KIBERXAVFSIZLIKDA ISOLATION FOREST YORDAMIDA ANOMALIYALARNI ANIQLASH

G‘ulomov Sherzod Rajaboyevich[1], Normirzayev Farrux Abdurahimovich[1]

[1]Muhammad al-Xorazmiy nomidagi Toshkent axborot texnologiyalari universiteti
E-mail: abdurahimovichfarrux@gmail.com

K E Y W O R D S

kiberxavfsizlik, anomaliya aniqlash, Isolation forest, chuqur o‘rganish, suqilib kirishni aniqlash, sun’iy intellekt

A B S T R A C T

Maqolada kiberxavfsizlik sohasida anomaliyalarni aniqlashning dolzarb muammolari va ularni hal qilishda sun’iy intellekt texnologiyalarining roli ko‘rib chiqiladi. Zamonaviy kiberhujumlarning murakkablashuvi va an’anaviy himoya usullarining cheklanganligi sababli, mashinaviy o‘rganishga asoslangan yondashuvlar, xususan, Isolation forest modellari muhim ahamiyat kasb etmoqda.

Jahon miqyosida kiberhujumlar natijasida yetkazilgan zarar 2024 yilda 9,5 trillion AQSH dollariga yetgan bo‘lib, 2025 yilda 10,5 trillion dollardan oshishi kutilmoqda. O‘zbekistonda esa 2023–2024 yillarda kiberjinoyatlar soni keskin oshib, yetkazilgan zarar milliardlab so‘mni tashkil etgan.

Maqolada Isolation forestning nazariy asoslari, arxitekturasi, rivojlanish tarixi, afzalliklari va kamchiliklari batafsil tahlil qilinadi. Ushbu modelning suqilib kirishlarni aniqlash tizimlari (IDS), zararli dasturlarni aniqlash, ichki tahdidlar va IoT xavfsizligidagi amaliy qo‘llanilishi misollar bilan ko‘rsatiladi. Kelajakda transformer va gibrid modellarning integratsiyasi orqali samaradorlikni oshirish imkoniyatlari muhokama qilinadi.

Kirish

Kiberxavfsizlik sohasida anomaliyani aniqlash texnologiyalari tizim va tarmoq infratuzilmasidagi g‘ayritabiiy faollikni aniqlash orqali kiberhujumlarni oldindan bashorat qilish va ularga qarshi kurashishda muhim rol o‘ynaydi. Zamonaviy hujumlar tobora murakkablashib, an’anaviy qoidaga asoslangan himoya usullarini chetlab o‘tmoqda. Shuning uchun SI va mashinaviy o‘qitish asosidagi anomaliyani aniqlash tizimlari hujumlarni oldindan bashorat qilish va real vaqtda javob berish uchun keng qo‘llanilmoqda.

Jahonda kiberxavfsizlikni ta’minlash vazifasi eng muhim masalalardan biri bo‘lib qolmoqda. So‘ngi yillarda kiberhujumlar bilan bog‘liq holatlar hamda axborot tizimlariga yetkazilgan zarar qiymati proporsional ravishda o‘sib bormoqda.

Xususan, 2024 yilda kiberhujumlar natijasida jahon bo‘ylab ko‘rilgan umumiy zarar 9.5 trln AQSH dollarini tashkil etdi, har bir tizim buzilishi o‘rtacha 4.88 mln AQSH dollariga teng bo‘ldi, buzilishlarining 46% autentifikasiya ma’lumotlarining buzilishi yoki o‘zgartirilishi sabab bo‘lgan. Soha ekspertlari 2025 yilda kiberhujumlar oqibatida yetkaziladigan zarar 10.5 trln AQSH dollaridan oshishi mumkinligini taxmin qilishmoqda.

O‘zbekiston Respublikasida 2023 yilda huquqni muhofaza qiluvchi idoralarga kiberjinoyatlar bilan bog‘liq 35 ming murojaat kelib tushgan va yetkazilgan umumiy zarar 404 mlrd so‘m miqdorida baholangan. Ushbu ko‘rsatkichlar 2024 yilda 58 ming murojaatni va 603 mlrd so‘m zararni tashkil qilgan [6],[7].

Ta’kidlash joizki, O‘zbekiston Respublikasi Prezidentining 2024 yil 14 oktabrdagi PQ-358-sonli qaror bilan tasdiqlangan “Sun’iy intellekt texnologiyalarini 2030-yilga qadar rivojlantirish strategiyasi”da firibgarlik holatlarini oldini olish, to‘lov tizimlari va xavflarini boshqarish yo‘nalishiga sun’iy intellekt texnologiyalarini joriy qilish o‘z aksini topgan.

Kiberxavfsizlikda anomaliyani aniqlash tizimlari an’anaviy signatura asosidagi

© Sh.R. G‘ulomov, F.A. Normirzayev

yondashuvlarga qaraganda ancha samarali. Sun’iy intellekt va mashinaviy o‘rganish modellari, ayniqsa Isolation forest zero-day hujumlar, ichki tahdidlar va murakkab kiberhujumlarni aniqlash imkoniyatini oshiradi. Quyida ta’kidlangan model haqida so’z boradi.

