, Amount	Oylama tabanlı topluluk modeli (RF, XGBoost, LR, SVM)	Oylama topluluk modeli tekil modellere göre F1 ve AUC artışı sağlamıştır.
Motie ve Raahemi (2024)	Sistematik derleme; GNN çalışmaları	Grafik yapıları, ilişkisel veriler	GNN, GAT, GCN tabanlı modeller	Grafik sinir ağları ilişkisel dolandırıcılık örüntülerini yakalamada üstünlük göstermiştir.
Sinap (2024)	Eylül 2013; 2 gün; açık veri	Sayısal işlem alanları	Ölçekleme, örnekleme, çoklu model	RF ve KNN görece güçlü sonuç üretmiştir.
Tang ve Liang (2024)	Kurumsal silolar + açık veri	İşlem ilişkileri, grafik temsilleri	Federated learning + GNN	Grafik tabanlı federated yapı ayrıştırma kalitesini artırmıştır.
Xie vd. (2024)	Büyük ölçekli gerçek + açık veri	İşlem sıraları, davranışsal temsiller	Zaman-duyarlı davranış temsili	Davranış temsili güçlü ayrıştırma başarımı üretmiştir.
Ahmed vd. (2025)	Kaggle açık veri; işlem düzeyi	Açık veri işlem alanları	SMOTE-ENN ile topluluk sınıflayıcı	Hibrit örnekleme ile topluluk modeli temel modellere göre daha iyi performans vermiştir.
Albalawi ve Dardouri (2025)	Kaggle + PaySim; işlem düzeyi	V1-V28, Time, Amount, PaySim alanları	SMOTE, LR, DT, RF, DL, focal loss	RF genel performansta, DL precision bakımından öne çıkmıştır.
Baisholan vd. (2025)	Açık veri; 284.807 işlem; 2 gün	V1-V28, Time, Amount, Class	RF + XGBoost topluluk modeli, eşik ayarı, SHAP, LIME	AUC-PR ve duyarlılık birlikte yükseltilirken açıklanabilirlik korunmuştur.
Btoush vd. (2025)	Açık veri; işlem düzeyi	V1-V28, Time, Amount	ML+DL hibrit topluluk modeli	Hibrit topluluk modeli tek başına ML veya DL'den daha iyi F1 üretmiştir.
Tayebi ve El Kafhali (2025)	Açık veri; işlem düzeyi	V1-V28, Time, Amount	XGBoost + Bayesyen hiperparametre optimizasyonu	Bayesyen optimizasyon ile XGBoost ayrıştırma kalitesini artırmıştır.
Tekin ve Yaman (2023)	2023; akıllı ev platformu; 435.815 örnek	Ağ trafiği; 7 saldırı + normal	XGBoost; diğer ML algoritmalarıyla karşılaştırma	%92,55 doğruluk (80:20); XGBoost başarılı
Güney ve Selvi (2023)	2006-2022; 8.224 abone; dönemsel tüketim	Sayaç, abone, tüketim, fatura, ödeme	Anomali analizi + DT, NB, KNN, LR, MLP, RF, GB	Anormal tüketim örüntüleri tespit edilebilmiştir.
Yıltırak ve Öztürk (2025)	CICIoT2023; 33 saldırı türü, 7 grup	İnternet trafiği özellikleri ve saldırı etiketleri	AdaBoost, CatBoost, XGBoost, RF, LightGBM, LR vb.	RF en iyi sonucu üretmiştir; ağaç tabanlı modeller güçlüdür.