

УДК 004.65:004.032.2

*Бабарицький Д. О., здобувач,  
Білова Т. Г., к.т.н., доцент*

*Харківський національний університет радіоелектроніки*

## **ФОРМУВАННЯ СИСТЕМИ АНАЛІТИЧНИХ ПОКАЗНИКІВ ДЛЯ ВИЯВЛЕННЯ АНОМАЛІЙ У ПУБЛІЧНИХ ЗАКУПІВЛЯХ**

Ефективність цифрових публічних закупівель безпосередньо залежить від своєчасного виявлення ризикових тендерів та системного проведення моніторингу [1]. Головною проблемою автоматизації аналізу тендерів у системі публічних закупівель «Prozorro» залишається різноманітність даних: поєднання значного обсягу неструктурованих текстових описів із кількісними ознаками, що унеможливує пряме алгоритмічне порівняння лотів і ускладнює пошук аномалій. Також ідентичні товари в різних тендерах можуть відображатися як унікальні об'єкти, що приховує штучне завищення цін.

Для вирішення проблеми запропоновано ієрархічну систему показників та процедуру їх динамічної агрегації для приведення даних до спільного математичного простору. Завдяки цьому пошук прихованих цінових аномалій відбувається всередині автоматично сформованих однорідних за описом підгруп, що дозволяє порівнювати однакові за предметом закупівель тендери.

Схему ієрархічної структури показників та процесу їх обробки представлено на рисунку 1.



Рис.1. Ієрархічна структура аналітичних показників та процедура їх агрегації для виявлення аномалій

Згідно з принципами моніторингу складних інформаційних систем [2], розроблена система будується за ієрархічним принципом і передбачає динамічне агрегування даних з нижчого рівня управління на вищий.

На початковому етапі збору інформації через API «Prozorro» формується масив базових показників: фактичний обсяг, очікувана вартість, наявність ПДВ та місяць оголошення. Далі ці кількісні метрики агрегуються, утворюючи похідні показники, зокрема, ціну за одиницю товару. З метою усунення статистичного шуму та зменшення впливу викидів застосовується логарифмічне згладжування цих показників.

Проблема неструктурованих описів вирішується шляхом введення групи семантичних показників. Текстові дані проходять очищення, лематизацію та векторизуються за алгоритмом TF-IDF. Оскільки кількість підгруп наперед невідома, для агрегації семантичних показників у вищі класифікаційні групи використовується алгоритм K-Means. Він автоматично розбиває векторний простір на підгрупи, де оптимальна кількість категорій визначається динамічно за критерієм максимізації коефіцієнта силуєту.

Процес перетворення показників повинен контролювати втрату інформації. Для цього до багатовимірного простору похідних та семантичних показників застосовується метод головних компонент. Він дозволяє знизити розмірність даних із мінімізацією втрати дисперсії.

На фінальному етапі прийняття рішень щодо ризиковості тендеру розгортається щільніший алгоритм DBSCAN. Він оцінює агреговані показники всередині кожної категорії, маркуючи як аномальні ті лоти, що демонструють нетипове співвідношення показників.

Отже, розроблена ієрархічна система аналітичних показників дозволяє формалізувати процес моніторингу закупівель. Застосування алгоритмів NLP, K-Means, PCA та DBSCAN для агрегації аналітичних показників дозволяє контролювати втрату інформації та автоматично дозволяється виявляти приховані цінові відхилення, формуючи обґрунтовані рішення для аудиторів.

#### **Список використаних джерел:**

1. Fazekas M., Tóth I. J., King L. P. An objective corruption risk index using public procurement data // *European Journal on Criminal Policy and Research*. 2016. Vol. 22, No. 3. P. 369–397. DOI: <https://doi.org/10.1007/s10610-016-9308-z>
2. Білова Т. Г., Ярута В. О., Побіженко І. О. Формування ієрархічної структури показників моніторингу складної інформаційної системи // *Системи обробки інформації*. 2015. № 12(137). С. 78–80.