1. Isolation Forest yordamida anomaliyalarni aniqlash

Isolation Forest – bu anomaliyalarni aniqlash uchun mo‘ljallangan, nazoratsiz (*ing. tilida - unsupervised*) o‘rganish algoritmi bo‘lib, ma’lumotlardagi noodatiy nuqtalarni ajratish tamoyiliga asoslanadi. Ushbu algoritm anomaliyalarni izolyatsiya qilish orqali ularni oddiy ma’lumotlardan ajratish jarayonini tezlashtiradi. An’anaviy anomaliya aniqlash usullari ko‘pincha ma’lumotlarning zichligi yoki masofaviy o‘zaro bog‘liqligiga tayanadi, Isolation Forest esa anomaliyalarni tabiiy ravishda izolyatsiya qilish tamoyili asosida aniqlaydi.

Bu usulda ma’lumotlar tasodifiy tanlangan xususiyatlar bo‘yicha, ularning yuqori-quyi oralig‘idagi tasodifiy bo‘linish qiymatlari asosida izolyatsiya qilinadi. Anomaliyalarning asosiy xususiyati shundaki, ular umumiy populyatsiyaga nisbatan kam uchraydi va boshqa nuqtalarga nisbatan ajralib turadi. Shu sababli, ularni izolyatsiya qilish uchun kamroq bo‘linish talab qilinadi, natijada, daraxt tuzilishida qisqa yo‘llar hosil bo‘ladi. Bu algoritmning samarali ishlashiga asosiy omil bo‘lib xizmat qiladi.

Isolation Forest katta hajmdagi ma’lumotlar to‘plami bilan yaxshi ishlaydi va hisoblash jihatdan samaradorligi yuqori bo‘lib, ayniqsa yuqori o‘lchamli ma’lumotlarda anomaliyalarni aniqlash uchun juda foydali hisoblanadi. Bu algoritmda asosiy tushuncha shundaki, normal ma’lumotlar ko‘p bo‘linishdan keyingina izolyatsiya qilinadi, anomaliyalar esa tezroq ajralib chiqadi. Shu bois, anomaliyalarni aniqlashda Isolation Forest juda tez va yengil usul hisoblanadi [1].

2. An’anaviy usullarga nisbatan afzalliklar

Isolation Forest algoritmi an’anaviy anomaliya aniqlash usullariga — masalan, zichlikka asoslangan (*ing. tilida - density-based*) yoki masofaga asoslangan (*ing. tilida - distance-based*) metodlarga — nisbatan sezilarli ustunliklarga ega bo‘lib, ayniqsa katta hajmli va murakkab ma’lumotlar bilan ishlaydigan zamonaviy ilovalarda afzal ko‘rilmoqda.

Birinchidan, algoritm hisoblash samaradorligi jihatidan ajralib turadi. U chiziqli vaqt murakkabligiga (O(n)) ega bo‘lib, faqat tasodifiy daraxtlar tuzadi va har bir namunani shu daraxtlar orqali sinovdan o‘tkazadi. An’anaviy usullar esa ko‘pincha kvadratik yoki undan yuqori murakkablikka ega bo‘lgani uchun katta ma’lumotlar to‘plamida sekin ishlaydi. Natijada, Isolation Forest real vaqtda monitoring talab qiladigan tizimlarda — masalan, kiberxavfsizlik platformalarida — resurslarni tejab, tezkor javob berish imkonini beradi [1].

Ikkinchidan, yuqori o‘lchovli (*ing. tilida - high-dimensional*) ma’lumotlar bilan ishlashda “o‘lchov la’nati” (*ing. tilida - curse of dimensionality*) muammosidan deyarli ta’sirlanmaydi. Klassik usullarda o‘lchamlar soni oshgani sayin masofalar yoki zichlik hisob-kitoblari ishonchsizlashadi. Isolation Forest esa har bir bo‘linishda xususiyatlarni mustaqil va tasodifiy tanlagani uchun bu muammoga duch kelmaydi. Bu xususiyat tufayli u kiberxavfsizlikda DDoS hujumlari, zararli dasturlar va moliyaviy firibgarliklarni aniqlashda, shuningdek sanoat monitoringi va tibbiyot diagnostikasida keng qo‘llanilmoqda [2].

Uchinchidan, algoritm shovqinlarga (*ing. tilida - noise*) yuqori chidamlilik va moslashuvchanlik ko‘rsatadi. Ko‘p an’anaviy metodlardan farqli o‘laroq, u deyarli giperparametrlarni sozlashni talab qilmaydi — masalan, radius yoki eng yaqin qo‘shnilar soni kabi qiymatlarni oldindan belgilash shart emas. Ensemble tamoyili (ko‘plab mustaqil daraxtlarning natijalarini birlashtirish) tufayli notekis taqsimlangan yoki shovqinli ma’lumotlarda ham barqaror va ishonchli natijalar beradi [3].

Bu uchta asosiy afzallik — yuqori samaradorlik, yuqori o‘lchovlilikka moslashuvchanlik hamda mustahkamlik — Isolation Forestni an’anaviy usullarga nisbatan amaliyotda ancha qulay va samaraliroq qiladi. Shu sababli, u kiberxavfsizlikdan tortib moliyaviy tahlil, sanoat avtomatlashtirish va tibbiyotgacha

© Sh.R. G‘ulomov, F.A. Normirzayev

bo‘lgan ko‘plab sohalarda tobora kengroq qo‘llanilmoqda.

