УДК 004.85:336.76

ЕКСПЕРИМЕНТАЛЬНІ ДОСЛІДЖЕННЯ ІНТЕЛЕКТУАЛЬНОЇ ІНФОРМАЦІЙНОЇ ТЕХНОЛОГІЇ ВИЯВЛЕННЯ ТА АНАЛІЗУ АНОМАЛІЙ У ФІНАНСОВИХ ЗВІТАХ

Яровий А.А., д.т.н., проф.[1]; Лановик А.С.[2]
[1,2] Вінницький національний технічний університет
[1,2] Україна, Вінниця
[1] a.yarovyy@vntu.edu.ua; [2] anastasialanovuk102@gmail.com

Анотація. *Представлено експериментальну перевірку модуля виявлення аномалій інтелектуальної інформаційної технології виявлення та аналізу аномалій у фінансових звітах. Початковий підхід на основі усереднення оцінок виявляв лише 25 % аномалій. Після переходу на процентильні ранги та щоденні показники помилок LSTM площа під кривою AP для набору даних становить 0,72, порівняно з 0,59 для алгоритму Isolation Forest. Додано умову валідності, – день позначається лише в тому випадку, якщо його оцінка суттєво відхиляється від норми (робастна z-оцінка, більша за 2,0). У ряді без аномалій це зменшує частку позначених днів з 10 % до 3,4 %.*

Ключові слова: *інтелектуальна інформаційна технологія; машинне навчання; аномалії; фінансові часові ряди; ансамбль детекторів; Isolation Forest; автоенкодер LSTM.*

Вступна частина. Аналіз фінансової звітності є одним із ключових інструментів забезпечення прозорості та довіри на фінансових ринках. Водночас зростання обсягу операцій і складності фінансових інструментів призводить до недостатньої ефективності ручного аудиту. Тому, автоматизація процесу виявлення аномалій із використанням методів штучного інтелекту є актуальним завданням. Існуючі комерційні системи (IBM Cognos Analytics, Oracle Financial Services) мають значні обмеження: вони аналізують ринкові дані або фінансові звіти окремо, використовують статичні порогові правила без адаптації і лише частково пояснюють характер виявлених розбіжностей.

У попередній роботі [1] обґрунтовано структурно-функціональну організацію інтелектуальної інформаційної технології для виявлення і аналізу аномалій та описано її архітектуру, зокрема, алгоритм Isolation Forest [2] виявляє відхилення в окремі моменти часу, автоенкодер LSTM [3] виявляє часові аномалії, алгоритм DBSCAN [4] кластеризує виявлені аномалії, а нечітка експертна система, оснащена базою знань, класифікує їхній тип за участю аналітика. Ця робота зосереджується виключно на першому етапі цього конвеєра: модуль виявлення, який об’єднує результати двох детекторів. Спочатку результати детекторів об’єднувалися за допомогою зваженої суми, а ефективність вказаного модуля оцінювалася лише теоретично.

Метою дослідження є експериментальні дослідження особливостей застосування ансамблевого підходу для удосконалення багатоетапного процесу виявлення та аналізу аномалій у фінансових звітах.

Основна частина. Методика експериментів. Основний експеримент було проведено на синтетичному ряді з 500 торгових днів (після обчислення 17 ознак залишається 481 день), у який було додано приблизно 10% аномальних днів (47 із 481) чотирьох типів (стрибок ціни, сплеск обсягу, розрив, тривалий рух) із відомою розміткою. Синтетичні дані було обрано тому, що для реальних котирувань відсутня розмітка аномальних днів [5], а без цього неможливо

обчислити частку правильно виявлених аномалій. Ряд обробляється з використанням тих самих модулів для обчислення 17 ознак та тих самих детекторів, що й для реальних котирувань. Дні, позначені системою, порівнювали з розміткою, і обчислювали Precision – частку позначених днів, які насправді є аномальними:

$$Precision = TP/(TP + FP), \quad (1)$$

де TP – відповідає справжнім аномаліям, позначеним системою, а FP – нормальним дням, позначеним системою. Потім обчислювали частку виявлених фактичних аномалій, як Recall:

$$Recall = TP/(TP + FN), \quad (2)$$

де FN позначає пропущені аномалії.

Також було обраховано гармонійне середнє цих двох показників, як F1:

$$F1 = 2 \cdot Precision \cdot Recall/(Precision + Recall) \quad (3)$$

Ці показники залежать від порогового значення θ, тому також було обчислено середню точність (AP) – площу під кривою «точність-повнота». Вона будується шляхом ранжування днів за рівнем аномалії та обчислення точності й повноти для кожної підмножини перших кількох днів. AP не залежить від порогового значення і для випадкового детектора дорівнює частці аномалій (приблизно 0,1). Крім того, набори даних було згенеровано з різними частками справжніх аномалій 0%, 5%, 10% та 19%. Кожен експеримент повторювали п’ять разів із різними початковими станами генераторів випадкових чисел, від яких залежать як згенерована послідовність, так і навчання LSTM; результати подано у вигляді середнього значення із врахуванням стандартного відхилення.

