УДК 324 : 004.8 : 351.9
DOI https://doi.org/10.32782/apfs.v040.2023.35

Ю. О. Яцина
ORCID ID: https://orcid.org/0000-0002-7286-4655
голова ГО «Союз соціальних технологів України»

ОСОБЛИВОСТІ ЗАСТОСУВАННЯ ІННОВАЦІЙНИХ АНАЛІТИКО-СТАТИСТИЧНИХ ТЕХНОЛОГІЙ В ІНФОРМАЦІЙНИХ СИСТЕМАХ З МОНІТОРИНГУ ТА ВИЯВЛЕННЯ КОРУПЦІЇ

Постановка проблеми

Ми живемо у часи безпрецедентного сплеску інтересу до інноваційних інформаційних технологій, а саме технологій машинного навчання та штучного інтелекту. Ці інструменти вже давно стали частиною повсякденного життя для мільярдів людей. Велика кількість цифрових продуктів та послуг, таких як, наприклад, інтерактивні карти, індивідуальна реклама та особисті голосові помічники безумовно є лише верхівкою айсбергу. Деякі прихильники штучного інтелекту навіть стверджують, що впровадження штучного інтелекту буде мати такі ж наслідки, що були у вогня чи електрики [7], що він змінить майже кожну сферу людської діяльності. Не залишилася осторонь і сфера державного (публічного) управління, в тому числі напрям протидії корупції. Аналіз стану наукової розробки проблеми інноваційних аналітико-статистичних технологій як інструменту протидії корупції в державі дозволяє охарактеризувати її як науковий напрям, орієнтований на практичне застосування у сфері контролю якості діяльності органів державної влади та місцевого самоврядування провідні інформаційні технології – в першу чергу технології штучного інтелекту та машинного навчання [14; 15].

Мета дослідження

Саме тому мета нашого дослідження полягає у визначенні особливостей впровадження інноваційних аналітико-статистичних технологій в інформаційних системах з моніторингу та виявлення корупції.

Виклад основного матеріалу

Почнемо з визначення основних понять.

Під корупцію ми розуміємо протиправне використання посадовою особою наданих їй управлінських ресурсів для особистої чи групової вигоди, що може мати як матеріальну, так і нематеріальну форму. При цьому протиправне використання – це порушення як формальних нормативно-юридичних установ, включаючи норми службової поведінки та етики, так і неформалізованих норм поведінки, етики і моралі. Під інструментами протидії корупції розуміються будь-які засоби щодо створенню перешкод вчиненню корупційних діянь, здійснення опору їх поширенню, а також реагування відносно тих діянь, які вже проявилися у фактично скоєних правопорушеннях.

Інноваційні аналітико-статистичні технології визначаються:

– в широкому значенні як сукупність методів та інструментів, що базуються на використанні математичних та статистичних методів аналізу даних з метою виявлення корисних залежностей та закономірностей в даних, підвищення ефективності прийняття рішень та виявлення аномалій у різних сферах діяльності;

– у вузькому – як процес використання найсучасніших методів та технологій аналізу даних, таких як машинне навчання, глибинне навчання, нейронні мережі, обробка природної мови, аналіз графів тощо з метою виявлення складних залежностей та корисних закономірностей в даних. До таких технологій відносяться також методи аналізу даних у режимі реального часу, які дозволяють отримувати швидкі та точні результати аналізу великих обсягів даних.

Інформаційні технології, в тому числі аналітико-статистичні технології, отримують своє практичне втілення у формі інформаційної системи або платформи, що об'єднує в собі як технічні засоби (комп'ютери, засоби та канали зв'язку, периферійні пристрої, пристрої введення / виводу даних тощо), так і програмні застосунки, що забезпечують функціонування технічної та аналітичної складової, включаючи статистику.

Організації, які використовують власні чи спеціально розроблені для них інформаційні (корпоративні) системи / платформи у своєї діяльності, умовно можна називати цифровими. Додатково до інформаційної системи / платформи такі організації також можуть використовувати мобільні робочі місця, хмарні сервіси з управління персоналом, бухгалтерію і тому подібне, просувати свої товари та послуги в Інтернеті, здійснювати транзакції щодо продажу своїх продуктів / послуг онлайн. Цифрові організації можуть бути представлені в будь-якій галузі економіки: фінансовій, нафтовій, енергетичній, транспортній, зв'язку тощо. І з кожним роком їх чисельність зростає, що призводить до появи цифрових урядів та держав.

