No: 01 2026
https://doi.org/10.64400/ea26-01-17
ISSN 2224-5529 2224-0829 2223-5124 2222-940X

NAXÇIVAN DÖVLƏT UNİVERSİTETİNİN ELMİ ƏSƏRLƏR JURNALI

MALİYYƏ FIRILDAQÇILIĞININ QARŞISININ ALINMASINDA EKSTREMAL QRADİYENT GÜCLƏNDİRMƏNİN (XGBOOST) ROLU

Zöhr BABAZADƏ
Elm və Təhsil Nazirliyi yanında İqtisadiyyat İnstitutu, Bakı, Azərbaycan
**Yazışılan müəllif: babazadezohr@gmail.com*

NƏŞR TARİXİ:
Qəbul edilmə tarixi: 03.03.2026
Nəşr edilmə tarixi: 17.03.2026

AÇAR SÖZLƏR:
XgBoost, fırıldaqçılıqla mübarizə, bank sistemi, maşın öyrənməsi, F1-score

XÜLASƏ
Maliyyə fırıldaqçılığının vaxtında və dəqiq şəkildə aşkarlanması banklar və maliyyə institutları üçün strateji əhəmiyyət daşıyır. Məqalədə maliyyə fırıldaqçılığının aşkarlanmasında Ekstremal Qradiyent Gücləndirmə (Extreme Gradient Boosting – XGBoost) modelinin rolu, üstünlükləri və tətbiq zamanı qarşılaşılan əsas çətinliklərin təhlili edilir. Xüsusilə, fırıldaqçılıq məlumatlarının ciddi sinif balanssızlığı şəraitində XGBoost modelinin yüksək əhatəlilik (recall) əldə etməsinə baxmayaraq, aşağı dəqiqlik (precision) səbəbindən çoxsaylı yalnış müsbət hallar yaratması əsas problem kimi ön plana çıxarılmışdır. Bu kimi məhdudiyyətlərin aradan qaldırılması üçün balanslaşdırma texnikalarının və hibrid yanaşmaların rolu empirik nəticələr əsasında qiymətləndirilmişdir. K-SMOTEENN kimi hibrid resampling metodunun tətbiqinin XGBoost modelinin dəqiqlik–əhatəlilik balansını əhəmiyyətli dərəcədə yaxşılaşdırdığı və beləliklə F1-score kimi əsas göstəricilərdə daha stabil nəticələr verdiyi araşdırmada qeyd edilmişdir. Qraf neyron şəbəkələri vasitəsilə gücləndirilmiş XgBoost modelinin səmərəliliyini artıraraq daha yüksək nəticələr əldə etməsi aşkarlanmışdır. Bundan əlavə, Random Forest və stacking ensemble modelləri ilə aparılan müqayisələr XGBoost-un təkbaşına deyil, daha geniş ansambl və hibrid arxitekturalar daxilində daha səmərəli fəaliyyət göstərdiyini sübut edir.

GİRİŞ

Maliyyə sisteminin qlobal miqyasda rəqəmsallaşdırılması istehlakçılar üçün böyük rahatlıq yaratmaqla yanaşı, həmçinin maliyyə cinayətləri üçün də imkanları əhəmiyyətli dərəcədə genişləndirmişdir. Kredit kartı fırıldaqçılığı üzrə itkilərin dinamikası hər il artan xarakter daşıyır. Statistikalar göstərir ki, qlobal itkilər 2011-ci ildə təxminən 9,84 milyard ABŞ dollarından 2021-ci ildə 32,34 milyard ABŞ dollarına qədər artaraq təxminən üç dəfə yüksəlmişdir (Mienye & Jere, 2024). Bu artım yalnız əməliyyat həcminin genişlənməsi ilə deyil, eyni zamanda cinayətkar şəbəkələrin fəaliyyətində təkmilləşdirilmiş metodların istifadəsini də özündə əks etdirir.