Виявлення аномалій. Початкова конфігурація виявляла лише 25 % аномалій (Precision = 0,60, Recall = 0,25). В ході експериментальних досліджень було виявлено три причини. По-перше, оцінки детекторів мали різні шкали (медіанна оцінка для автоенкодера LSTM – 0,019 проти 0,173 для Isolation Forest). При такому поєднанні, детектор із вищою шкалою приглушує інший. Саме тому було обрано підхід – поєднувати їх не на основі їхніх значень, а відповідно до їхніх процентильних рангів. Ранг вказує на позицію дня за рівнем аномальності серед усіх днів у ряді та не залежить від шкали оцінювання [6]. По-друге, оцінка LSTM розраховувалася як середня похибка реконструкції для всіх вікон, що містять даний день. Отже, день з відхиленням некоректно впливав на оцінки сусідніх днів. Розрахунок похибки для самого дня підвищив показник AP автоенкодера з 0,30 до 0,80. По-третє, детектори можна правильно порівнювати лише за умови, що частка позначених днів є однаковою; інакше вища повнота просто є наслідком більшої кількості позначених днів. Саме тому алгоритм тепер обраховує зважену суму рангів і позначає дні, ранг яких перевищує θ, так щоб частка позначених днів дорівнювала 1 – θ. Поріг θ визначає навантаження на аналітика (θ = 0,9 означає перевірку 10 % днів). У таблиці 1 наведено результати для θ = 0,9 без будь-яких умов щодо z-оцінки, щоб детектори можна було порівнювати за рівних навантажень. Залежність точності від повноти показано на рис. 1.

Таблиця 1 – Показники виявлення аномалій на синтетичних даних (θ = 0,9, 10%, 5 незалежних прогонів)

Детектор	Precision	Recall	F1	AP
Isolation Forest	0,562 ± 0,035	0,574 ± 0,032	0,568 ± 0,033	0,592 ± 0,023
Ансамбль	0,633 ± 0,013	0,660 ± 0,019	0,646 ± 0,016	0,723 ± 0,028

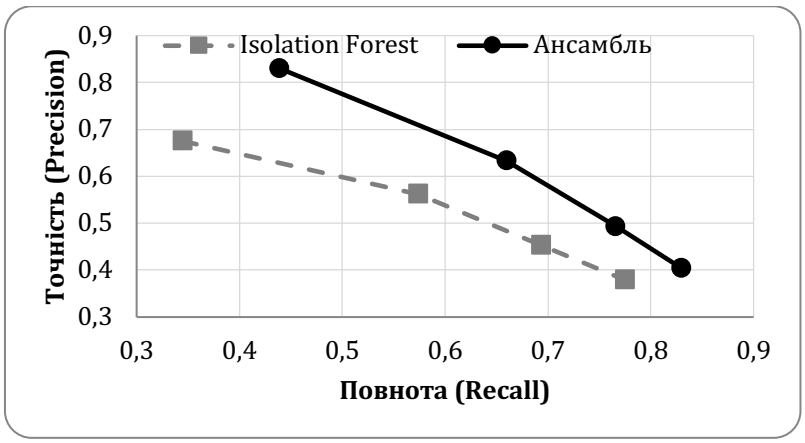


Рисунок 1 – Залежність точності від повноти за порогів θ = 0,80–0,95

При однаковому навантаженні ансамбль перевершує Isolation Forest за всіма показниками (AP 0,72 проти 0,59; F1 0,65 проти 0,57), а для θ = 0,8 його повнота становить 0,83 проти 0,78. Однак, на синтетичних даних автоенкодер LSTM окремо демонструє дещо кращі результати, ніж ансамбль (AP 0,8). Оскільки реальні дані не містять міток, неможливо заздалегідь визначити кращий детектор.

Адаптація до частки аномалій. Ранг лише впорядковує дні; отже, у часовому ряді без аномалій критерій рангу в будь-якому разі визначив би ті самі 10 % днів. Щоб уникнути цього, було додано другу умову: доказовість. День позначається лише тоді, коли його робастна оцінка z (відхилення від медіани в одиницях MAD, усереднене за обома детекторами) перевищує 2,0. Як показано в таблиці 2, у ряді без аномалій позначається 3,4 % днів замість 10,1 %, тоді як для 5- та 10-відсоткових аномалій частка позначених днів (8,3 та 9,8 %) близька до фактичної частки. У випадку 19 % аномалій система позначає лише 8,4 % днів, при цьому показник повноти становить 0,39. Причина в порозі доказовості, а не в θ: для θ = 0,75 (до 25 % днів) та порогу 2,0 показник повноти для такого ряду становить лише 0,4. Зниження порогу доказовості до 1,0 збільшує показник повноти до 0,67, але в рядах з незначним відсотком аномалій тоді позначається 15,6% днів замість 3,4 %. Таким чином, поріг доказовості забезпечує баланс між повнотою в рядах, що містять багато аномалій, та кількістю помилкових спрацьовувань у рядах без аномалій.