Як відомо, ідея переводу уряду в режим онлайн з'явилася ще в 1997 році в Естонії [2]. Пройшло майже 20 років і в Україні з'явився проект «Держава у смартфоні» із застосунком ДІЯ, що призвело до того, що у 2021 році саме Україна стала

першою країною у світі [4], яка на законодавчому рівні ввела електронні паспорти на рівні із паперовими. Крім цього, сервіс дозволив використовувати на рівні із паперовими аналогами водійські посвідчення, свідоцтва про народження, оформлення права на користування автотранспортним засобом та навіть сертифікати про вакцинацію проти COVID-19, а с початку введення воєнного стану ДІЯ почала застосовуватись як аналог посвідчення особи [1]. На цьому перелік функцій цієї системи не завершується. Незважаючи на це, кожна нова функція цієї системи / платформи об'єднані одним єдиним гаслом – перемогти корупцію. Якщо подивитися більш уважно, то функціонування цієї системи / платформи підпорядковується загальній схемі впровадження інформаційних технологій у сфері боротьби з корупцією [5, с. 6]:

1. Етап «цифровізації» – на цьому етапі інформаційні технології впроваджуються у сферу боротьби з корупцією з метою автоматизації процесів. Наприклад, запровадження електронних баз даних, електронного документообігу та електронної звітності.

2. Етап «відкритих даних» – на цьому етапі створюються платформи для публічного доступу до даних про державні закупівлі, бюджетні витрати, доходи та майно посадових осіб тощо, що дозволяє зацікавленим громадянам легше відслідковувати та виявляти випадки корупції.

3. Етап «цифрової ідентифікації» – на цьому етапі проваджуються засоби та технології для електронної ідентифікації громадян та посадових осіб, що дозволяє підвищити прозорість та відстежуваність державних процесів.

4. Етап «штучного інтелекту та аналітики» – на даному етапі відбувається широке впровадження систем та технологій аналітики даних та штучного інтелекту, що дозволяє більш ефективно виявляти та запобігати випадкам корупції, наприклад, за рахунок аналізу великих обсягів даних та виявлення аномалій в діяльності посадових осіб та державних структур.

5. Етап «цифрової економіки» – на даному етапі активно розвиваються блокчейн-технології, які дозволяють забезпечити надійний захист від підробок, маніпуляцій та фальсифікацій, включаючи документообіг, електронне голосування та інші процеси, пов'язані з боротьбою з корупцією.

Ці етапи є умовними та можуть перекриватися один з одним. Вони можуть різнитися у різних країнах, залежно від рівня розвитку інформаційних технологій і соціально-політичних особливостей тієї чи іншої держави. Якщо подивитися на тренди сучасного світу, можна з повною впевненістю стверджувати, що сучасні країни активно займаються питаннями цифрової трансформації, в якій Україна хоче зайняти не останнє місце, особливо в сфері «мілтек».

Що стосується сфери протидії корупції, то в даному напрямку основна увага дослідників останніх років (2018-2022) зосереджена на технологіях машинного навчання та штучного інтелекту. Машинне навчання та штучний інтелект – це дві тісно пов'язані, але різні концепції в галузі комп'ютерних наук. У загальному сенсі штучний інтелект – це область науки, яка прагне створити машини, які можуть діяти за допомогою інтелекту, аналогічного людському. Машинне навчання – це одна з технологій, що використовуються для створення таких машин [13].

Якщо більш конкретно, то машинне навчання – це методологія, що дозволяє комп'ютерам навчатися з урахуванням наявних даних без застосування явного програмування. Замість того, щоб людина написала програму, яка вирішує певне завдання, алгоритми машинного навчання використовуються для навчання комп'ютера певним патернам у даних. Комп'ютер може використовувати цю інформацію для прийняття рішень або вирішення завдань, які він не бачив раніше.

Штучний інтелект, з іншого боку, є більш загальним поняттям, яке охоплює всі технології, спрямовані на створення комп'ютерних систем, які можуть діяти інтелектуально, тобто сприймати, обробляти та використовувати знання та вирішувати завдання, які потребують людського інтелекту. Технології штучного інтелекту можуть використовувати методи машинного навчання, але можуть включати інші підходи, такі як системи експертних знань, вирішення завдань на основі знань і нейронні мережі [19].

Таким чином, машинне навчання є одним із методів, що використовуються в галузі штучного інтелекту. Обидва ці поняття використовуються в різних галузях, включаючи пошук корупції, де машинне навчання може використовуватися для аналізу великих обсягів даних, а штучний інтелект може використовуватися для розробки систем, які можуть приймати рішення на основі цього аналізу.

Окрім цих технологій, на думку багатьох авторів, для успішної боротьби з корупцією необхідне застосування наступних інформаційних технологій [5, с. 10]:

1. Блокчейн – технологія, яка може забезпечити прозорість та незамінність інформації. Завдяки цьому вона може бути застосована для створення систем електронного голосування, угод, контрактів, які неможливо підробити.

2. Децентралізовані системи зберігання даних – дозволяють зберігати дані без централізованого управління, що забезпечує додатковий захист від несанкціонованого доступу та зміни даних.

