

МЕТОД ВИЯВЛЕННЯ АНОМАЛІЙ У ФІНАНСОВИХ ЗВІТАХ

Вінницький національний технічний університет

Анотація

Запропоновано метод виявлення аномалій у фінансових звітах на основі алгоритму машинного навчання Isolation Forest, який дозволяє підвищити точність ідентифікації потенційного шахрайства, помилок обліку та нетипових фінансових операцій. Досліджено вплив параметрів алгоритму на ефективність виявлення аномалій у різних типах фінансових даних.

Ключові слова: аномалії, фінансова звітність, штучний інтелект, Isolation Forest, виявлення шахрайства, фінансовий аналіз.

Abstract

A method for detecting anomalies in financial reports based on the Isolation Forest machine learning algorithm is proposed, which allows to increase the accuracy of identifying potential fraud, accounting errors and atypical financial transactions. The influence of the algorithm parameters on the efficiency of detecting anomalies in different types of financial data is studied.

Keywords: anomalies, financial reporting, artificial intelligence, Isolation Forest, fraud detection, financial analysis.

Вступ

На сьогоднішній день проблема виявлення аномалій у фінансових звітах досі лишається актуальною. Це питання постає гостро як перед регуляторами, так і перед керівництвом компаній та аудиторів [1]. Традиційні методи виявлення аномалій вже не дають бажаної точності та мають обмежену ефективність, оскільки базуються на наборі заздалегідь визначених правил та порогових значень. Це особливо помітно на наборі великої кількості даних та при виявленні аномалій нестандартних типів [2].

Глобальна цифровізація усіх рівнів людської діяльності вимагає впровадження інформаційних технологій інтелектуального аналізу даних, що дозволить підвищити рівень ідентифікації шахрайства з персональними даними. Тому актуальним є використання методів на базі штучного інтелекту, які дозволять точніше визначати різні види аномальних операцій, що свідчить про наявність шахрайських операцій, та здійснювати ефективну обробку великого об'єму фінансових даних.

Метою роботи є розробка методу виявлення аномалій у фінансових звітах з використанням алгоритму Isolation Forest та оцінка його ефективності порівняно з традиційними методами.

Результати дослідження

Алгоритм Isolation Forest базується на принципі, що аномалії легше ізолювати від нормальних даних завдяки їх відмінним характеристикам, тобто при розподілі дерева, ми отримаємо менший шлях до аномалій, що допоможе їх відокремити від загальних даних [3]. Серед переваг даного методу, можна виділити те, що дані не потребують попередньої обробки перед початком аналізу [4].

Метод Isolation Forest для виявлення аномалій можна описати функцією показника аномальності:

$$S(x) = 2^{\frac{-E(h(x))}{c(n)}}, \quad (1)$$

де $h(x)$ – функція середньої глибини ізоляційних даних x , $E(h(x))$ – функція очікуваної глибини, $c(n)$ – нормалізуючий коефіцієнт для набору даних розміром n .

Також для визначення системи рейтингу аномалій за ступенем потенційного ризику було введено функцію ризику:

$$R = S(x) \cdot I \cdot P, \quad (2)$$

де I – оцінка впливу (вагомість показника у фінансовій звітності), P – ймовірність того, що аномалія

не є випадковою похибкою даних.

У ході дослідження було проаналізовано набір фінансових звітів 50 компаній за період 2019 - 2023 років. Масив даних за цей період містить як звичайні транзакції, так і заздалегідь ідентифіковані аномалії. Основні фінансові показники, які аналізувалися, включають об'єми торгів, суму дивідендів та денну прибутковість.

Найбільшу ефективність алгоритм Isolation Forest показав при аналізі таких типів аномалій:

- необґрунтовані коливання в обсягах продажів між послідовними звітними періодами;
- нетипові значення у співвідношеннях між статтями балансу;
- порушення встановлених фінансових закономірностей;
- розбіжності між грошовими потоками та звітами про прибутки і збитки.

Дослідження показало, що продуктивність алгоритму залежить також і від набору вхідних параметрів. При включенні додаткових параметрів таких як врахування сезонності, волатильності та темпів росту, ефективність та точність виявлення аномалій зростає.

