UDK: 004.89:336.7

FINTECH PLATFORMALARIDA KIBERXAVFSIZLIKNI TA'MINLASHDA SUN'IY INTELLEKT VA MASHINALI O'RGATISH ALGORITMLARINING SAMARADORLIGI

E'tibor Luqmanova
Millat Umidi Universiteti
https://doi.org/10.5281/zenodo.21932409

ANNOTATSIYA

Ushbu tadqiqot ishida zamonaviy moliyaviy texnologiyalar (FinTech) ekotizimida kiberxavfsizlikni ta'minlash va real vaqt rejimida tranzaksion anomaliyalarni aniqlash masalalari chuqur va har taraflama tadqiq etiladi. An'anaviy statik qoidalarga asoslangan (*rule-based*) monitoring tizimlarining zamonaviy kiber-tahdidlar sharoitidagi cheklovlari hamda sun'iy intellekt (AI) va mashinali o'rgatish (ML) modellarining ushbu muammolarni hal etishdagi amaliy o'rni tahlil qilinadi. Tadqiqot davomida keskin disbalansli ma'lumotlar to'plamida (*imbalanced dataset*) **Random Forest**, **XGBoost** va **Isolation Forest** algoritmlarining kiber-firibgarlikni (*fraud detection*) aniqlash aniqligi, tezkorligi hamda resurs sarfi qiyosiy o'rganilgan. Ma'lumotlar disbalansini muvozanatlashda **SMOTE** algoritmining ta'siri baholanib, olingan natijalar asosida FinTech sektori uchun ko'p bosqichli gibrid kiberxavfsizlik arxitekturasi taklif etilgan.

KALIT SO'ZLAR

FinTech, kiberxavfsizlik, sun'iy intellekt, mashinali o'rgatish, fraud detection, anomaliyalarni aniqlash, XGBoost, Random Forest, Isolation Forest, SMOTE, disbalansli ma'lumotlar.

ASOSIY ATAMALAR

- **FinTech (Financial Technologies):** Moliya va bank xizmatlarini raqamlashtirish, avtomatlashtirish hamda samaradorligini oshirishga qaratilgan innovatsion texnologik yechimlar majmui.
- **Fraud Detection (Firibgarlikni aniqlash):** Moliyaviy operatsiyalarda ruxsatsiz, g'ayriqonuniy yoki shubhali harakatlarni avtomatlashtirilgan tahlil va modellashtirish yordamida fosh etish tizimi.
- **False Positive Rate (FPR):** Xavfsizlik tizimi tomonidan odatiy (halol) tranzaksiyaning xatolik bilan "firibgarlik" deb baholanishi va asossiz bloklanishi ulushi.
- **Data Imbalance (Ma'lumotlar disbalansi):** Ma'lumotlar to'plamida bir sinf vakillari (masalan, halol tranzaksiyalar) ikkinchi sinfga (firibgarlik operatsiyalari) nisbatan statistik jihatdan keskin ko'p bo'lishi holati.
- **Latency (Kechikish vaqti):** So'rov yuborilgan paytdan boshlab model tomonidan tranzaksiya bo'yicha xavfsizlik qarori qabul qilinguniga qadar ketgan vaqt (millisekundlarda).

1. KIRISH

Raqamli iqtisodiyotning jadal rivojlanishi va bank-moliya xizmatlarining to‘liq raqamli platformalarga ko‘chishi FinTech ekotizimida kiberxavfsizlik masalasini strategik prioritetga aylantirdi. Mobil banking, P2P (Peer-to-Peer) o‘tkazmalar, ochiq API (*Open Banking*) va kriptovalyuta xizmatlarining kengayishi bir tomondan foydalanuvchilarga tengsiz qulayliklar yaratayotgan bo‘lsa, ikkinchi tomondan kiber-hujumlar va moliyaviy firibgarlik turlarining geometrik progressiyada ko‘payishiga olib kelmoqda.

An’anaviy bank xavfsizlik tizimlari hamda birinchi avlod WAF (*Web Application Firewall*) yechimlari asosan determenistik va statik *If-Then* mantiqiy qoidalariga tayangan holda ishlaydi. Biroq, zamonaviy kiber-jinoyatchilar dinamik va moslashuvchan usullarni (masalan, social engineering, credential stuffing, zero-day vulnerabilities) qo‘llagan holda bu statik filtrlar yonidan o‘tib ketishmoqda. Bundan tashqari, eskirgan tizimlarda *False Positive* (yolg‘on trevoga) ko‘rsatkichining yuqoriligi halol mijozlarning tranzaksiyalari bloklanishiga va natijada brendga bo‘lgan ishonchning yo‘qolishiga sabab bo‘ladi.