3. Big Data – дані можуть бути використані для створення прогностичних моделей корупційних схем, визначення ключових факторів, що сприяють корупції, та моніторингу динаміки корупційних процесів.

4. GIS-технології – дозволяють використовувати просторові дані для дослідження корупційних явищ та виявлення зв'язків між ними та конкретними місцями на карті.

5. Інтернет речей (IoT) – може використовуватися для збору даних, моніторингу та контролю за ключовими об'єктами, такими як громадські будівлі, дороги, транспорт тощо.

6. Голосові технології та розпізнавання мови – можуть бути використані для створення систем розпізнавання голосу та подальшої автоматизації державних процесів.

В чому полягає специфіка застосування інноваційних аналітико-статистичних технологій в інформаційних системах з моніторингу та виявлення корупції? Річ у тому, що проникнення інформаційних технологій, з одного боку, дає цифровим організаціям, підприємствам та установам перевагу у швидкості надання послуг, їх якості та ціні тощо, але з іншого боку, зростають ризики кібербезпеки. Для мінімізації ризиків кібербезпеки застосовують різноманітні технічні рішення. Найбільш повний перелік інноваційних аналітико-статистичних технологій, а точніше виробників подібних систем можна знайти тут [8], що розподілений за 12 напрямками:

– безпека мереж та інфраструктури (network & infrastructure security);
– моніторинг та адміністрування інструментів захисту інформації (MSSP);
– управління ідентифікацією та доступом (identity & access management);
– веб-безпека (web security);
– безпека даних (data security);
– реагування на інциденти та заходи безпеки (security ops & incident response);
– розвідка загроз (threat intelligence);
– управління цифровими ризиками (digital risk management);
– безпека блокчейн (blockchain);
– захист кінцевих точок (endpoint security);
– безпека застосунків (application security);
– безпека мобільних пристроїв (mobile security);
– інтернет речей (IoT);
– безпека засобів комунікації (messaging security);
– послуги та консультації з питань безпеки (security consulting & services);
– безпека транзакцій та захист від шахрайства (fraud & transaction security);
– хмарна безпека (cloud security);
– ризик і комплаєнс (risk & compliance).

Проблема полягає в тому, що існуючі інформаційно-технічні засоби ефективно захищають від зовнішніх загроз (DDOS-атаки, підбір паролів, обхід мережевих екранів, зараження вірусами), в той час як корупція є внутрішньою загрозою. В контексті нашого дослідження ми маємо справу із питанням захисту інтересів організації (будь-то звичайна компанія чи система органів державної влади) від внутрішнього зловмисника (корупціонера, шахрая), для якого в інформаційній науці було запозичено англомовний термін «інсайдер» (від англ. "insider") – «особа, яка завдяки своєму службовому становищу або спорідненим зв'язкам має доступ до конфіденційної інформації ..., що недоступна широкій громадськості, та може використати її у власних цілях з метою збагачення, одержання неконкурентних переваг, привілеїв тощо» [3, с. 312].

Інсайдерами можуть бути: 1) фізичні особи (власники істотної участі, управлінський персонал (персонал із статусом ОПР, працівники внутрішнього аудиту, члени ревізійної комісії, контролери), асоційовані особи (рідні брати, сестри, батьки, чоловік, дружина або повнолітні діти керівників, контролерів організації та акціонерів-власників істотної участі (від 10%)); 2) юридичні особи (власники істотної участі, афілійовані, споріднені особи та асоційовані особи.

В свою чергу, наявність інсайдерів створює інсайдерські загрози – шкідливі для організації активності, які походять від співробітників всередині організації, зокрема – від діючих працівників, колишніх працівників, підрядників, ділових партнерів і навіть завербованих працівників або працівників, навмисно впроваджених в організацію, які мають доступ до конфіденційної інформації в рамках своїх посадових обов'язків та які мають уявлення про систему управління інформаційною безпекою організації.

Саме через наявність адекватних засобів захисту від зовнішніх атак, значними витратами на їх здійснення, більш економічно вигідним є використання інсайдерів, які можуть не тільки, наприклад, своїми руками чи руками колег несанкціоновано вивантажити конфіденційну інформацію на власний носій, переслати через електронну пошту, вилучити зашифрований жорсткий диск з робочої станції або сервера тощо, але й можуть провести низку легальних процедур щодо виведення інформації із захищених контурів та сховищ за рахунок використання власних посадових повноважень та мережевих привілеїв. Найбільшу цінність в сучасних умовах цифровізації країн для інсайдерів становлять об'єкти концентрації даних – сховища даних або Big Data.

