

Олег Шубалий

Аспірант кафедри комп'ютерних наук
Тернопільський національний технічний університет імені Івана Пулюя,
м. Тернопіль, Україна
ORCID 0009-0003-2345-9148
oleh_shubalyi1606@tntu.edu.ua

Леся Дмитроца

К.т.н, доцент, доцент кафедри комп'ютерних наук
Тернопільський національний технічний університет імені Івана Пулюя,
кафедра комп'ютерних наук, м. Тернопіль, Україна
ORCID 0000-0003-2583-3271
dmytrotsa.lesya@gmail.com

МЕТОДОЛОГІЧНИЙ АНАЛІЗ ПІДХОДІВ В АНАЛІТИЦІ BIG DATA

Анотація. Здійснено порівняльний аналіз методів провадження аналітики великих даних. Запропоновано створення комплексної методології аналітики Big Data із застосуванням кращих підходів та їх поєднанням.

Ключові слова: Big Data; аналітика; методологія.

1. ВСТУП

Проблематика великих даних включає у себе не лише питання інфраструктури та технологій розподіленої обробки, але й, очевидно, як найважливіший аспект, їх ефективний аналітичний аналіз. По при сучасні тенденції всебілшого поширення методів машинного навчання, класичні статистичні, емпіричні та сигнатурні методи можуть мати перевагу над ними в аналітиці Big Data в залежності від контексту задач та вимог.

Постановка проблеми. Систематизація підходів та визначення кращих шляхів провадження аналітики великих даних є актуальною задачею як для наукових досліджень, так і для практичного застосування.

Аналіз останніх досліджень і публікацій. Одним з найбільш відомих дослідників, який звертає увагу на непрозорість машинного навчання на протидію статистичній обґрунтованості у Big Data є професор Bin Yu з університету Берклі. Зокрема в [1] ним запропоновано концепцію PCS: Predictability, Computability, Stability як основу для надійного аналізу даних. Важливим та актуальним для розуміння класичних методів аналітики є джерело [2]. Детальний виклад аналітики великих даних подано у [3]. Актуальна аналітика Big Data засобами машинного навчання розкрита в [4].

Мета публікації. Метою публікації є розробка нових підходів до формування комплексної методології аналітики великих даних, яка поєднує класичні методи та методи штучного інтелекту.

2. РЕЗУЛЬТАТИ ТА ОБГОВОРЕННЯ

В реальних сферах застосувань до аналітики Big Data ставляться вимоги прозорості та контролю. Важливу роль відіграє й безпека чутливих даних. Тому застосування методів машинного навчання для аналітики при таких обмеженнях може бути недопустиме. Окрім того, класичні статистичні, евристичні чи сигнатурні підходи більш ефективніші в задачах потокової обробки реального часу, оскільки споживають значно менше обчислювальних ресурсів ніж технології штучного інтелекту. Через непрозорість алгоритмів, особливо у глибоких моделях, інтерпритованість результатів машинного навчання може бути низькою і, відповідно, неприйнятною для аудиту. Крім

того, методи штучного інтелекту мають критичну залежність від якості навчальних даних, яка також впливає і стабільність результатів.

В свою чергу, класичні методи аналітики поступаються машинному навчанню у гнучкості і є менш ефективними при високій варіативності даних у Big Data та складних нелінійних залежностях. Нейромережі здатні самостійно виявляти закономірності в даних без попереднього знання їх структури, в той час як для сигнатурних та евристичних методів необхідне глибоке розуміння домену.

Більш наочний порівняльний аналіз класичних методів аналітики Big Data із методами машинного навчання наведено у таблиці 1.

Таблиця 1

Порівняння методів аналітики Big Data

Критерій порівняння	Класичні методи	Методи машинного навчання
Прозорість та контрольованість	Висока	Низька
Споживання ресурсів	Помірне	Високе
Робота в режимі реального часу	Висока придатність	Менша придатність
Гнучкість та адаптивність	Обмежена	Висока
Потреба у навчальних даних	Відсутня	Наявна
Стабільність	Висока	Визначається якістю навчальних даних
Обробка даних складної структури	Обмежена	Ефективна при якісних навчальних даних

Хоча кожен із описаних методологічних підходів аналітики великих даних має свої недоліки і вибір для застосувань у практичній сфері визначається накладеними вимогами, однак перспективним є розвиток гібридних методологій, що використовуватимуть переваги кожного із підходів.

При цьому можливі варіанти як послідовного застосування класичних методів та методів штучного інтелекту у різних комбінаціях, так і реалізації паралельних підходів. Ще одним можливим варіантом гібридних методологій є використання класичних методів для пояснення результатів, отриманих технологіями штучного інтелекту. До прикладу, для пояснення рішень, отриманих ШІ, вже на основі статистичних або евристичних оцінок важливості ознак.

У випадку каскадної (послідовної) інтеграції можливі варіанти, коли одні з методів використовуватимуться для попередньої обробки великих даних, а подальший аналіз здійснюватиметься іншими методами. Так, якщо спочатку використати для очистки та підготовки Big Data класичні методи (зокрема, статистичні – для нормалізації та зменшення розмірності, а евристичні – для фільтрації за заданими правилами та пороговими значеннями), то завдяки покращеній якості даних підвищиться точність та машинного навчання і, відповідно якість результатів ШІ. Якщо ж навпаки для попередньої обробки Big Data вибрати нейронні мережі, то їх можна застосувати щоб виділити ознаки у високо варіативних даних із складними залежностями, які в подальшому можна буде проаналізувати класичними статистичними методами.