Maliyyə institutları uzun müddət ənənəvi sistemlərə, yəni fırıldaqçılıq tipologiyalarından irəli gələn if–then məntiqinə əsaslanan qayda əsaslı yanaşmalara istinad etmişdir. Bu sistemlər şəffaf olmalarına baxmayaraq, struktur etibarilə dəyişiklərə reaksiya problemi ilə üzləşirlər. Yeni və əvvəllər müşahidə olunmamış fırıldaqçılıq nümunələrini aşkar etmək qabiliyyətinə malik olmamaqla, yüksək yalnış müsbət nəticələrə gətirib çıxarmaqla, aşağı dəqiqliyə səbəb olur (Manchev, 2021). Müasir dövrdə kart əməliyyatlarının həcminin çox böyük olması, iri banklar üçün saatda milyonlarla əməliyyata çata bilməsi qaydaların əl ilə saxlanılmasını həm yoxlama və nəzarət baxımından, həm də insan resursları baxımından praktiki olaraq qeyri-mümkün edir.

Bu boşluqların aradan qaldırılması məqsədilə bank sektoru məlumat əsaslı Maşın Öyrənməsi (Machine Learning – ML) metodlarına keçid etmiş və ya araşdırmalar tərəfindən tövsiyyə edilmişdir. Qaydaya əsaslanan sistemlərdən fərqli olaraq, ML alqoritmləri keçmiş məlumatlar üzərindən müvafiq qərar sərhədlərini avtomatik şəkildə öyrənir ki, bu da fırıldaqçılıq davranışına xas olan əlaqələrin aşkar edilməsinə imkan yaradır. Bu sahədə inkişaf tək xətti modellərdən (məsələn, Logistik Reqressiya, Qərar Ağacları) daha mürəkkəb və yüksək performanslı ansambl metodlarına doğru mərhələli şəkildə yönəlmişdir (Damanik & Liu, 2025).

Bu yanaşmalar arasında Ekstremal Qradiyent Gücləndirmə (Extreme Gradient Boosting – XGBoost) əsas maşın öyrənməsi metodlarından biri hesab olunur. Qradiyentlə gücləndirilmiş qərar ağaclarının (Gradient Boosted Decision Trees – GBDT) miqyaslana bilən reallaşdırılması kimi

təqdim edilən XGBoost, icra sürəti, modelin proqnozlaşdırma qabiliyyəti mexanizmlərinin optimal kombinasiyası sayəsində sənaye üzrə göstəricilərdə və data elmi müsabiqələrində (məsələn, Kaggle IEEE-CIS Fraud Detection) davamlı şəkildə istifadə olunan əsas metod olmuşdur (Zhang et al., 2020). Bununla belə, bank sektorunda XGBoost modelinin tətbiqi sadə proses hesab edilmir. Bu model bir sıra ciddi çətinliklərlə üzləşir ki, onların əsasını balanssız məlumat bazası təşkil edir. Belə ki, fırıldaqçılıq əməliyyatları məlumat bazasında son dərəcə kiçik paya malikdir (<~0,2%) (Khadka, 2025). Eyni zamanda, requlyativ çərçivələr üzrə də modelin izah oluna bilməsinə dair tələblər də XGBoost-un praktik tətbiqini məhdudlaşdıran amillərdən biri hesab edilir (Baisholan et al., 2025).

XGBOOSTİNG VƏ İKİNCİ DƏRƏCƏLİ DÜZÜLÜŞ OPTİMİZASİYASI

Gücləndirmə (boosting) yanaşmasının əsas prinsipi modeldə zəif öyrənən qərar ağaclarının ardıcıl şəkildə qurulmasına əsaslanır. Bu prosesdə hər bir növbəti ağac əvvəlki ansamblın qalıq səhvlərini düzəltməyə yönəlir. Ənənəvi Qradiyent Gücləndirmə Modellərində (GBM) bu mexanizm birinci dərəcəli törəmələrdən (qradiyentlərdən) istifadə edilməklə idarə olunduğu halda, XGBoost itki funksiyasının ikinci dərəcəli Teylor açılımından istifadə edir.

$$\mathcal{L}(t) = \sum_{i=1}^{n} l\left(y_i, \widehat{y_i^{(t-1)}} + f_t(x_i)\right) + \Omega(f_t)$$

Burada,

$\mathcal{L}(t)$ nömrəli iterasiyada modelin itki funksiyasıdır. Bu funksiya modelin cari mərhələdə proqnoz keyfiyyətini ölçür və optimallaşdırma prosesinin əsas hədəfini müəyyən edir.