Взагалі, на нашу думку, найбільш дотичними до проблеми протидії корупції є інструменти з реагування на інциденти та засоби безпеки та оцінки

ризику і комплаєнсу. Застосунки з реагування на інциденти та заходи безпеки поділяються на:

– управління інформацією про безпеку та події безпеки (security information and event management)

– реагування на інциденти безпеки (security incident response)

– аналіз безпек (security analytics).

В свою чергу ризик і комплаєнс поділяється на рішення з:

– оцінки та візуалізація ризиків;

– вимірювання ризику;

– симуляції атак та взломів (BAS – breach and attack simulation);

– врядування, управління ризиками та дотримання вимог законодавства (GRS – governance, risk management and compliance);

– освіти з питань інформаційної безпеки.

Таким чином, інноваційними аналітико-статистичними технологіями, що можуть бути використані корпоративних системах моніторингу та протидії корупції в державі можна вважати ті технології, що сфокусовані на пошуку ознак аномальної (зловмисної) активності користувачів у великих масивах даних, що відображають результати діяльності публічних (урядових) організацій, підприємств та установ. Ці інформаційні масиви формуються з трьох джерел:

– даних, що характеризують основну господарську діяльність органів публічної влади, в тому числі дані постійного моніторингу режиму їх роботи;

– персональних даних співробітників органів публічної влади, які одночасно здійснюють свою активну професійну діяльність;

– дані щодо поточних операцій з клієнтами та контрагентами органів державної влади та місцевого самоврядування, що здійснюються зазначеними співробітниками в рамках виконання покладених на них посадових обов'язків.

В умовах становлення цифрової цивілізації для боротьби з внутрішніми зловмисниками інформаційних систем використовується ряд технічних засобів, таких як DLP (data loss prevention (запобігання втраті даних) або data leakage prevention (запобігання витоку даних)), SIEM (security information and event management – об'єднання двох термінів, що позначають використання програмного забезпечення: SIM (security information management) – управління інформацією про безпеку та SEM (security event management) – управління подіями безпеки) тощо. Але ці засоби проводять моніторинг нелегальних каналів, в той час як корупціонери чи фінансові шахраї використовують легальні канали руху інформації. Їх злочинні дії, прикриті діяльністю щодо виконання посадових обов'язків, з точки зору науки про інформацію класифікуються як інсайдерські атаки. Вони стають не тільки більш економічно вигідним способом злочинних дій в інформаційному просторі, а все частіше і єдиними можливими. В більшості випадків злочинні (корупційні, шахрайські) дії в умовах цифрової організації пов'язані із маніпуляціями чи махінаціями з інформацією – її крадіжкою, фальсифікацією чи банальним приховуванням.

Тому для пошуку ознак корупційної (інсайдерської) активності в масивах даних необхідно мати навички оперувати великими обсягами поточної інформації, аналізуючи її та формуючи рекомендації для прийняття відповідних управлінських рішень в умовах жорстких часових обмежень. Саме цим і обумовлено використання в подібному аналізі інноваційних аналітико-статистичних засобів, в тому числі, систем штучного інтелекту, за допомогою яких забезпечується інтелектуальний аналіз даних у відповідному часовому режимі. При цьому необхідно, щоб сформовані в процесі такого аналізу рекомендації були зрозумілі не лише експертам, а й керівникам відділів служб безпеки, що займаються як питаннями інформаційної безпеки, так і питаннями протидії шахрайству і корупції.

Отже, в умовах цифрової організації особа, що зловживає своїм посадовим становищем з метою отримання власного зиску чи зиску для третіх сторонніх осіб, в контексті інформаційних систем протидії їй, відноситься до категорії інсайдерів.

Наявність інсайдеру передбачає ризик інсайдерської загрози або «дій інсайдера, які наражають на ризик організацію або її ресурси» [16]. Як зазначено в [20] зловмисні інсайдерські загрози поділяються на дві групи: зрадників та маскувальників. Зрадники мають повне уявлення про системи, з якими вони працюють щодня, а також фактичну безпекову політику. Вони зазвичай діють від свого імені, тому використовують власні повноваження для зловмисних дій. Маскувальники можуть мати набагато менші знання про організацію, ніж зрадники. Це зловмисники, які крадуть облікові дані іншого законного користувача, а потім використовують їх для здійснення зловмисної дії від його імені.