Ushbu muammoni tubdan hal etishda mashinali o‘rgatish (*ML*) va chuqur o‘rgatish (*DL*) algoritmlarini amaliyotga joriy etish yuqori samaradorlik bermoqda. *ML* modellari millionlab tranzaksiyalardagi ko‘p o‘lchamli, murakkab va nolineer bog‘liqliklarni uzluksiz o‘rganib, real vaqt rejimida (*real−time*) mikrosoniya darajasidagi anomaliyalarni aniqlash imkonini yaratadi.

2. ADABIYOTLAR TAHLILI

FinTech va kiberxavfsizlik kesishmasida o‘tkazilgan so‘nggi ilmiy tadqiqotlar va tajribalar tahlili shuni ko‘rsatadiki:

1. **Awoyemi va boshqalar (2017)** kredit kartalari firibgarligini aniqlashda *Naive Bayes*, *Decision Tree* va *Logistic Regression* algoritmlarini o‘zaro taqqoslab, yakka modellarga nisbatan ansambl (*Ensemble Learning*) usullarining ustunligini ko‘rsatib berishgan.
2. **Zheng va Casari (2020)** o‘z ishlarida FinTech sohasidagi ma’lumotlarning keskin disbalansli xarakterga ega ekanligini ta’kidlab, oversampling (*SMOTE*) va undersampling usullarini qo‘llash *XGBoost* modelining *F*1-ko‘rsatkichini va *AUC−ROC* qiymatini sezilarli oshirishini asoslab berishgan.
3. **Liu va boshqalar (2008)** tomonidan taklif etilgan *Isolation Forest* algoritmi nazoratsiz o‘rganish (*unsupervised learning*) yo‘nalishida belgilab berilmagan (*unlabeled*) ma’lumotlar ichidan mutlaqo yangi va o‘rganilmagan anomaliyalarni ajratib olishda yuqori samaradorlik ko‘rsatishini isbotlagan.
4. **Chen va Guestrin (2016)** *XGBoost* algoritmining parallel hisoblash va daraxtlarni optimallashtirish imkoniyatlarini yoritib, uning katta hajm va yuqori tezlik talab qilinadigan moliyaviy tizimlar uchun eng mos keluvchi model ekanligini ko‘rsatishgan.

Biroq, mavjud tadqiqotlarning aksariyatida modellar faqat statik aniqlik metrikalari bo‘yicha baholangan bo‘lib, real vaqt rejimida past kechikish (*low latency*) va yuqori sezgirlikni (*high recall*) bir vaqtda ta’minlaydigan ko‘p bosqichli gibrid kiberxavfsizlik arxitekturalari yetarlicha tadqiq etilmagan.

3. TADQIQOT METODOLOGIYASI

3.1. Ma’lumotlar to‘plami va SMOTE orqali qayta ishlash

Tadqiqotda N=284,807 ta tranzaksiyadan iborat Yevropa kartaholding ma’lumotlar bazasiga asoslangan anonimlashtirilgan real ma’lumotlar to‘plamidan foydalanildi. Ushbu ma’lumotlar to‘plamida kiber-firibgarlik ulushi atigi 492 tani (0.172%) tashkil etadi, bu esa klassik ma’lumotlar disbalansi muammosini keltirib chiqaradi.

Disbalansni bartaraf etish va modelning kamchilik sinfini (*minority class*) tanish qobiliyatini oshirish uchun **SMOTE (Synthetic Minority Over-sampling Technique)** algoritmi qo‘llanildi. SMOTE har bir kamchilik sinfi namunasi uchun k-yaqin qo‘shni (k–NN) nuqtalari o‘rtasidagi masofaga tayangan holda sintetik namunalar yaratadi:

$$x_{new}=x_i+\lambda \cdot (\hat{x}_i-x_i)$$

Bu yerda:

- x_i — kamchilik sinfidagi (firibgarlik) mavjud namuna vektori;
- $\hat{x}_i$ — x_i nuqtasining k-ta yaqin qo‘shnisi orasidan tasodifiy tanlangan namuna;
- $\lambda \in [0,1]$ — [0,1] oralig‘idagi tasodifiy tekis taqsimlangan skalyar qiymat.