$\sum_{i=1}^{n} x$ İtkinin bütün müşahidələr üzrə (n) cəmlənməsini ifadə edir. Bu modelin performansını yalnız tək müşahidə üzrə deyil, bütün məlumat dəsti üzrə qiymətləndirməyə imkan verir.

$l(.)$ İtki funksiyasıdır. Modelin verdiyi proqnoz ilə real müşahidə arasındakı fərqi ölçür.

y_i $-$i nömrəli müşahidə üçün real və ya nəticə dəyişədir.

$\widehat{y_i^{(t-1)}}$ Modelin əvvəlki iterasiyaya qədər yaratdığı kumulyativ proqnozdur.

$f_t(x_i)$ - t nömrəli iterasiyada əlavə edilən zəif öyrənən modelin (qərar ağacı) xi müşahidəsi üzrə verdiyi proqnozdur.

$\Omega(f_t)$ – requlyarizasiya terminidir. Modelin overfitting (həddindən artıq uyğunlaşma) olmasının qarşısını alır.

Fırıldaqçılığın aşkarlanması kontekstində, anomaliya (məsələn, səyahət zamanı xaricdə alış veriş edilməsi ilə) ilə real fırıldaq əməliyyatı arasındakı qərar sərhədi son dərəcə incə olduğundan, bu cür dəqiqlik Həqiqi Müsbətlərin Aşkarlanma Dərəcəsinin (True Positive Rate – TPR) maksimuma çatdırılması və Yanlış Müsbətlərin (False Positives) minimallaşdırılması baxımından həlledici əhəmiyyətə malikdir (Kabane, 2024).

Fırıldaqçılıq kimi balanssız məlumatların modelləşdirilməsində ən mühüm risklərdən biri həddindən artıq uyğunlaşmadır (overfitting). Fırıldaq əməliyyatlarının müşahidə sayı çox az olduğundan, requlyarizasiyasız bir model fırıldaqçılığın ümumi və xarakterik xüsusiyyətlərini öyrənmək əvəzinə, məlumat dəstində mövcud olan azsaylı fırıldaq nümunələrinin spesifik xüsusiyyətlərini "əzbərləyə" bilər.

XGBoost alqoritmi bu baxımdan klassik Gradient Boosted Decision Trees (GBDT) və Random Forest modellərindən əsaslı şəkildə fərqlənir. Belə ki, XGBoost-da requlyarizasiya mexanizmi birbaşa məqsəd (itki) funksiyasına inteqrasiya edilmişdir, bu da modelin mürəkkəbliyinə sistemli nəzarət etməyə və oxşar lakin eyni olmayan hadisələrin aşkarlanmasında daha ümumiləşdirilə bilən nəticələr əldə etməyə imkan verir (Kabane, 2024).

XGBOOST-UN DİGƏR MODELLƏRLƏ MÜQAYİSƏSİ

XGBoost ilə Random Forest arasındakı fərqlər bu sahədə ən geniş sənədləşdirilmiş müqayisələrdən biri hesab olunur. Hər iki yanaşma ağac (tree) əsaslı ansambl modelləri olsa da, onların qurulma mexanizmləri gücləndirmə (boosting) və torbalama (bagging) məlumatların xüsusiyyətlərindən asılı olaraq fərqlənir.

Das və həmkarları tərəfindən aparılmış müqayisəli tədqiqatda kredit kartı əməliyyatlarından ibarət məlumat dəsti əsasında XGBoost modeli maksimum 99,96% dəqiqlik səviyyəsinə nail olmuş və Random Forest modelinin 99,95%-lik nəticəsini cüzi fərqlə üstələmişdir (Das et al., 2023). Bu fərq ilk baxışda əhəmiyyətsiz görünsə də, gündə milyonlarla əməliyyatın icra edildiyi bank mühitində 0,01%-lik fərq yüzlərlə səhv təsnif edilmiş əməliyyata ekvivalentdir.