Згідно з дослідженням [6] інсайдерські загрози можна класифікувати за такими типами:

– IT-саботаж (sabotage), під час якого інсайдер використовує інформаційні технології для заподіяння конкретної шкоди організації чи фізичній особі. Такими інсайдерами зазвичай є незадоволені співробітники з відповідною технічною освітою, за умов, що в них є адміністративні привілеї. Прикладом може слугувати логічна бомба, яка активується після звільнення співробітника;

– розкрадання інтелектуальної власності або так зване шпигунство (theft), яке зазвичай здійснюється технічним персоналом, наприклад,

інженерами та розробниками, а також нетехнічним персоналом (клерки та продавці). Злочинці можуть викрадати інформацію, до якої вони мають щоденний доступ, і забрати її з собою в разі звільнення (наприклад, використовуючи власний IP для своєї справи, передаючи її новому роботодавцю або іншій організації);

– шахрайство (fraud) – коли інсайдер використовує інформаційні технології для несанкціонованої зміни, додавання чи видалення даних організації для особистої вигоди чи крадіжки. Інсайдерське шахрайство зазвичай здійснюється співробітниками низького рівня з нетехнічним досвідом, наприклад, співробітниками відділу кадрів або служби підтримки. Причиною цього часто є жадібність чи фінансові труднощі, і це вид злочинів, зазвичай, носить довгостроковий характер. Вербування таких шахраїв зовнішніми структурами також дуже поширене.

Саме в контексті виявлення інсайдерів в інформаційних системах використовуються найновіші аналітико-статистичні технології, засновані на технології машинного навчання. Зміст таких технологій полягає в тому, щоб «навчити» інформаційну систему розпізнавати дії інсайдерів, тобто виявляти аномалії. Аномалії можна вважати тими шаблонами в даних, що відхиляються від очікуваних. Методи виявлення аномалій використовують контрольований, напівконтрольований або неконтрольований підхід. Контрольовані підходи мають вищі показники виявлення і менші обсяги помилкових сигналів, ніж неконтрольовані підходи. Однак, неконтрольовані підходи можуть виявити невідому поведінку, але контрольовані – не можуть. Отримання точних і репрезентативних даних про всі види поведінки часто є затратним. Маркування зазвичай проводиться експертом. Отримання аномальних даних поведінки складніше, ніж отримання даних про номінальну поведінку. Дані часто мають динамічний характер, наприклад, можуть виникати нові аномалії, а старі втрачати свій статус [9].

Найбільш повний перелік актуальних (робочих) на даний час методів пошуку аномалій можна знайти в репозиторії www.github.com, а саме – ADBench [10]. ADBentch є спільним проектом дослідників Шанхайського університету фінансів та економіки (SUFE) і Університету Карнегі-Меллона (CMU). Проект розроблено авторами найбільш популярних бібліотек виявлення аномалій, включаючи виявлення аномалій для табличних даних чи баз даних (PyOD), часових рядів (TODS) та графів (PyGOD).

За результатами проекту було проведена оцінка продуктивності 30 алгоритмів виявлення аномалій в масивах даних з використанням 57 наборів даних в кількості 98 436 експериментів за трьома параметрами:

– типу методу машинного навчання (supervision): тести працездатності алгоритмів включають 14 алгоритмів контрольованого, 7 напівконтрольованого і 9 неконтрольованого навчання;

– характеру аномалії, що досліджується: локальні, глобальні, кластерні, залежні;

– стійкості і стабільності алгоритму в умовах наявності інформаційного шуму чи неповних даних.

Як було вказано, проєкт поєднав в собі розробки за трьома напрямами – виявлення аномалій в табличних даних, часових та графових.

1. PyOD представляє собою бібліотеку Python для виявлення аномальних об'єктів у багатовимірних даних. Оригінальний PyOD включає понад 40 алгоритмів виявлення, починаючи від класичного LOF до найсвіжішого ECOD [18; 22].

2. TODS – це універсальна система автоматичного машинного навчання для виявлення викидів (аномалій) в даних багатовимірних часових рядів. TODS включає модулі для побудови систем виявлення аномалій на основі машинного навчання, включаючи: обробку даних, обробку часових рядів, аналіз ознак (добування), алгоритми виявлення і модуль посилення. Функціональні можливості, надані через ці модулі, включають попередню обробку даних для загальних цілей, згладжування/перетворення даних часових рядів, витягування ознак з часових/частотних блоків даних, різні алгоритми виявлення, включаючи алгоритми залучення експертів (людського досвіду) для калібрування системи [11; 21].

3. PyGOD – бібліотека Python для виявлення викидів (аномалій) в графах. PyGOD включає в себе понад 10 алгоритмів виявлення аномалій в графах, таких як DOMINANT або GUIDE [12; 17].

Однією з переваг даного комплексу алгоритмів є простота їх застосування в сенсі обсягу використання коду – для запуску більшості алгоритмів достатньо 5 строчок коду.

Загалом, сучасні підходи до виявлення інсайдерських загроз базуються на методах криміналістики (forensic) і зазвичай обмежують себе вивченням журналів кібербезпеки для застосування алгоритмів виявлення аномалій або пошуку відбитків (сигнатур). Ці алгоритми необхідні для виявлення інсайдерської загрози, але вони лише є складовою повного рішення. Тому більш ефективними слід вважати інтегровані рішення з виявлення та запобігання інсайдерським загрозам, що використовують семантично марковані дані з інфраструктури кібернетичного та фізичного контролю доступу і забезпечують проактивне (упереджене) виявлення порушників.