3. Nazariy asoslari

Isolation Forest algoritmi o‘zining o‘ziga xos ishlash tamoyillari va nazariy asoslari bilan boshqa daraxt asosli algoritmlardan ajralib turadi. Ushbu algoritm ma’lumotlarni izolyatsiya qilish (*ing. tilida — isolation*) tamoyiliga asoslanib, anomaliyalarni tez va samarali aniqlashga mo‘ljallangan bo‘lib, nazoratsiz o‘qitish (*ing. tilida — unsupervised learning*) usullarining eng muvaffaqiyatlilaridan biri hisoblanadi [1].

4. Isolation Forestning asosiy tamoyillari

Isolation Forest algoritmi anomaliyalarni ajratib ko‘rsatishda bir nechta o‘zaro bog‘liq tamoyillarga tayanadi, ularning har biri algoritmning samaradorligini ta’minlaydi.

Algoritmning markaziy g‘oyasi shuki, anomaliyalar umumiy populyatsiyada kam uchraydi va boshqa nuqtalardan osongina ajralib turadi. Shu sababli, tasodifiy bo‘linishlar orqali ularni normal kuzatuvlarga nisbatan ancha tezroq izolyatsiya qilish mumkin. Bu tamoyil algoritmning hisoblash samaradorligi va tezkorligining asosiy poydevori bo‘lib, anomaliyalarning tabiiy “ajraluvchanligi”dan kelib chiqadi [1].

Algoritm tasodifiy tanlangan xususiyat (*ing. tilida — feature*) bo‘yicha va ushbu xususiyatning minimal va maksimal qiymatlari oralig‘idagi tasodifiy bo‘linish nuqtasi asosida ma’lumotlarni rekursiv ravishda ikkiga ajratadi. Har bir bo‘linish yangi tugun (*ing. tilida — node*) hosil qiladi va jarayon har bir nuqta to‘liq yakkalanmaguncha yoki oldindan belgilangan chuqurlik chegarasiga yetguncha davom etadi. Anomaliyalar uchun bu jarayon odatda kamroq bo‘linish talab qiladi, ya’ni ular daraxt tuzilmasida qisqa yo‘llarga (*ing. tilida — shorter path lengths*) ega bo‘ladi [4].

Anomaliya darajasini baholashda algoritm kuzatuvning ansanbl daraxtlaridagi o‘rtacha yo‘l uzunligidan
(*ing. tilida — average path length*) foydalanadi. Normal kuzatuvlar chuqurroq tugunlarda joylashgan bo‘lsa, anomaliyalar yuzaki tugunlarda izolyatsiya qilinadi. Hisoblangan anomaliya skori (*ing. tilida — anomaly score*) quyidagi formula asosida aniqlanadi:

$$s(x,n) = 2^{-\frac{E(h(x))}{c(n)}}$$

bu yerda: E(h(x)) - o‘rtacha yo‘l uzunligi;

c(n) - normalizatsiya koeffitsienti.

Skor 0,5 ga yaqin bo‘lsa, kuzatuv normal deb hisoblanadi; 1 ga yaqin bo‘lsa, aniq anomaliya; 0,5 dan past bo‘lsa, butunlay normal deb baholanadi [1].

Amaliy qo‘llashda anomaliya aniqlash skorning oldindan belgilangan chegara qiymati (*ing. tilida — threshold*) bilan solishtiriladi. Bu chegara ma’lumotlar to‘plamining tabiati, shovqin darajasi va vazifaning talablariga qarab moslashtiriladi. Masalan, yuqori sezuvchanlik talab qilinadigan kiberxavfsizlik tizimlarida chegara pastroq (masalan, 0,6 dan yuqori skor anomaliya deb belgilanadi) qo‘yilishi mumkin, bu esa ko‘proq potensial tahdidlarni aniqlashga imkon beradi [3].

Isolation Forestning ushbu nazariy asoslari uni katta hajmli, yuqori o‘lchovli (*ing. tilida — high-dimensional*) va shovqinli ma’lumotlar bilan ishlashda boshqa klassik anomaliya aniqlash usullariga nisbatan tezkor, samarali va mustahkam qiladi. Aynan shu xususiyatlari tufayli algoritm kiberxavfsizlik, moliyaviy firibgarlik aniqlash, sanoat monitoringi va tibbiyot diagnostikasi kabi sohalarda tobora keng qo‘llanilmoqda [4].

5. Rivojlanish tarixi

Isolation Forest algoritmi anomaliyalarni aniqlash sohasidagi muhim yutuqlardan biri bo‘lib, uning rivojlanish tarixi an’anaviy usullarning cheklovlarini bartaraf etishga qaratilgan tadqiqotlar va takomillashtirish bosqichlarini o‘z ichiga oladi. Algoritm dastlab hisoblash samaradorligi pastligi, giperparametrlarga yuqori sezgirlik va yuqori o‘lchovli (*ing. tilida — high-dimensional*) ma’lumotlardagi “o‘lchov la’nati” (*ing. tilida — curse of dimensionality*) kabi muammolarni hal qilish maqsadida ishlab chiqilgan bo‘lib, keyingi yillarda turli takomillashtirilgan variantlari bilan keng qo‘llanilmoqda [1].