Məlumatların ciddi şəkildə balanssız olduğu fırıldaqçılığın aşkarlanması sahəsində dəqiqlik (accuracy) göstəricisi aldadıcı xarakter daşıyır. Belə ki, bütün əməliyyatları "qanuni" kimi proqnozlaşdıran model belə 99,8% dəqiqlik səviyyəsinə çata bilər. Bu kontekstdə həlledici qiymətləndirmə meyarları Dəqiqlik (Precision) yəni şübhəli kimi işarələnmiş əməliyyatların neçə faizinin həqiqətən fırıldaqçılıq olduğu və Əhatəlilik (Recall) yəni faktiki fırıldaqçılıq hallarının neçə faizinin model tərəfindən aşkarlanması hesab olunur.

Tədqiqat nəticələri göstərir ki, XGBoost modeli adətən Əhatəlilik (Recall) göstəricisi üzrə üstün performans nümayiş etdirir. IEEE Access jurnalında dərc olunmuş bir araşdırmada XGBoost-un 0,98 səviyyəsində recall dəyərinə nail olduğu və bu göstəricinin sınaqdan keçirilmiş digər klassifikatorlar arasında ən yüksək nəticə olduğu qeyd edilmişdir. Lakin bu üstünlük Dəqiqlik (Precision) hesabına əldə edilmişdir. Dəqiqlik göstəricisi 0,47 olmuş və nəticədə F1-göstəricisi (0,64) həmin konfiqurasiyada balanslaşdırılmış Random Forest modeli ilə müqayisədə daha aşağı səviyyədə qeydə alınmışdır (Damanik & Liu, 2025). Bu fakt XGBoost-un fırıldaqçılığın aşkarlanmasında yüksək aqressivliyə malik olduğunu, yəni çoxsaylı qanuni əməliyyatları da şübhəli kimi işarələyə bildiyini (yalnış müsbət nəticələr – False Positives) göstərir. Banklar üçün bu vəziyyət kompromis yaradır: yüksək əhatəlilik (recall) birbaşa fırıldaqçılıq itkilərini minimuma endirsə də, aşağı dəqiqlik (precision) əməliyyat xərclərinin (əl ilə yoxlamaların artması) və müştəri narazılığının (kart əməliyyatlarının əsassız rədd edilməsi) yüksəlməsinə səbəb ola bilər.

Fırıldaqçılığın aşkarlanmasında əsas fundamental problem kəskin balanssızlığıdır. Geniş şəkildə etalon (benchmark) kimi istifadə olunan Avropa Kredit Kartı məlumat bazasında 284.807 əməliyyatdan cəmi 492-si fırıldaqçılıq xarakterlidir ki, bu da ümumi əməliyyatların yalnız 0,172%-ni təşkil edir. Belə bir məlumat dəsti üzərində təlim keçirilmiş standart XGBoost modeli bütün əməliyyatları "qanuni" kimi proqnozlaşdırmaqla 99,8% dəqiqlik səviyyəsinə çata bilər ki, bu da praktik baxımdan tamamilə faydasız bir nəticə deməkdir (Khadka, 2025). Elmi ədəbiyyatda XGBoost-un bu dərəcədə qeyri-bərabər paylanmış mühitdə səmərəli fəaliyyət göstərə bilməsi üçün bir sıra kritik müdaxilələrin zəruri olduğu vurğulanır.

Modelin işləmə məntiqində balanssızlığın nəzərə alınması məqsədilə scaling_pos_weight adlı daxili hiperparametr mövcuddur. Bu parametr geriyə müsbət sinfin, yəni fırıldaqçılıq əməliyyatlarının qradiyentlərini miqyaslandırmaqla, modelin nadir sinfə qarşı həssaslığını artırmağa xidmət edir (XGBoost Developers, 2019).

$$scale\_pos\_weight = \frac{Say(Qanuni\ əməliyyatlar)}{Say(Fırıldaqçılıq\ əməliyyatları)}$$