Таблиця 2 – Частка днів, позначених як аномальні, залежно від частки фактичних аномалій (θ = 0,9, 5 прогонів)

Справжніх аномалій, %	Лише ранг: позначено, %	Повна система: позначено, %	Precision	Recall	F1
0	10,1	3,4	0,00	–	–
4,9	10,1	8,3	0,44	0,71	0,54
9,8	10,2	9,8	0,64	0,64	0,64
19,4	10,1	8,4	0,89	0,39	0,54

Обмеження. Усі показники були отримані на основі синтетичних даних, а для перевірки відповідності результатів фактичним цінам необхідні розмічені дані. Поріг достовірності 2,0 було обрано на основі синтетичного ряду, вільного від аномалій. Отже, його слід калібрувати, використовуючи період стабільної динаміки цін з історичних даних щодо реальних активів. Обидва пороги мають свої недоліки. Верхня межа навантаження 1 − θ гарантує, що аналітик не переглядатиме більше ніж 10 % днів (θ = 0,9), але якщо фактичних аномалій буде більше,

деякі з них будуть пропущені: для 19,4 % аномалій повнота не може перевищувати $10/19{,}4 \approx 0{,}52$ (результат становить 0,39). Поріг доказовості при значенні 2,0 зменшує хибні спрацювання, але знижує повноту в рядах з багатьма аномаліями. Тим часом поріг 1,0 діє навпаки. Іншими словами, менше навантаження на аналітика та менша кількість помилкових позначок досягаються ціною пропуску деяких аномалій.

Крім того, модель демонструє низьку ефективність у виявленні тривалих рухів (trend_break, три дні поспіль із щоденним зсувом ціни на 4–7 %). Повнота становить лише 0,03 для $\theta = 0{,}90$ та 0,38 для $\theta = 0{,}80$, тоді як повнота виявлення цінових стрибків, розривів та сплесків обсягу для $\theta = 0{,}90$ коливається від 0,75 до 1,00.

Висновки. Експериментальна перевірка показала, що початкова схема об'єднання оцінок була не надто ефективною (Recall = 0,25). Перехід на процентильні ранги, щоденний розрахунок помилки LSTM та повторне ранжування суми підвищили показник *AP* до 0,72 порівняно з 0,59 для Isolation Forest при однаковому навантажені. Додатково поріг θ встановлює верхню межу навантаження на аналітика, тоді як умова, пов'язана зі робастною z-оцінкою, запобігає позначенню днів у ряді з незначною кількістю аномалій (3,4 % замість 10,1 %). Подальші дослідження будуть зосереджені на розмічених реальних даних, а також на аналізі решти модулів та їхньому вдосконаленні.

Використання інструментів ШІ. Не використовувались.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Лановик, А. С., & Яровий, А. А. (2026). *Багатокомпонентна система виявлення та аналізу аномалій у фінансових звітах на основі методів машинного навчання*, Молодь в науці: дослідження, проблеми, перспективи (МН-2026). Вінниця: ВНТУ. Взято з https://conferences.vntu.edu.ua/index.php/mn/mn2026/paper/view/30153

[2] Liu, F. T., Ting, K. M., & Zhou, Z.-H. (2008). Isolation Forest. In *Proceedings of the 8th IEEE International Conference on Data Mining* (pp. 413-422). IEEE. https://doi.org/10.1109/ICDM.2008.17

[3] Malhotra, P., Ramakrishnan, A., Anand, G., Vig, L., Agarwal, P., & Shroff, G. (2016). *LSTM-based encoder-decoder for multi-sensor anomaly detection*, ICML 2016 Anomaly Detection Workshop. New York, NY. Retrieved from https://doi.org/10.48550/arXiv.1607.00148

[4] Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *In Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining (KDD-96)* (pp. 226-231). Portland, OR: AAAI Press.

[5] Aggarwal, C. C., & Sathe, S. (2017). *Outlier ensembles: An introduction.* Cham, Switzerland: Springer. https://doi.org/10.1007/978-3-319-54765-7

[6] Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys, 41(3), Article 15.* https://doi.org/10.1145/1541880.1541882

Yarovyi A.A., Lanovyk A.S.

Experimental research on intelligent information technology for the detection and analysis of anomalies in financial statements

Abstract. *This paper presents an experimental evaluation of an anomaly detection module based on artificial intelligence technology for detecting and analyzing anomalies in financial statements. The initial approach, based on averaging scores, detected only 25% of anomalies. After switching to percentile ranks and daily LSTM error metrics, the area under the AP curve for the dataset is 0.72, compared to 0.59 for the Isolation Forest algorithm. A validity condition was added: a day is flagged only if its score deviates significantly from the norm (a* robust *z-score greater than 2.0). In a series without anomalies, this reduces the proportion of flagged days from 10% to 3.4%.*

Keywords: *intelligent information technology; machine learning; anomalies; financial time series; ensemble of detectors; Isolation Forest; LSTM autoencoder.*