В деяких реалізаціях система виявлення загроз передбачає та формулює гіпотези поведінки, що вказує на поточні інсайдерські атаки, шляхом

збору необроблених цифрових і фізичних даних, аналізу спостережень з необроблених даних, і виявлення підозрілої поведінки. В окремих випадках висновки системи потребують вивчення даних спостережень незвичних закономірностей, що вказують на зміну звичок / ролей людини, навмисне або ненавмисне порушення політик, неналежну конфігурацію системи контролю доступу або активну зловмисну поведінку, тобто, можливе виникнення емпіричних закономірностей.

Для успішної ідентифікації інсайдерів потрібні набори даних, що описують області діяльності і життя людини в різних сферах, її поведінку на роботі, як користувача інформаційної системи і так далі. Особливо важливо враховувати характеристики співробітників, які потрапляють в коло підозрюваних при оцінці можливого залучення до витоків інформації. Ефективність аналізу тільки однорідних даних в таких випадках викликає сумніви, оскільки інсайдерами зазвичай є достатньо розумні, професійні люди, яким довіряють доступ до обробки цінної інформації, тому саме вони знають способи обходу систем контролю.

Тому аналіз інсайдерської діяльності повинен здійснюватися на всьому діапазоні збору даних – починаючи з даних технічних систем та систем управління, закінчуючи даними особистого життя персоналу, оперативної інформації служб безпеки або детективних агентств, баз даних різного призначення (різних видів довідкових систем, в тому числі баз даних правоохоронних та інших державних органів, соціальних мереж, даних рекрутингових агентств, державних реєстрів тощо). Чим різноманітніше вихідні дані – тим більш точна робота моделей, і тим більш ймовірність проаналізувати непомітні відхилення або взаємозв'язки.

В контексті протидії корупції, система пошуку аномалій має зосереджуватися на пошук особливих аномалій, які називаються змовою. Наприклад, збір цінної інформації працівниками з використанням особистих відносин є серйозною проблемою інформаційної безпеки в урядових і комерційних організаціях. У цьому випадку часто порушник інформаційної безпеки має доступ до частини цінної інформації, але не має права на її отримання повністю. Прикладом такої ситуації може бути поділ спільної проблеми на різні підрозділи організації. Тоді інсайдер буде намагатися отримати відсутню інформацію у своїх друзів з інших підрозділів або піти на змову щодо збору та подальшого продажу цінної інформації.

Висновки

Отже, головною метою впровадження інноваційних аналітико-статистичних технологій як інструменту протидії корупції в сучасній державі є автоматизація рутинних процесів обробки і аналізу даних у зв'язку із зростаючим рівнем цифровізації повсякденного життя людини. Автоматизація потрібна саме в тих сферах, за результатами діяльності яких формуються значні обсяги даних (текстових, табличних, даних реального часу тощо), більшу частину яких ефективно проаналізувати обмеженим людським мозком неможливо у зв'язку із дією принципу технічної сингулярності. Результатом автоматизації має стати більш ефективний процес пошуку аномалій в масивах даних, що передбачає проведення:

– аналізу первинних даних, що постійно накопичуються (в процесі функціонування публічної організації, підприємства, держави), на предмет моніторингу неявно присутніх відомостей про взаємодію співробітників з інформаційними ресурсами;

– активації алгоритмів виявлення в цих даних неявних описів взаємодій, що мають ознаки (явні або потенційні) шкідливих наслідків.

В дійсності, мова йде про проблему фільтрації даних про існуючі взаємодії, які постійно оновлюються, що характеризується необхідністю:

– постійно мати справу з дуже великими обсягами первинних даних;

– враховувати (наприклад, в частині організації ряду послуг – інформаційного пошуку, підтримки в актуальному стані поточного профілю загроз і моделі порушників тощо) постійне поповнення таких даних новою інформацією;

– враховувати при цьому наявні часові обмеження щодо аналізу даних та прийняття відповідних управлінських рішень (наприклад, про організацію протидії ідентифікованим загрозам тощо);

– забезпечувати «прозорість» висновків та рекомендацій, сформульовану комп'ютерною системою захисту для фахівців служби безпеки та осіб із статусом ОПР.

Література

1. єДокумент – новий тимчасовий цифровий документ на період воєнного стану. URL: https://diia.gov.ua/news/yedokument-novij-timchasovij-cifrovij-dokument-na-period-voyennogo-stanu (дата звернення: 10.04.2023)

2. Люфкин Б. Первое в мире цифровое правительство. *BBC News Україна*. URL: https://www.bbc.com/ukrainian/vert-fut-russian-41746784 (дата звернення: 10.04.2023)

3. Термінологічний словник з питань запобігання та протидії легалізації (відмиванню) доходів, одержаних злочинним шляхом, фінансуванню тероризму, фінансуванню розповсюдження зброї масового знищення та корупції / Чубенко А.Г., Лошицький М.В., Павлов Д.М., Бичкова С.С., Юнін О.С. Київ: Ваіте, 2018. 826 с.