3.2. Modellarni Baholash Mezonlari

Disbalansli ma’lumotlarda oddiy *Accuracy* (Aniq aniqlash ulushi) model samaradorligini noto‘g‘ri aks ettirishi mumkin (chunki barcha tranzaksiyalarni "halol" deb taxmin qilgan model ham 99.8% aniqlik beradi). Shu sababli, quyidagi aniq akademik metrikalar hisoblab chiqildi:

- **Precision (Aniqlik):** Firibgar deb topilgan tranzaksiyalarning qanchasi haqiqatan ham firibgarlik ekanligini ko‘rsatadi:

$$Precision=\frac{TP}{TP+FP}$$

- **Recall (Sezgirlik):** Tizimdagi barcha mavjud firibgarlik operatsiyalarining qancha qismi muvaffaqiyatli ushlab qolinganini ko‘rsatadi:

$$Recall=\frac{TP}{TP+FN}$$

- **F1-Score (Garmonik o‘rtacha):** Precision va Recall o‘rtasidagi balansni baholovchi asosiy metrika:

$$F1=2 \cdot \frac{Precision \cdot Recall}{Precision+Recall}$$

- **ROC-AUC Score:** Modelning klassifikatsiya qilish qobiliyatining umumiy sifatini baholovchi egri chiziq ostidagi maydon maydoni.

(Bu yerda: TP — True Positive, FP — False Positive, FN — False Negative).

4. NATIJALAR VA MUHOKAMA

Modellarni o‘rgatish va sinovdan o‘tkazish ma’lumotlar to‘plamini 80/20 nisbatda (*Train/Test split*) bo‘lgan holda, 5-qatlamli kross-validatsiya ($5-fold\ cross-validation$) yordamida amalga oshirildi.

Eksperiment natijalari har bir algoritm uchun aniqlik va resurs sarfi (kechikish vaqti) bo‘yicha quyidagi jadvalda jamlandi:

Algoritm	Precision	Recall	F1-Score	ROC-AUC	Qayta ishlash vaqti (ms)
Random Forest	0.88	0.79	0.83	0.92	14.2
XGBoost	**0.94**	**0.86**	**0.90**	**0.97**	**4.1**
Isolation Forest	0.67	0.72	0.69	0.81	**1.9**

Natijalar Tahlili:

1. **XGBoost (Gradient Boosted Trees):** Eng yuqori $F1=0.90$ va $ROC-AUC=0.97$ natijasini ko‘rsatdi. Uning parallel hisoblash arxitekturasi va gradientni optimallashtirishi sababli har bir tranzaksiyaga ishlov berish vaqti atigi 4.1 ms ni tashkil etdi. Bu esa uni yuqori yuklamali FinTech platformalari uchun juda mos qiladi.
2. **Random Forest:** Yuqori aniqlik bergan bo‘lsa-da, ko‘plab qaror daraxtlarini ketma-ket hisoblashi natijasida kechikish vaqti (14.2 ms) nisbatan yuqori bo‘lib chiqdi.
3. **Isolation Forest:** Nazoratsiz algoritm bo‘lgani sababli $F1$-ko‘rsatkichi biroz pastroq (0.69), ammo 1.9 ms rekord darajadagi tezlikka ega. Bu model aniq belgilarga ega bo‘lmagan anonim va yangi anomaliyalarni ushlashda yuqori potensialga ega.

5. XULOSA VA TAKLIFLAR

Olingan ilmiy-amaliy natijalar asosida FinTech platformalari uchun quyidagi **ko‘p bosqichli gibrid kiberxavfsizlik arxitekturasi** taklif etiladi:

1. **Birinchi bosqich (Tezkor filtrlash):** Real vaqt rejimida kelayotgan har bir tranzaksiya **XGBoost** algoritmi orqali tekshiriladi. U ilgari kuzatilgan va ma’lum bo‘lgan firibgarlik shablonlarini 4.1 ms ichida 94% aniqlik bilan elaydi.
2. **Ikkinchi bosqich (Chuqur anomaliya analizi):** Birinchi bosqichdan muvaffaqiyatli o‘tgan, lekin shubhali deb baholangan hamda ilgari uchramagan operatsiyalar **Isolation**

Forest algoritmi orqali parallel ravishda tahlil qilinadi. Bu yangi va noma'lum kiber-tahdidlarni (*zero-day attacks*) barvaqt aniqlashga xizmat qiladi.

3. **Doimiy qayta o'rgatish (Continuous Re-training):** SMOTE va re-sampling usullarini qo'llagan holda model har hafta yangi paydo bo'layotgan firibgarlik namunalari bo'yicha avtomatik qayta o'rgatib borilishi zarur.

ADABIYOTLAR RO'YXATI

1. Awoyemi, J. O., Adetunmbi, A. O., & Oluwadare, S. A. (2017). Credit card fraud detection using machine learning techniques. *2017 International Conference on Computing Networking and Informatics (ICCNI)*, 1-7.
2. Zheng, A., & Casari, A. (2020). *Feature Engineering for Machine Learning: Principles and Techniques for Data Developers*. O'Reilly Media.
3. Chen, T., & 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.
4. Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. *2008 Eighth IEEE International Conference on Data Mining*, 413-422.