

ВИКОРИСТАННЯ BIG DATA ТА АЛГОРИТМІВ ШІ ЯК ІНСТРУМЕНТІВ АВТОМАТИЗОВАНОГО МОНІТОРИНГУ В СИСТЕМІ ЗАПОБІГАННЯ КОРУПЦІЇ

Вінницький національний технічний університет

Анотація

У дослідженні розглянуто проблеми та можливості впровадження технологій великих даних (Big Data) і машинного навчання (Machine Learning) для автоматизованого моніторингу корупційних ризиків. Особливу увагу приділено аналізу відкритих державних даних, які використовуються для створення предиктивних систем контролю. Описано практичну ефективність методів аналізу графів у виявленні прихованих зв'язків та алгоритмів детекції аномалій для ідентифікації підозрілих фінансових операцій. Наголошено, що впровадження передових аналітичних систем є ключовим фактором переходу від реактивного реагування на правопорушення до їх проактивного запобігання.

Ключові слова: Big Data, штучний інтелект, антикорупційний моніторинг, машинне навчання, відкриті реєстри.

Abstract

The study examines the problems and opportunities of implementing Big Data and Machine Learning technologies for automated monitoring of corruption risks. Particular attention is paid to the analysis of open government data used to create predictive control systems. The practical effectiveness of graph analysis methods in identifying hidden connections and anomaly detection algorithms for identifying suspicious financial transactions is described. It is emphasized that the implementation of advanced analytical systems is a key factor in the transition from reactive response to offenses to their proactive prevention.

Keywords: Big Data, artificial intelligence, anti-corruption monitoring, machine learning, open registries.

Вступ

У сучасному цифровому суспільстві ефективність запобігання та протидії корупції значною мірою залежить від здатності органів публічної влади швидко опрацювати та аналізувати великі обсяги інформації. Традиційні підходи, зокрема ручні аудити та вибіркові перевірки, втратили свою ефективність у контексті стрімкого збільшення даних у публічних електронних системах. Реактивна модель, що зосереджується на розслідуванні вже скоєних правопорушень, неспроможна забезпечити систематичну профілактику корупції. Це створює нагальну потребу в інтеграції інструментів автоматизованого предиктивного моніторингу, які ґрунтуються на використанні технологій Big Data та штучного інтелекту. Такі інструменти дають змогу постійно аналізувати цифрові сліди, виявляти тенденції та формувати попередження щодо ризиків ще на початкових стадіях їх виникнення. Варто зазначити, що зазначені вище відомості останнім часом є ключовим доказом вчинення корупційних правопорушень або порушень, пов'язаних із корупцією.

Результати дослідження

Основою автоматизованого антикорупційного моніторингу є консолідація розрізаних даних з відкритих державних реєстрів (ЄДР юридичних осіб, фізичних осіб-підприємців, громадських формувань, Державний реєстр речових прав на нерухоме майно, система публічних закупівель Prozorro, реєстр декларацій НАЗК та ін.) в єдине інформаційне середовище, а саме – «озеро даних» (Data Lake). Зважаючи на неоднорідність форматів, першочерговим завданням є налаштування ETL-процесів (Extract, Transform, Load) для очищення та структуривання інформації.

Ключові технологічні вектори застосування алгоритмів штучного інтелекту для завдань антикорупційного моніторингу:

1. Аналіз соціальних та фінансових графів (*Network/Graph Analysis*). Корупційні схеми рідко є прямолінійними; вони базуються на складних ланцюжках посередників, офшорних

юрисдикціях та підставних особах. Традиційні реляційні бази даних не здатні ефективно обробляти такі багатомірні запити. Розгляньмо використання граф-орієнтованих баз даних (наприклад, *Neo4j*). У такій моделі суб'єкти (посадовці, їхні родичі, юридичні особи) виступають вузлами, а їхні взаємодії (родинні зв'язки, спільне володіння майном, фінансові транзакції) – ребрами. Алгоритми обходу графів здатні за доли секунди виявити приховану афілійованість між посадовою особою, що формує тендерні вимоги, та кінцевим бенефіціаром компанії-переможця, автоматично фіксуючи конфлікт інтересів.