Експериментально визначено, що оптимальними параметрами для аналізу фінансових звітностей є використання 100-200 дерев та порогове значення аномальності – 0,2.

Важливим аспектом дослідження є інтерпретація виявлених аномалій. Для цього було розроблено метод графічного представлення аномальних даних у вигляді відповідних точок та графіки (рис. 1), що відображають відхилення по різних параметрах, що були проаналізовані.

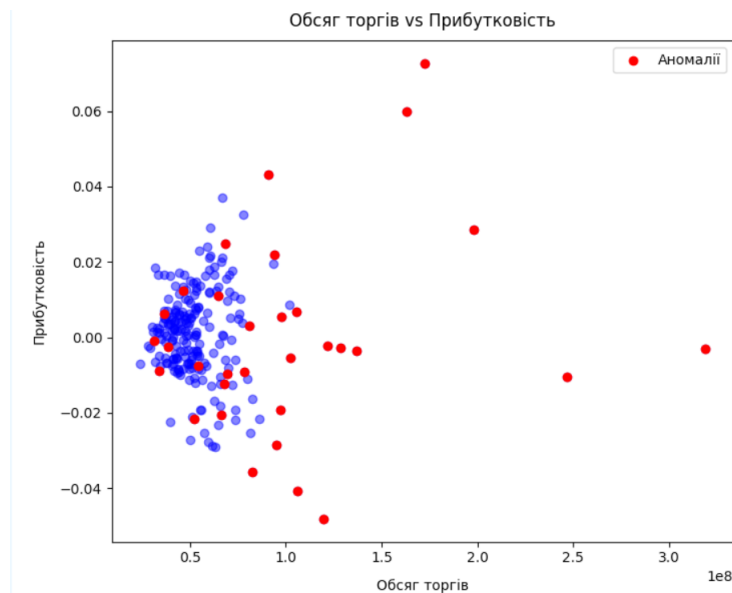


Рис. 1. Визначені аномалії в обсягах торгів та розподілі прибутковості

На рисунку 1 представлений результат виявлення аномалій для даних обсягів торгів та прибутковості. Аномалії виділені червоним кольором, також чітко видно, що самі аномалії досить розкидані і не згуртовані з загальними даними. Це дозволяє аналітикам краще розуміти природу виявлених відхилень та полегшує подальший їхнє вивчення.

Висновки

Встановлено, що запропонований підхід дозволяє підвищити загальну точність виявлення аномалій у фінансових звітах порівняно з традиційними статичними методами, при цьому забезпечує меншу кількість хибних напрацювань, що задовольняє потреби як аудиторів, так і регуляторного контролю.

Також, використання даного методу дозволяє виявляти й не очевидні аномалії, які проявляються у взаємодії декількох фінансових показників, що є важливими в сучасних реаліях ведення бізнесу.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Abdul Kader, L. Anomaly detection in financial transaction time series data [Master's thesis, Uppsala University]. - Uppsala University, 2023. - 56 с.

2. М. В. Дубініна, І. С. Норова. Особливості виявлення та попередження помилок у фінансовій звітності: стаття Миколаїв, 2013р. - 6 с.
3. Alexander Bakumenko Detecting anomalies in financial data using Machine Learning - Luleå University of Technology Department of Computer Science, Electrical and Space Engineering, 2022. - 68 с.
4. What is Isolation Forest? [Електронний ресурс] – Режим доступу: <https://www.geeksforgeeks.org/what-is-isolation-forest/>.

Лановик Анастасія Сергіївна – студента групи КН-21Б, факультет інформаційних інтелектуальних технологій та автоматизації, Вінницький технічний національний університет, Вінниця

Науковий керівник: **Іванчук Ярослав Володимирович** – д-р техн. наук, професор кафедри комп'ютерних наук, Вінницький національний технічний університет, м. Вінниця

Lanovyk Anastasia Serhiivna – Department of Information Intellectual Technologies and Automation, Vinnytsia National Technical University, Vinnytsia

Supervisor: Ivanchuk Yaroslav V. — Dr. Sc. (Eng.), Professor, Professor of the Department of Computer Sciences, Vinnytsia National Technical University, Vinnytsia