4. Українці найбільше довіряють програмі «Держава у смартфоні» та цифровізації – опитування «Рейтинг». *Міністерство цифрової трансформації*

України. URL: https://thedigital.gov.ua/news/ukraintsi-naybilshe-doviryayut-programi-derzhava-u-smartfoni-ta-tsifrovizatsii-opituvannya-reyting (дата звернення: 10.04.2023)

5. Artificial Intelligence in International Development: A Discussion Paper. URL: https://www.idiainnovation.org/s/AIandinternationalDevelopment.pdf (дата звернення: 10.04.2023)

6. Cappelli D., Moore A., Trzeciak R. How to Prevent, Detect, and Respond to Information Technology Crimes (Theft, Sabotage, Fraud). Westford, Massachusetts: Pearson Education, Inc., 2012. 97 p.

7. Clifford C. Google CEO: A.I. is more important than fire or electricity. *CNBC.com*. URL: https://www.cnbc.com/2018/02/01/google-ceo-sundar-pichai-ai-is-more-important-than-fire-electricity.html (дата звернення: 10.04.2023)

8. CYBERscape.Momentumcyber.com.URL:https://momentumcyber.com/docs/CYBERscape_4.7.22.pdf (дата звернення: 10.04.2023)

9. Gogoi P., Bhattacharyya D.K., Borah B., Kalita J.K. A Survey of Outlier Detection Methods in Network Anomaly Identification. *The Computer Journal*. 2011. #54(4). PP. 570-588. URL: https://doi.org/10.1112/comjnl/bxr026 (дата звернення: 10.04.2023)

10. Han S., Hu X., Huang H., Jiang M., Zhao Y. ADBench: Anomaly Detection Benchmark. *NeurIPS 2022*. 2022. #45. URL: https://doi.org/10.48550/arXiv.2206.09426 (дата звернення: 10.04.2023)

11. Lai K.-H., Zha D., Wang G., Xu J., Zhao Y., Kumar D., ... Hu X. TODS: An Automated Time Series Outlier Detection System. 2021. URL: https://doi.org/10.48550/arXiv.2009.09822 (дата звернення: 10.04.2023)

12. Liu K., Dou Y., Zhao Y., Ding X., Hu X., Ding R.Z.K., ... Yu P.S. BOND: Benchmarking Unsupervised Outlier Node Detection on Static Attributed Graphs. *NeurIPS 2022*. 2022. URL: https://doi.org/10.48550/arXiv.2206.10071 (дата звернення: 10.04.2023)

13. Machine Learning Applications for Accounting Disclosure and Fraud Detection. Hershey, PA: IGI Global, 2021. 291 p.

14. Managing Machine Learning Projects in International Development: a Practical Guide. URL: https://www.usaid.gov/sites/default/files/2022-05/Vital_Wave_USAID-AIML-FieldGuide_FINAL_VERSION_1.pdf (дата звернення: 10.04.2023)

15. Paul A., Jolley C., Anthony A. Reflecting the Past, Shaping the Future: Making AI Work for International Development. URL: https://www.usaid.gov/sites/default/files/2022-05/AI-ML-in-Development.pdf (дата звернення: 10.04.2023)

16. Predd J., Pfleeger S.L., Hunker J., Bulford C. Insiders Behaving Badly. *IEEE Security & Privacy Magazine*. 2008. #6(4). PP. 66-70.

17. PyGOD. *Github.com*. URL: https://github.com/pygod-team/pygod (дата звернення: 10.04.2023)

18. Python Outlier Detection (PyOD). *Github.com*. URL: https://github.com/yzhao062/pyod (дата звернення: 10.04.2023)

19. Russell S.J., Norvig P. Artificial Intelligence: A Modern Approach: Pearson Education Limited, 2022. 1167 p.

20. Salem M.B., Hershkop S., Stolfo S.J. A Survey of Insider Attack Detection Research. *Insider Attack and Cyber Security*. 2009. PP. 69-90.

21. TODS: Automated Time-series Outlier Detection System. *Github.com*. URL: https://github.com/datamllab/tods (дата звернення: 10.04.2023)

22. Zhao Y., Nasrullah Z., Li Z. PyOD: A Python Toolbox for Scalable Outlier Detection. *Journal of Machine Learning Research*. #20. PP. 1-7. URL: https://doi.org/10.48550/arXiv.1901.01588 (дата звернення: 10.04.2023)

Анотація

Яцина Ю. О. **Особливості застосування інноваційних аналітико-статистичних технологій в інформаційних системах з моніторингу та виявлення корупції.** – Стаття.