(cc) BY © Sh.R. G‘ulomov, F.A. Normirzayev

Isolation Forest algoritmi birinchi marta 2008 yilda Fei Tony Liu, Kai Ming Ting va Zhi-Hua Zhou tomonidan ishlab chiqilgan va IEEE International Conference on Data Mining (ICDM) konferensiyasida taqdim etilgan [1]. Ushbu ishda algoritm anomaliyalarni “izolyatsiya qilish” (*ing. tilida — isolation*) tamoyiliga asoslangan yangi model sifatida tavsiflanadi, bu an’anaviy usullardan farqli o‘laroq, normal nuqtalarni profil yaratish o‘rniga anomaliyalarni to‘g‘ridan-to‘g‘ri ajratib ko‘rsatishga qaratilgan. Tadqiqotchilar algoritmning chiziqli vaqt murakkabligi (*ing. tilida — linear time complexity*) va parametrlarga kam bog‘liqligini ta’kidlab, uni katta hajmli ma’lumotlar bilan ishlash uchun ideal vosita deb baholaganlar [1].

Algoritmning yaratilishiga asosiy sabab an’anaviy anomaliya aniqlash usullarining — masalan, Local Outlier Factor (LOF) kabi zichlikka asoslangan yoki One-Class SVM kabi masofaga asoslangan yondashuvlarning — bir qator cheklovlari edi. Bu usullar ko‘pincha kvadratik yoki undan yuqori vaqt murakkabligiga ega bo‘lib, katta ma’lumotlar to‘plamlarida hisoblash jihatidan samarasiz ishlaydi hamda ko‘plab giperparametrlarni (masalan, qo‘shnilar soni yoki radius) qo‘lda sozlashni talab qiladi [1]. Bundan tashqari, yuqori o‘lchovli ma’lumotlarda masofalar yoki zichlik hisob-kitoblari ishonchsizlashib, noto‘g‘ri natijalarga olib kelishi mumkin edi. Isolation Forest esa ushbu muammolarni tasodifiy daraxtlar ensemble (*ing. tilida — random forest ensemble*) va izolyatsiya tamoyili orqali hal qilib, parametrlarga deyarli bog‘liq bo‘lmagan va tezkor yechim taklif etdi [2].

2012 yilda xuddi shu mualliflar tomonidan algoritmning to‘liq nazariy asoslari va qiyosiy tahlillari ACM Transactions on Knowledge Discovery from Data jurnalida chop etilib, uning ishonchliligi va amaliy samaradorligi yanada mustahkamlandi [2]. Keyingi muhim takomillashtirish 2018–2019 yillarda Sahand Hariri, Matias Carrasco Kind va Robert J. Brunner tomonidan “Extended Isolation Forest” (EIF) nomi bilan taqdim etilgan variant bo‘ldi [3]. Ushbu kengaytmada asl algoritmning daraxt bo‘linishidagi yo‘nalishli bias (*ing. tilida — branching bias*) muammosi hal qilinib, hyperplane (*ing. tilida — hyperplane*) asosidagi tasodifiy bo‘linishlar joriy etildi, bu esa anomaliya skorlarining izchilligini va yuqori o‘lchovli ma’lumotlardagi aniqlikni oshirdi [3].

So‘nggi yillarda algoritmning chuqur o‘qitish bilan integratsiyalangan variantlari (*ing. tilida — deep Isolation Forest*) va streaming ma’lumotlar uchun moslashtirilgan modellari ham paydo bo‘ldi, bu esa uni real vaqtda ishlaydigan tizimlarda yanada samarali qildi [9]. Isolation Forest va uning variantlari hozirda kiberxavfsizlik, moliyaviy firibgarlik aniqlash, sanoat monitoringi va astronomik ma’lumotlar tahlilida keng qo‘llanilmoqda, bu algoritmning anomalialarni aniqlash sohasidagi yetakchi o‘rnini tasdiqlaydi [5], [7], [8].

6. Asosiy takomillashtirishlar

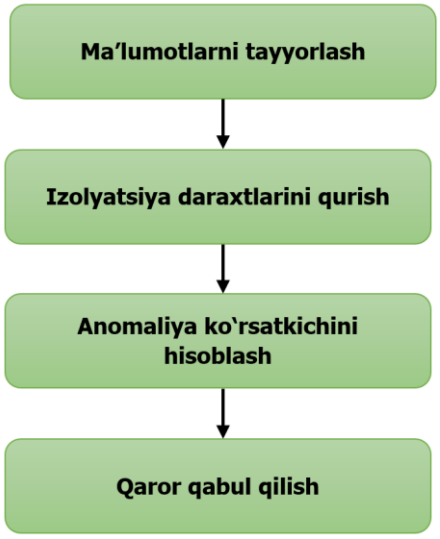
Isolation Forest algoritmi anomal ma’lumotlarni aniqlashda samarali, unsupervised usul sifatida ilmiy hamjamiyat tomonidan keng qabul qilinganidan keyin [1], uning aniqligini, ishonchliligini va qo‘llash doirasini oshirishga qaratilgan bir qator takomillashtirishlar amalga oshirildi. Ushbu yaxshilanishlar algoritmning dastlabki cheklovlarini bartaraf etishga, xususan, yuqori o‘lchamli va murakkab ma’lumotlarda barqarorlikni ta’minlashga yo‘naltirilgan bo‘lib, natijada algoritm zamonaviy real-vaqt ilovalarida yanada mustahkam va moslashuvchan vosita sifatida o‘rnini mustahkamladi.