Bu parametr əhatəliliyi artırmaqla, fırıldaqçılıq hallarının modeldəki dəyərini artırır. Beləliklə, model müxtəlif fırıldaqçılıq hallarını aşkarlamaq qabiliyyətinə malik olsa belə, bu aşağı dəqiqlik (precision) hesabına baş verir. Daha çox yanlış müsbətlər ortaya çıxır. Lakin, əhatəlilik və dəqiqlik balanssız məlumat bazasında göstərici kimi kifayət etmir. Bunun üçün F1 göstəricisi əsas hesab olunur. F1 göstərici (F1 score) yalnız hər iki göstərici (əhatəlilik və dəqiqlik) yüksək olduqda yüksək nəticə göstərir.

$$F1 = 2 * \frac{Dəqiqlik\ (precision) * Əhatəlilik\ (recall)}{Dəqiqlik\ (precision) + Əhatəlilik\ (recall)}$$

Yuxarıda qeyd olunan düstura əsasən, əgər bir göstərici aşağı olarsa, F1 göstəricisi əsaslı dərəcədə aşağı nəticə göstərir. Bu isə modelin səmərəli işləyib işləmədiyini bilmək üçün əsas göstəricilərdən biri hesab olunur.

XGBOOST VƏ QRAF NEYRON ŞƏBƏKƏLƏRİ

NVIDIA tərəfindən paylaşılan bir məqalədə qeyd olunur ki, XGBoost kimi ənənəvi maşın öyrənməsi modelləri əməliyyat səviyyəsində fırıldaqçılığın aşkarlanmasında yüksək səmərəliliyə

malik olsa da, əlaqəli və şəbəkə xarakterli fırıldaqçılıq sxemlərini tam şəkildə əhatə edə bilmir. Maliyyə fırıldaqçılığı adətən tək-tək əməliyyatlardan ibarət olmur. Bu adətən bir neçə hesab və şəxslər arasında koordinasiyalı fəaliyyət formasında baş verir. Bu boşluğu doldurmaq üçün məqalə Qraf Neyron Şəbəkələrini (Graph Neural Networks – GNN) təqdim edir. GNN-lər tranzaksiyaları əlaqəli qraf strukturu kimi modelləşdirərək ənənəvi modellərin görə bilmədiyi gizli əlaqələri üzə çıxarır və bu, mürəkkəb fırıldaqçılıq sxemlərinin aşkarlanmasını əhəmiyyətli dərəcədə gücləndirir (Naim et al., 2025).

GNN-lər qraf əsaslı vektor təsvirlər yaradır və bunlar XGBoost modelinə giriş dəyişənləri kimi ötürülür. Beləliklə, GNN-lərin əlaqə əsaslı analitik gücü XGBoost-un sürətli, izaholunan və miqyaslana bilən təsnifat imkanları ilə birləşdirilir. Nəticədə daha yüksək əhatəlilik (recall), daha az yalnış pozitiv və real vaxt rejimində daha səmərəli fırıldaqçılıq nəzarəti əldə olunur. Məqalə göstərir ki, XGBoost bu sistemlərdə əvəz edilmir, əksinə qraf əsaslı dərin öyrənmə ilə gücləndirilmiş əsas model rolunu oynayır (Naim et al., 2025).

Şəkil 1: Qraf əsaslı və GNN-lə gücləndirilmiş fırıldaqçılıq aşkarlama sistemlərində istifadə olunan maliyyə subyektləri və onların əlaqələrinin konseptual təsviri
Mənbə: https://www.slideteam.net/financial-fraud-detection-graph-analytics-model.html - Şəkil Nano Banana modeli vasitəsilə tərcümə edilərək generasiya edilmişdir.

Şəkil 1-də düyünlər bank hesabları, kredit kartları, telefon nömrələri, sosial sığorta nömrələri (SSN), ünvanlar və sintetik şəxsiyyətlər kimi maliyyə subyektlərini, kənarlar isə bu subyektlər arasında mövcud olan əlaqələri (paylaşılan məlumatlar və ya maliyyə münasibətləri) təmsil edir. Vizual struktur bir neçə hesabın və şəxsiyyətin dolayı yollarla eyni atributlar üzərindən əlaqələndiyini göstərərək sintetik şəxsiyyət fırıldaqçılığı və koordinasiyalı təminatsız kredit sui-istifadəsi kimi mürəkkəb risk nümunələrini üzə çıxarır.