Стаття присвячено проблемі визначення особливостей впровадження інноваційних аналітико-статистичних технологій в інформаційних системах з моніторингу та виявлення корупції. Визначено, що інноваційні аналітико-статистичні технології в широкому значенні є сукупністю методів та інструментів аналізу даних, які використовуються з метою пошуку аномалій в даних, виявлення корисних закономірностей в масивах даних, що може сприяти підвищенню якості управлінських рішень. У вузькому значенні під цими технологіями розуміються процес використання найсучасніших методів та технологій аналізу даних з метою виявлення складних залежностей та корисних закономірностей в даних. До таких технологій на сучасному етапі розвитку людства відносяться технології машинного навчання, глибинного навчання, нейронні мережі, обробки природної мови (NLP), аналіз графів, технології аналізу даних у режимі реального часу. Аналітико-статистичні технології отримують своє практичне втілення у формі інформаційної системи або платформи, що об'єднує в собі як технічні засоби (комп'ютери, засоби та канали зв'язку, периферійні пристрої, пристрої введення / виводу даних тощо), так і програмні застосунки, що забезпечують функціонування технічної та аналітичної складової. Інноваційними аналітико-статистичними технологіями, що можуть бути використані системах моніторингу та протидії корупції в державі можна вважати ті технології, що сфокусовані на пошуку ознак аномальної (зловмисної) активності користувачів (інсайдерів) у великих масивах даних, що відображають результати діяльності публічних (урядових) організацій, підприємств та установ. Наявність інсайдерів створює інсайдерські загрози – шкідливі для організації активності, які походять від співробітників всередині організації, зокрема – від діючих працівників, колишніх працівників, підрядників, ділових партнерів і навіть завербованих працівників або працівників, навмисно впроваджених в організацію, які мають доступ до конфіденційної інформації в рамках своїх посадових обов'язків та які мають уявлення про систему управління інформаційною безпекою організації. Сучасні підходи до виявлення інсайдерських загроз базуються на методах криміналістики (forensic) і зазвичай обмежують себе вивченням журналів кібербезпеки для застосування алгоритмів виявлення аномалій або

пошуку відбитків (сигнатур). Ці алгоритми необхідні для виявлення інсайдерської загрози, але вони лише є складовою повного рішення. Тому більш ефективними слід вважати інтегровані рішення з виявлення та запобігання інсайдерським загрозам, що використовують семантично марковані дані з інфраструктури кібернетичного та фізичного контролю доступу і забезпечують проактивне (упереджене) виявлення порушників (шахраїв, корупціонерів).

Ключові слова: протидія корупції, криміналістика, інсайдерська загроза, машинне навчання, штучний інтелект.

Summary

Yatsyna Yu. O. **Features of applying innovative analytical-statistical technologies in information systems for monitoring and detecting corruption.** – Article.

The article is dedicated to the problem of determining the features of implementing innovative analytical-statistical technologies in information systems for monitoring and detecting corruption. It is determined that innovative analytical-statistical technologies, in a broad sense, are a set of methods and tools for data analysis used for anomalies identification, useful patterns discovery in datasets that can contribute to improving the quality of decision-making process. In a narrow sense, these technologies involve the process of using the most advanced methods and technologies for data analysis to identify complex dependencies and useful patterns in the data. At the current stage of human development, such technologies include machine learning, deep learning, neural networks, natural language processing, graph analysis, and real-time data analysis technologies. Analytical-statistical technologies find their practical implementation in the form of an information system or platform that combines both technical means (computers, communication means and channels, peripheral devices, data input/output devices, etc.) and software applications that ensure the functioning of technical and analytical components. Innovative analytical-statistical technologies that can be used in systems for monitoring and countering corruption in the state are those technologies focused on searching for signs of abnormal (malicious) user activity (insiders) in large datasets which reflect the results of the activities of public (government) organizations, enterprises, and institutions. The presence of insiders creates insider threats – harmful activities that come from employees within the organization, including current employees, former employees, contractors, business partners, and even recruited employees or employees intentionally introduced into the organization who have access to confidential information as part of their job duties and who have an understanding of the organization's information security management system. Modern approaches to detecting insider threats are based on forensic methods and usually limit themselves to studying cybersecurity logs for the application of anomaly detection algorithms or searching for fingerprints (signatures). These algorithms are necessary for detecting insider threats, but they are only a component of a complete solution. Therefore, integrated solutions for detecting and preventing insider threats, which use semantically labeled data from cyber and physical access control infrastructure and provide proactive detection of offenders (fraudsters, corrupt officials), should be considered more effective.

Key words: corruption counteraction, forensic, insider threat, machine learning, artificial intelligence.