Різні варіанти паралельного (ансамблевого) поєднання класичних методів та методів ШІ можуть, до прикладу, застосовуватися для гібридного виявлення аномалій

(коли сигнатурними методами виявляють відомі типи збоїв, на яких навчатиметься модель), або у стекінгу (коли класичними статистичними методами та методами машинного навчання генеруються окремі прогнози, які потім об'єднуються моделлю ШІ, що виступає мета-класифікатором).

Основні переваги кожного з наведених варіантів гібридної методології згруповано у таблиці 2.

Таблиця 2

Основні переваги гібридної методології в аналітиці Big Data

Види гібридної методології	Деталі методології	Переваги
Послідовна (каскадна)	Попередня обробка Big Data класичними методами, машинне навчання на якісних даних	Підвищення точності та швидкості ML
	Виділення складних варіативних ознак із сирих даних, фінальний аналіз класичними методами	Використання потужностей ШІ для аналізу класичними методами
Паралельна (ансамблева)	Стекінг шляхом здійснення базового прогнозу класичними та машинними методами, фінальний прогноз мета-класифікатором ШІ	Максимальна надійність прогнозних результатів
	Гібридне виявлення аномалій сигнатурними методами та методами ШІ	Швидкість та гнучкість у виявленні загроз та критичних відхилень
Інтерпретаційна післяобробка	Кількісна оцінка важливості ознак та факторів класичними методами за результатами ШІ-моделей	Збільшення довіри до результатів ШІ та їх прийнятність для аудиту

ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

Виконано методологічний аналіз підходів в аналітиці Big Data. Відображено переваги та недоліки кожного з них. Запропоновано створення комплексної методології аналітики великих даних на основі гібридних підходів, які будуть розвиватися у наступних дослідженнях.

ПОСИЛАННЯ

1. B. Yu and R. L. Barter, *Veridical Data Science: The Practice of Responsible Data Analysis and Decision Making*. Cambridge, MA: The MIT Press, 2024.
2. G. James, D. Witten, T. Hastie, and R. Tibshirani, *An Introduction to Statistical Learning: with Applications in R*, 2nd ed. Stanford University, 2023. [Online]. Available: https://hastie.su.domains/ISLR2/ISLRv2_corrected_June_2023.pdf.download.html. [Accessed: Nov. 9, 2025].
3. J. Leskovec, A. Rajaraman, and J. D. Ullman, *Mining of Massive Datasets*, 3rd ed. Cambridge: Cambridge University Press, 2020.
4. K. Banachewicz and L. Massaron, *The Kaggle Book: Data Analysis and Machine Learning for Competitive Data Science*. Birmingham: Packt Publishing, 2022.

MINISTRY OF EDUCATION
AND SCIENCE OF UKRAINE

NATIONAL UNIVERSITY
OF LIFE AND ENVIRONMENTAL
SCIENCES OF UKRAINE

FACULTY OF INFORMATION
TECHNOLOGY

МІНІСТЕРСТВО ОСВІТИ
І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ
БІОРЕСУРСІВ І
ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ
ТЕХНОЛОГІЙ

PROCEEDINGS

XIII International scientific
and practical conference

**GLOBAL AND
REGIONAL PROBLEMS OF
INFORMATIZATION IN
SOCIETY AND
NATURE USING
'2025**

13-14 November 2025

Kyiv, NULES of Ukraine

Kyiv 2025

МАТЕРІАЛИ

XIII Міжнародної науково-
практичної конференції

**ГЛОБАЛЬНІ ТА
РЕГІОНАЛЬНІ ПРОБЛЕМИ
ІНФОРМАТИЗАЦІЇ В
СУСПІЛЬСТВІ І
ПРИРОДОКОРИСТУВАННІ
'2025**

13-14 листопада 2025 року

Київ, НУБіП України

Київ 2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ
І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ
ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

МАТЕРІАЛИ

ХІІІ Міжнародної науково-практичної конференції

ГЛОБАЛЬНІ ТА РЕГІОНАЛЬНІ ПРОБЛЕМИ ІНФОРМАТИЗАЦІЇ В СУСПІЛЬСТВІ І ПРИРОДОКОРИСТУВАННІ '2025

13-14 листопада 2025 року

Київ, НУБіП України

Київ 2025

УДК 004

Рекомендовано до друку вченою радою факультету інформаційних технологій Національного університету біоресурсів і природокористування України (протокол № 4 від 18.12.2025).

Укладач: д.т.н., доцент Шкарупило В.В.

Збірник матеріалів XIII Міжнародної науково-практичної конференції "Глобальні та регіональні проблеми інформатизації в суспільстві і природокористуванні '2025", 13–14 листопада 2025 року, НУБіП України, Київ. – К.: НУБіП України, 2025. – 206 с.

Відповідальність за зміст публікацій несуть автори.

© Національний університет біоресурсів
і природокористування України, 2025