Bu şəkil birbaşa neyron şəbəkəni təsvir etməsə də, Qraf Neyron Şəbəkələrinin (GNN) işləmə məntiqi ilə uyğun gəlir. GNN yanaşmasında hər bir düyün qonşu düyünlərdən məlumat toplayaraq əlaqə əsaslı risk siqnallarını öyrənir. Məsələn, eyni telefon nömrəsi və ya ünvanla əlaqəli bir neçə hesabın mövcudluğu yüksək fırıldaqçılıq riski göstəricisi kimi modelləşdirilə bilər. Nəticədə əldə olunan düyünlərin vektor təsvirləri (embeddings) sonrakı emal mərhələsində XGBoost kimi təsnifat modellərinə ötürülərək fırıldaqçılığın daha dəqiq aşkarlanmasına imkan yaradır.

XGBOOST VƏ K-SMOTEENN

K-SMOTEENN sinif balanssızlığı problemini həll etmək üçün K-means klasterləşdirməni, SMOTE əsaslı sintetik nümunə yaradılmasını və Edited Nearest Neighbors (ENN) ilə modeldə səs-küyün təmizlənməsini birləşdirən hibrid resampling metodudur. Bu yanaşma azlıq sinfinin (fırıldaq əməliyyatlarının) təmsilolunmasını artırmaqla yanaşı, yanlış və qeyri-reprezentativ nümunələri aradan qaldıraraq modelin ümumiləşmə qabiliyyətini gücləndirir (Damanik & Liu, 2025). Son tədqiqatlar göstərir ki, XGBoost fırıldaqçılığın aşkarlanmasında yüksək əhatəlilik (recall) əldə etsə də, ciddi sinif balanssızlığı şəraitində dəqiqlik (precision) göstəricisi zəifləyir. Damanik və Liu (2025) bu məhdudiyyəti empirik şəkildə təsdiqləyərək, K-SMOTEENN ilə gücləndirilmiş stacking ensemble modellərinin precision–recall balansını əhəmiyyətli dərəcədə yaxşılaşdırdığını nümayiş etdirir. Bu nəticələr XGBoost-un təkbaşına deyil, daha geniş və hibrid aşkarlama arxitekturaları daxilində daha səmərəli olduğunu göstərir.

Cədvəl 1

Xüsusiyyət konstruksiyası və K-SMOTEENN tətbiq edilmədən əldə olunan performans göstəriciləri

Model	Düzgünlük (Accuracy)	F1 score	Əhatəlilik (Recall)	Dəqiqlik (Precision)	AUPRC	ROC
Random Forest	1	0.79	0.68	0.95	0.84	0.96
XgBoost	1	0.41	0.98	0.26	0.92	1

Cədvəl 2

Xüsusiyyət konstruksiyası və K-SMOTEENN tətbiq edildikdən sonra performans göstəriciləri

Model	Düzgünlük (Accuracy)	F1 score	Əhatəlilik (Recall)	Dəqiqlik (Precision)	AUPRC	ROC
Random Forest	1	0.86	0.79	0.96	0.76	0.97
XgBoost	1	0.64	0.98	0.47	0.94	1

Mənbə: Damanik, N., & Liu, C.-M. (2025). https://doi.org/10.1109/ACCESS.2025.3528079

Əgər Cədvəl 1-ə nəzər yetirsək, əhatəliliyin XgBoost modeli üzrə əhatəliliyin 0.98 olması o deməkdir ki, bütün fırıldaq əməliyyatları aşkar edilir. Lakin dəqiqliyin 0.26 olması aşkar edilən əməliyyatların böyük hissəsinin yanlış müsbətlər olması deməkdir. Bu isə özünü aşağı F1-scoreda (0.41) göstərir. Random Forest isə daha konservativ qərar mexanizmi quraraq yalnış pozitivləri minimuma endirir, lakin bu, müəyyən fırıldaq hallarının aşkar edilməməsi bahasına baş verir.