Eng muhim takomillashtirishlardan biri Kengaytirilgan Isolation Forest (*ing. Extended Isolation Forest, EIF*) hisoblanadi [2]. An’anaviy Isolation Forest faqat o‘q bo‘ylab parallel bo‘linishlarga (*ing. axis-parallel splits*) asoslangan bo‘lsa, EIF tasodifiy yo‘nalishdagi gipertekisliklar (*ing. random hyperplanes*) orqali bo‘linish mexanizmini joriy etadi. Bu o‘zgartirish anomal ballarini (*ing. anomaly scores*) yanada izchil va ishonchli hisoblash imkonini beradi, chunki klassik usulda yuqori o‘lchamli fazolarda (*ing. high-dimensional spaces*) bo‘linish yo‘nalishining tasodifiyligi ballar tarqalishida artefaktlar keltirib chiqarishi mumkin edi. EIF turli zichlikdagi klasterlarga ega ma’lumotlarda lokal anomaliyalarni (*ing. local anomalies*) aniqlashda sezilarli ustunlikka ega bo‘lib, astronomiya, kibertahdidlarni aniqlash va moliyaviy fraud monitoring kabi sohalarda keng qo‘llanilmoqda.

© Sh.R. G‘ulomov, F.A. Normirzayev

So‘nggi yillarda algoritmning masshtablanuvchanligi (*ing. scalability*) va real-vaqt ilovalariga moslashuvi bo‘yicha ham muhim yutuqlarga erishildi. Dastlabki versiyada Isolation Forest chiziqli vaqt murakkabligiga (*ing. linear time complexity, O(n)*) ega bo‘lib, katta hajmdagi ma’lumotlar uchun mos edi [1]. Keyinchalik scikit-learn kutubxonasidagi realizatsiyada ko‘p yadroli parallel hisoblash (n_jobs parametri) qo‘llab-quvvatlandi, shuningdek, Apache Spark platformasida distributed versiyalar ishlab chiqildi, bu real-vaqt (*ing. near real-time*) anomal aniqlashda algoritmning samaradorligini oshirdi. 2025-yilda taklif etilgan Online Isolation Forest varianti esa oqim ma’lumotlari (*ing. streaming data*) bilan ishlash uchun maxsus optimallashtirilgan bo‘lib, dinamik muhitlarda doimiy o‘rganish imkonini berdi.

Isolation Forestning ish jarayoni anomal nuqtalar kam uchraydigan va osongina izolyatsiya qilinadigan tabiiy xususiyatidan kelib chiqib, izchil va samarali tarzda tashkil etilgan. Jarayonning asosiy bosqichlari 1-rasmda blok-sxema shaklida tasvirlangan bo‘lib, u ma’lumotlarni tayyorlashdan boshlab, qaror qabul qilishgacha bo‘lgan ketma-ketlikni aniq ko‘rsatadi.



1-rasm. Isolation Forest algoritmining ish jarayoni bosqichlari

Rasmda tasvirlanganidek jarayon ma’lumotlarni oldindan tayyorlash bilan boshlanadi, bu bosqich modelning ishonchliligi va aniqligini ta’minlashda hal qiluvchi ahamiyatga ega. Ma’lumotlarda bo‘sh qiymatlar mavjud bo‘lsa, ular imputatsiya yoki olib tashlash yo‘li bilan bartaraf etiladi; xususiyatlar turli o‘lchov birliklarida bo‘lsa, min-max normalizatsiya yoki z-score standartlashtirish usullari qo‘llaniladi; shuningdek, shovqin keltirib chiqaradigan yoki anomal aniqlashga bevosita ta’sir qilmaydigan belgilar feature selection usullari orqali filtrlanadi. Ushbu tayyorgarlik noto‘g‘ri ijobiy va salbiy natijalarni minimallashtirib, modelning umumiy samaradorligini sezilarli oshiradi.

Keyingi bosqichda izolyatsiya daraxtlari (ing. *isolation trees*) quriladi, bu algoritmning markaziy mexanizmini tashkil etadi. Har bir daraxt mustaqil ravishda tasodifiy tanlangan subsampl ma’lumotlar asosida rekursiv bo‘linish orqali yaratiladi: tasodifiy xususiyat tanlanadi, uning qiymatlari oralig‘ida tasodifiy nuqta belgilanadi va ma’lumotlar ikkiga ajratiladi. Jarayon belgilangan chuqurlik chegarasiga yetguncha yoki barcha kuzatuvlar izolyatsiya qilinmaguncha davom etadi. Anomal nuqtalar zich klasterlardan uzoqroq joylashganligi sababli odatda qisqaroq yo‘llarda izolyatsiya qilinadi, bu esa algoritmning asosiy afzalligini - normal va anomal nuqtalar o‘rtasidagi tabiiy farqni samarali ekspluatatsiya qilishni ta’minlaydi.

Ushbu bosqichdan so‘ng har bir kuzatuv uchun anomal ko‘rsatkichi (*ing. anomaly score*) hisoblanadi, bu ko‘rsatkich o‘rmondagi barcha daraxtlarda o‘rtacha yo‘l uzunligiga (*ing. average path length*) asoslanadi. Qisqa yo‘lga ega nuqtalar yuqori ball oladi va anomal deb baholanadi; uzun yo‘llar esa normal nuqtalarga xosdir. Hisoblash ikkilik qidiruv daraxtidagi muvaffaqiyatsiz qidiruv uzunligi bilan normallashtirishga asoslangan bo‘lib, ball 0.5 atrofida o‘rtacha, 1 ga yaqin anomal holatni ko‘rsatadi [2].