2. Детекція аномалій у масивах даних (*Anomaly Detection*). Для моніторингу публічних закупівель найдоцільніше застосовувати методи машинного навчання без вчителя (*Unsupervised Learning*), зокрема алгоритми типу *Isolation Forest*. Звернімо увагу на принцип їх роботи: моделі навчаються на масивах історичних даних про тисячі конкурентних і прозорих тендерів, формуючи математичний профіль «нормальної» закупівлі. Відхилення від цього профілю – наприклад, нетипово великий крок зниження ціни, участь компаній-клонів зі спільними IP-адресами, або ж штучний поділ одного великого тендеру на декілька допорогових сум для уникнення відкритих торгів – маркуються як «червоні прапорці» (*Red Flags*) і миттєво передаються на детальну перевірку аудиторам.
3. Інтелектуальний аналіз природної мови (*Natural Language Processing – NLP*). Значний масив корупційних ризиків приховано саме в неструктурованих текстових документах: технічних специфікаціях, договорах чи поясненнях до електронних декларацій. Детальніше розглянемо цей аспект: застосування сучасних мовних моделей та алгоритмів семантичного аналізу дозволяє автоматизувати пошук дискримінаційних вимог у тендерній документації. Система розпізнає випадки, коли технічні умови штучно прописуються під характеристики товару конкретного постачальника. Крім того, інструменти NLP ефективні для зіставлення офіційно задекларованих доходів із текстовими згадками про спосіб життя чиновника у відкритих джерелах.
4. Побудова систем предиктивного скорингу (*Predictive Risk Scoring*). Синтез вищезгаданих методів дозволяє перейти до оцінювання ризиків у режимі реального часу. Кожній транзакції, декларації чи тендеру алгоритм автоматично присвоює індекс ризикованості. Це забезпечує пріоритетизацію роботи антикорупційних органів: людський ресурс спрямовується виключно на об'єкти з найвищим показником скорингу.

Висновок

Інтеграція технологій великих даних та штучного інтелекту зазнає глибоких змін у сфері антикорупційного моніторингу, перетворюючи його з пасивного інструменту на активну систему підтримки прийняття управлінських рішень. Автоматизований початковий аналіз сприяє зниженню впливу суб'єктивних факторів, забезпечує комплексне охоплення інформації та формує психологічний ефект неминучості цифрового контролю. Для подальшого розвитку цього напрямку необхідним є підвищення якісних характеристик відкритих даних, а також постійне вдосконалення алгоритмів через навчання на нових типах економічних правопорушень.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Терещук М. М. Використання штучного інтелекту та аналітики для виявлення корупційних схем. *Антикорупційна система в Україні*: матеріали Всеукр. наук.-практ. конф. (м. Біла Церква, 2023). С. 21-23. URL: https://rep.btsau.edu.ua/bitstream/BNAU/10051/1/Vykorystannia_sh.pdf
2. Євтушенко А. І. Формування та реалізація антикорупційної політики в умовах глобалізації: дис. ... д-ра філософії. Харківський національний університет імені В. Н. Каразіна. Харків, 2023. 248 с. URL: https://karazin.ua/storage/static-content/source/documents/aspirantura/zakhysty/jevtushenko_a/%D0%94%D0%B8%D1%81%D0%B5%D1%80%D1%82%D0%B0%D1%86%D1%96%D1%8F%20%D0%84%D0%B2%D1%82%D1%83%D1%88%D0%B5%D0%BD%D0%BA%D0%BE.pdf
3. Пюнтковський Костянтин Анатолійович. ДОСВІД ВИКОРИСТАННЯ ІНФОРМАЦІЙНОКОМУНІКАЦІЙНИХ ТЕХНОЛОГІЙ В ЗАПОБІГАННІ ТА ПРОТИДІЇ КОРУПЦІЇ В УКРАЇНІ ТА СВІТІ: магістерська робота. Національний університет «Києво-Могилянська академія», Київ, 2023. 104 с. URL: <https://ekmair.ukma.edu.ua/server/api/core/bitstreams/7b348ab4-884d-43c9-addf-8bcca3bb682b/content>

Павловська Анастасія Вячеславівна – студентка групи ІБС-24Б, факультет інформаційних технологій та комп'ютерної інженерії. Вінницький національний технічний університет, Вінниця, email: nastyapav2006@gmail.com

Науковий керівник: **Ковальчук Іван Васильович** – старший викладач кафедри суспільно-політичних наук, факультету інформаційних електронних систем, Вінницький національний технічний університет, м. Вінниця, e-mail: ivankovalchuk@vntu.edu.ua

Pavlovska Anastasiia Vyacheslavovna – student of group IBS-24B, Faculty of Information Technologies and Computer Engineering. Vinnytsia National Technical University, Vinnytsia, email: nastyapav2006@gmail.com

Scientific supervisor: **Kovalchuk Ivan Vasilyevich** – senior lecturer of the Department of Social and Political Sciences, Faculty of Information Electronic Systems, Vinnytsia National Technical University, Vinnytsia, email: ivankovalchuk@vntu.edu.ua

ORCID: 0009-0007-74675317