Cədvəl 2-ə nəzər yetirsək görərik ki, K-SMOTEENN tətbiqi XGBoost-un əsas zəif cəhətini yəni aşağı dəqiqlik problemini qismən aradan qaldırır. Model yüksək əhatəlilik səviyyəsini qoruyaraq daha az yalnış pozitiv yaradır. Bu, balanslaşdırılmış məlumatın XGBoost üçün kritik əhəmiyyət daşıdığını göstərir. Random Forest balanslaşdırmadan sonra həm daha çox fırıldaqçılıq hallarını aşkar edir, həm də yüksək dəqiqlik səviyyəsini saxlayır. Bu, modelin ümumiləşmə qabiliyyətinin gücləndiyini göstərir.

NƏTİCƏ

Aparılan araşdırmalar göstərir ki, XGBoost yüksək proqnozlaşdırma qabiliyyəti, sürətli icra mexanizmi və güclü optimallaşdırma xüsusiyyətləri sayəsində fırıldaqçılıq aşkarlanmasında səmərəli alət olsa da, ciddi sinif balanssızlığı şəraitində təkbaşına istifadə edildikdə əməliyyat baxımından risklər yaradır. Modelin yüksək əhatəlilik (recall) əldə etməsi çox vaxt aşağı dəqiqlik (precision) hesabına baş verir ki, bu da bank mühitində yalnış pozitivlərin artmasına, əməliyyat xərclərinin yüksəlməsinə və müştəri məmnuniyyətinin azalmasına səbəb ola bilər.

Tədqiqat nəticələri balanslaşdırma texnikalarının və məlumat səviyyəli müdaxilələrin XGBoost-un səmərəliliyində həlledici rol oynadığını göstərir. K-SMOTEENN kimi hibrid resampling metodlarının tətbiqi XGBoost-un dəqiqlik-əhatəlilik balansını əhəmiyyətli dərəcədə

yaxşılaşdıraraq modelin ümumiləşmə qabiliyyətini gücləndirir. Cədvəllər üzrə aparılan müqayisəli analizlər sübut edir ki, balanslaşdırılmış məlumat üzərində təlim keçmiş XGBoost modeli daha stabil F1-score göstəriciləri nümayiş etdirir. Bununla yanaşı, Random Forest kimi ansambl modellərinin daha konservativ qərar mexanizmi quraraq yalnış pozitivləri azaltması, lakin müəyyən fırıldaq hallarını qaçırması, model seçiminin bankların risk iştahına və əməliyyat prioritetlərinə uyğun şəkildə aparılmasının vacibliyini bir daha vurğulayır.

ƏDƏBİYYAT

1. Mienye, I. D., & Jere, N. (2024). Deep learning for credit card fraud detection: A review of algorithms, challenges, and solutions. IEEE Access, 12, 96893–96910. https://doi.org/10.1109/ACCESS.2024.3426955
2. Manchev, N. (2021, April 21). Credit card fraud detection using XGBoost, SMOTE, and threshold moving. Domino Data Lab. https://domino.ai/blog/credit-card-fraud-detection-using-xgboost-smote-and-threshold-moving
3. Damanik, N., & Liu, C.-M. (2025). Advanced fraud detection: Leveraging K-SMOTEENN and stacking ensemble to tackle data imbalance and extract insights. IEEE Access, 13, 10356–10370. https://doi.org/10.1109/ACCESS.2025.3528079
4. Zhang, Y., Tong, J., Wang, Z., & Gao, F. (2020). Customer transaction fraud detection using XGBoost model. 2020 International Conference on Computer Engineering and Application (ICCEA), 554–558. https://doi.org/10.1109/ICCEA50009.2020.00122
5. Khadka, A. (2025). Credit card fraud detection via model retraining and fine-tuning [Article, University of Texas at Arlington]. MavMatrix. https://mavmatrix.uta.edu/cse_studentresearch/1
6. Baisholan, N., Dietz, J. E., Gnatyuk, S., Turdalyuly, M., Matson, E. T., & Baisholanova, K. (2025). FraudX AI: An interpretable machine learning framework for credit card fraud detection on imbalanced datasets. Computers, 14(4), Article 120. https://doi.org/10.3390/computers14040120
7. Brenndoerfer, M. (2025, July 18). XGBoost: Complete guide to extreme gradient boosting with mathematical foundations, optimization techniques & Python implementation. Michael Brenndoerfer. https://mbrenndoerfer.com/writing/xgboost-extreme-gradient-boosting-complete-guide-mathematical-foundations-python-implementation
8. Kabane, S. (2024). Impact of sampling techniques and data leakage on XGBoost performance in credit card fraud detection. arXiv. https://doi.org/10.48550/arXiv.2412.07437
9. Das, S. R., Bin Sulaiman, R., & Butt, U. (2023). Comparative analysis of machine learning algorithms for credit card fraud detection. FMDB Transactions on Sustainable Computing Systems, 1(4), 225–244. https://www.fmdbpub.com/uploads/articles/170961178515234.%20FTSCS-106-2023.pdf
10. XGBoost Developers. (2019). XGBoost parameters (Version 0.82) [Computer software documentation]. Read the Docs. https://xgboost.readthedocs.io/en/release_0.82/parameter.html
11. Naim, M., Rees, B., & Liu, S. (2025, June 2). Supercharging fraud detection in financial services with graph neural networks [Blog post]. NVIDIA Technical Blog. https://developer.nvidia.com/blog/supercharging-fraud-detection-in-financial-services-with-graph-neural-networks/