Qaror qabul qilish bosqichida oldindan belgilangan chegara qiymati (*ing. threshold*) asosida yakuniy tasnif amalga oshiriladi. Anomal ball chegara qiymatidan yuqori bo‘lsa, kuzatuv anomal deb belgilanadi; aks holda, normal deb qabul qilinadi. Chegara qiymati domen bilimiga, vazifa talablariga yoki precision-recall egri chizig‘i kabi eksperimental baholashga asoslanib tanlanadi, bu algoritmning yuqori moslashuvchanligini ta’minlaydi.

© Sh.R. G‘ulomov, F.A. Normirzayev

Amaliy misol

Quyidagi Python kodi **Scikit-learn** kutubxonasidan foydalangan holda Isolation Forest modelini qanday qo‘llash mumkinligini ko‘rsatadi:

```
from sklearn.ensemble import IsolationForest
import numpy as np
import pandas as pd
# Sun'iy ma’lumotlar yaratish
np.random.seed(42)
X = np.random.randn(100, 2) # 100 ta normal nuqta
X_anomaly = np.random.uniform(low=-4, high=4, size=(5, 2)) # 5 ta anomaliya
# Ma’lumotlarni birlashtirish
X_total = np.vstack([X, X_anomaly])
# Isolation Forest modelini yaratish
model = IsolationForest(n_estimators=100, contamination=0.05, random_state=42)
model.fit(X_total)
# Anomaliya ko‘rsatkichlarini hisoblash
anomaly_scores = model.decision_function(X_total)
predictions = model.predict(X_total)
# Natijalarni DataFrame sifatida ko‘rish
df = pd.DataFrame(X_total, columns=['Feature 1', 'Feature 2'])
df['Anomaly Score'] = anomaly_scores
df['Prediction'] = predictions # -1 anomaliya, 1 normal
print(df.head(10))
```

Tahlil:

n_estimators=100 – 100 ta izolyatsiya daraxti yaratadi.

contamination=0.05 – ma’lumotlarning taxminan 5% anomaliya bo‘lishi kutilmoqda.

decision_function() – har bir nuqta uchun anomaliya skori hisoblanadi.

predict() – -1 (anomaliya) yoki 1 (normal) deb belgilaydi.

Isolation Forest algoritmining samaradorligi uning anomaliyalarni tez aniqlash, katta hajmdagi ma’lumotlar bilan ishlash va minimal o‘qitish talab qilish kabi afzalliklari bilan bog‘liq. Shu bilan birga, murakkab ma’lumotlar to‘plamlarida samaradorlikning pasayishi, tasodifiy natijalar va aniqlikdagi cheklovlar kabi kamchiliklari ham mavjud. Isolation Forestning afzalliklari va kamchiliklari quyidagi 1-jadvalda keltirilgan:

1-jadvali.

Isolation Forestning afzalliklari va kamchiliklari

Afzalliklar	Kamchiliklar
Anomaliyalarni tez aniqlaydi, chunki u tasodifiy bo‘linishlardan foydalanadi va murakkab hisob-kitoblarga hojat qoldirmaydi.	Juda murakkab ma’lumotlar to‘plamlarida (masalan, ko‘p o‘lchovli yoki bir-biriga bog‘liq xususiyatlar bo‘lsa) samaradorligi pasayishi mumkin.
Katta hajmdagi ma’lumotlar bilan ishlaganda samarali, chunki u o‘rmon qurishda faqat tasodifiy namunadan foydalanadi.	Anomaliyalar bir-biriga yaqin joylashgan bo‘lsa, ularni ajratishda qiyinchilik tug‘dirishi mumkin.
O‘qitish jarayoni minimal talab qiladi va ko‘p parametr sozlashga ehtiyoj sezmaydi.	Normal ma’lumotlar va anomaliyalar o‘rtasidagi farq kam bo‘lsa, noto‘g‘ri tasniflash ehtimoli yuqori bo‘ladi.
Har xil turdagi ma’lumotlar (masalan, sonli yoki kategoriyali) bilan moslashuvchan ishlaydi.	Tasodifiy bo‘linishlarga tayanadiganligi sababli, natijalar har safar biroz farq qilishi mumkin (qaytariluvchanlik past).

Isolation Forest yuqori samaradorlik va oddiy ishlash tamoyiliga ega bo‘lsa-da, uni dinamik va murakkab muhitlarda qo‘llashdan oldin ushbu cheklovlarni e’tiborga olish kerak.

7. Kiberxavfsizlikdagi qo‘llanilishi

Kiberxavfsizlik sohasida anomal ma’lumotlarni aniqlash zamonaviy tahdidlarning murakkabligi va doimiy evolyutsiyasi tufayli hal qiluvchi ahamiyat kasb etmoqda, chunki an’anaviy imzo asosidagi (*ing. signature-based*) usullar noma’lum yoki zero-day hujumlarga qarshi samarasiz qolmoqda. Isolation Forest algoritmi

© Sh.R. G‘ulomov, F.A. Normirzayev

unsupervised yondashuvi, chiziqli vaqt murakkabligi (*ing. linear time complexity*) va yuqori o‘lchamli ma’lumotlarda barqaror ishlashi tufayli intruziya aniqlash, zararli dasturlar monitoringi, firibgarlik deteksiyasi hamda ichki tahdidlarni baholash kabi vazifalarda keng qo‘llanilmoqda. Algoritmning asosiy afzalligi anomal nuqtalar kam uchraydigan va osongina izolyatsiya qilinadigan tabiiy xususiyatidan kelib chiqib, normal xatti-harakatlar modelini talab qilmasdan yangi tahdidlarni aniqlash imkonini berishidadir [3].

Intruziya aniqlash tizimlarida (*ing. Intrusion Detection Systems, IDS*) Isolation Forest tarmoq trafigidagi g‘ayritabiiy namunalarni samarali ajratib, erta bosqichda hujumlarni aniqlashga yordam beradi. Algoritm tarmoq paketlari yoki oqim ma’lumotlaridagi og‘ishlarni izolyatsiya qilish orqali normal trafikdan farqli ravishda ishlaydigan harakatlarni topadi, bu esa DDoS, SYN flooding yoki boshqa hujumlarni real-vaqt rejimida (*ing. near real-time*) aniqlashda yuqori aniqlik va recall ko‘rsatkichlarini ta’minlaydi. Ayniqsa, IoT muhitlarida Isolation Forest va One-Class SVM kombinatsiyasi qo‘llanilganda, robust anomal deteksiya frameworki yuqori samaradorlikka erishgan, bu esa tarmoq intruziyalariga qarshi moslashuvchan himoyani kuchaytiradi. An’anaviy qoida asosidagi IDS'larga nisbatan Isolation Forest yangi va noma’lum tahdidlarga qarshi ustunlikka ega, chunki u oldindan belgilangan imzalar talab qilmaydi va web-traffik anomalilarini aniqlashda ham muvaffaqiyatli qo‘llanilgan. Bundan tashqari, algoritmning hybrid versiyalari intruziya aniqlashda qo‘shimcha samaradorlik ko‘rsatgan.

Zararli dasturlarni (*ing. malware*) aniqlashda Isolation Forest dasturiy jarayonlarning xatti-harakatlarini tahlil qilish orqali g‘ayritabiiy faollikni izolyatsiya qiladi, bu esa statik imzo asosidagi antiviruslardan farqli o‘laroq, polimorf yoki yangi malware turlarini ham aniqlash imkonini beradi. Algoritm tizim resurslaridan foydalanish (CPU, memory), tarmoq aloqalari va fayl operatsiyalaridagi og‘ishlarni baholab, anomal ballarni hisoblaydi, natijada yuqori aniqlik ko‘rsatkichlariga erishilgan [4]. Ushbu yondashuv IoT qurilmalarida yoki mobil platformalarda malware deteksiyasida ayniqsa samarali bo‘lib, one-class klassifikatorlar bilan integratsiya qilinganda yanada yuqori ishonchlilik ta’minlaydi. Isolation Forestning unsupervised tabiati tufayli u nol kunlik (*ing. zero-day*) malware tahdidlariga qarshi moslashuvchan vosita sifatida zamonaviy endpoint protection tizimlarida keng integratsiya qilinmoqda. So‘nggi ishlanmalarda Isolation Forest va OCSVM kombinatsiyasi IoT malware aniqlashda yuqori natijalar bergan [5].

Moliyaviy firibgarlikni (*ing. fraud detection*) aniqlashda Isolation Forest tranzaksiya ma’lumotlaridagi noodatiy namunalarni tez va samarali izolyatsiya qilish orqali kredit karta operatsiyalari, onlayn to‘lovlar yoki kredit so‘rovlaridagi anomal holatlarni topadi. Algoritmning random partitioning mexanizmi firibgarlarning doimiy o‘zgaruvchan sxemalariga moslashib, global va lokal anomalialarni aniqlaydi, bu esa an’anaviy rule-based yondashuvlarga nisbatan yuqori precision va recall beradi. Kiberxavfsizlik kontekstida firibgarlik ko‘pincha tarmoq yoki foydalanuvchi xatti-harakatlaridagi og‘ishlar bilan bog‘liq bo‘lib, Isolation Forest bu vazifada moliyaviy tizimlarning real-vaqt monitoringida muhim afzallik taqdim etadi.

Ichki tahdidlarni (*ing. insider threats*) aniqlashda Isolation Forest foydalanuvchi xatti-harakatlaridagi og‘ishlarni (*ing. User and Entity Behavior Analytics, UEBA*) tahlil qilish orqali ruxsatsiz ma’lumot ko‘chirish, shubhali kirishlar yoki administrator faoliyatidagi anomalialarni erta bosqichda aniqlaydi. Algoritm class imbalance muammosini algoritm darajasida hal qilib, normal xatti-harakatlar ichidan kam uchraydigan ichki tahdidlarni izolyatsiya qiladi, bu esa CERT datasetlari asosidagi tajribalarda yuqori samaradorlik ko‘rsatgan. Ichki tahdidlar tashqi hujumlarga nisbatan kamroq e’tiborni tortganligi sababli, Isolation Forestning unsupervised yondashuvi bunday vazifalarda ayniqsa qimmatli bo‘lib, tashkilot ichidagi ma’lumot oqishlarini minimallashtirishga yordam beradi [5].