ABSTRACT

THE ROLE OF EXTREME GRADIENT BOOSTING (XGBOOST) IN THE PREVENTION OF FINANCIAL FRAUD

Zohr Babazade

The timely and accurate detection of financial fraud is of strategic importance for banks and financial institutions. This paper analyzes the role, advantages, and main challenges encountered during the implementation of the Extreme Gradient Boosting (XGBoost) model in financial fraud detection. Specifically, the issue of the XGBoost model generating numerous false positives due to low precision—despite achieving high recall under conditions of severe class imbalance in fraud data—is highlighted as a primary concern. To mitigate these limitations, the role of balancing techniques and hybrid approaches has been evaluated based on empirical results. The study notes that the application of a hybrid resampling method, such as K-SMOTEENN, significantly improves the precision-recall balance of the XGBoost model, thereby yielding more stable results in key metrics like the F1-score. It was also found that enhancing the XGBoost model through Graph Neural Networks increases its efficiency and leads to superior results. Furthermore, comparisons with Random Forest and stacking ensemble models demonstrate that XGBoost performs more effectively within broader ensemble and hybrid architectures rather than in isolation.

Keywords: *XGBoost, fraud detection, banking system, machine learning, F1-score*

РЕЗЮМЕ

РОЛЬ ЭКСТРЕМАЛЬНОГО ГРАДИЕНТНОГО БУСТИНГА (XGBOOST) В ПРЕДОТВРАЩЕНИИ ФИНАНСОВОГО МОШЕННИЧЕСТВА

Зохр Бабазаде

Своевременное и точное выявление финансового мошенничества имеет стратегическое значение для банков и финансовых институтов. В статье анализируются роль, преимущества и основные трудности, возникающие при применении модели Extreme Gradient Boosting (XGBoost) для обнаружения финансового мошенничества. В частности, в качестве основной проблемы выделяется то, что в условиях серьезного дисбаланса классов данных о мошенничестве модель XGBoost, несмотря на достижение высокой полноты (recall), генерирует множество ложноположительных срабатываний из-за низкой точности (precision). Для устранения подобных ограничений на основе эмпирических результатов была оценена роль методов балансировки и гибридных подходов. В исследовании отмечается, что применение гибридного метода ресемплинга, такого как K-SMOTEENN, значительно улучшает баланс точности и полноты модели XGBoost, обеспечивая более стабильные результаты по ключевым показателям, таким как F1-score. Было выявлено, что усиление модели XGBoost с помощью графовых нейронных сетей повышает ее эффективность и позволяет достичь более высоких результатов. Кроме того, сравнения с моделями Random Forest и стекингом ансамблей (stacking ensemble) доказывают, что XGBoost работает более эффективно не изолированно, а в составе более широких ансамблевых и гибридных архитектур.

Ключевые слова: *XGBoost, борьба с мошенничеством, банковская система, машинное обучение, F1-score*