Xulosa va istiqboldagi yo‘nalishlari

Isolation Forest algoritmi anomaliyalarni aniqlash sohasida o‘zining yuqori samaradorligi, chiziqli vaqt murakkabligi (*ing. linear time complexity*) va unsupervised tabiati bilan zamonaviy kiberxavfsizlik muammolarini hal qilishda muhim o‘rin egallaganligini isbotladi. Algoritmning asosiy afzalligi anomal nuqtalar

© Sh.R. G‘ulomov, F.A. Normirzayev

kam uchraydigan va osongina izolyatsiya qilinadigan tabiiy xususiyatidan kelib chiqib, normal xatti-harakatlar modelini oldindan talab qilmasdan yangi va noma'lum tahdidlarni aniqlash imkonini berishidadir, bu esa intruziya aniqlash tizimlari (ing. Intrusion Detection Systems, IDS), zararli dasturlar monitoringi, moliyaviy firibgarlik deteksiyasi hamda ichki tahdidlarni baholashda real-vaqt rejimida (ing. near real-time) samarali ishlatilishini ta'minlaydi. Kiberxavfsizlik kontekstida algoritm yuqori o'lchamli (*ing. high-dimensional*) va shovqinli ma'lumotlarda barqarorlik ko'rsatib, an'anaviy imzo asosidagi (*ing. signature-based*) usullarning cheklovlarini bartaraf etadi, natijada zero-day hujumlar va polimorf zararli dasturlarga qarshi moslashuvchan himoya yaratadi.

Biroq, algoritmning kuchli tomonlariga qaramay, ba'zi cheklovlari mavjud bo'lib, ularni hisobga olish kelajakdagi ilovalarda muhim ahamiyatga ega. Masalan, haddan tashqari shovqinli ma'lumotlar (*ing. noisy data*) yoki concept drift holatlarida samaradorlik pasayishi mumkin, chunki model vaqt o'tishi bilan o'zgaruvchan muhitlarga moslashish uchun qayta o'qitishni talab qiladi. Bundan tashqari, adversarial hujumlar (*ing. adversarial attacks*) ga nisbatan zaiflik kuzatiladi, bunda raqiblar maxsus tayyorlangan namunalar orqali anomal ballarni manipulyatsiya qilib, tahdidlarni yashirishi mumkin. Ushbu muammolar tadqiqotchilar tomonidan faol o'rganilmoqda va robustlikni oshirishga qaratilgan takomillashtirishlar, masalan, adversarial training yoki filtering mexanizmlari joriy etilmoqda.

Istiqbolli yo'nalishlar nuqtai nazaridan Isolation Forestning streaming ma'lumotlar (*ing. streaming data*) va real-vaqt tizimlariga moslashuvi eng muhim tendensiyalardan biri bo'lib qolmoqda. Online Isolation Forest kabi yangi variantlar dinamik muhitlarda concept driftni kuzatib, doimiy o'rganish imkonini berib, IoT va tarmoq monitoringida yuqori samaradorlikka erishmoqda. Shu bilan birga, gibrid modellarning integratsiyasi — masalan, Isolation Forestni LSTM, autoencoderlar yoki transformer asosidagi chuqur o'rganish (*ing. deep learning*) modellari bilan birlashtirish — lokal va global anomalialarni yanada aniqroq aniqlashga, shuningdek, adversarial robustnessni kuchaytirishga yo'l ochmoqda. SIEM amaliy qo'llanilishini yanada kengaytirib, kiberxavfsizlikning kelajak avlod echimlarida markaziy rol o'ynashini ta'minlaydi.

Xulosa qilinadigan bo'lsa, Isolation Forest kiberxavfsizlikda anomaliyalarni aniqlashning mustahkam va moslashuvchan vositasi sifatida o'z o'rnini mustahkamladi, ammo uning to'liq salohiyatini ochish uchun doimiy takomillashtirish va gibrid yondashuvlar zarur. Ushbu algoritmning rivojlanishi nafaqat texnologik, balki strategik jihatdan ham kiber tahdidlarga qarshi global himoyani kuchaytirishga xizmat qiladi.

Foydalanilgan adabiyotlar ro'yxati:

1. Liu, F. T., Ting, K. M., Zhou, Z. H. (2008). Isolation Forest. In: 2008 Eighth IEEE International Conference on Data Mining, pp. 413–422. DOI: 10.1109/ICDM.2008.17.
2. Hariri, S., Carrasco Kind, M., Brunner, R. J. (2019). Extended Isolation Forest. IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 4, pp. 1479–1489. DOI: 10.1109/TKDE.2019.2947676.
3. Xu, H., Pang, G., Wang, Y., Wang, Y. (2023). Deep Isolation Forest for Anomaly Detection. IEEE Transactions on Knowledge and Data Engineering, vol. 35, no. 10, pp. 10486–10496. DOI: 10.1109/TKDE.2022.3228181.
4. Liu, F. T., Ting, K. M., Zhou, Z. H. (2012). Isolation-based anomaly detection. ACM Transactions on Knowledge Discovery from Data, vol. 6, no. 1, pp. 3:1–3:39. DOI: 10.1145/2133360.2133363.
5. Ahmad, M., Imran, K., Attar, H., Solyman, A. (2025). Isolation Forest-based anomaly detection through generation offers enhanced network security. Computers & Security, vol. 142.
6. Key Cyber Security Statistics for 2025. SentinelOne; 2025. manzil: https://www.sentinelone.com/cybersecurity-101/cybersecurity/cyber-security-statistics
7. Cybersecurity Statistics 2025: Rising Threats and Industry Impact. Fortinet; 2025. manzil: https://www.fortinet.com/resources/cyberglossary/cybersecurity-statistics

(cc) BY © Sh.R. G'ulomov, F.A. Normirzayev