

ДОСЛІДЖЕННЯ ІНСТРУМЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Вінницький національний технічний університет

Анотація

У роботі досліджено сучасний інструментарій штучного інтелекту (ШІ), призначений для інтелектуального аналізу даних (Data Mining). Проаналізовано ключові класи алгоритмів, включаючи методи класифікації, кластеризації, регресійного та асоціативного аналізу. Здійснено порівняльну характеристику популярних програмних засобів: бібліотек Scikit-learn та TensorFlow, а також автоматизованих платформ RapidMiner та Google AutoML. Визначено критерії вибору оптимальних аналітичних рішень залежно від обсягу даних, необхідної точності та рівня технічної підготовки користувача. Результати дослідження підтверджують, що глибока інтеграція ШІ у процеси аналітики є критично важливою для виявлення прихованих закономірностей та прийняття управлінських рішень в умовах цифровізації.

Ключові слова: штучний інтелект, інтелектуальний аналіз даних, Data Mining, машинне навчання, нейронні мережі, Scikit-learn, TensorFlow, AutoML, цифровізація.

Abstract

The paper explores modern artificial intelligence (AI) tools designed for intelligent data analysis (Data Mining). Key classes of algorithms are analyzed, including methods of classification, clustering, regression, and associative analysis. A comparative characterization of popular software tools is carried out: Scikit-learn and TensorFlow libraries, as well as RapidMiner and Google AutoML automated platforms. Criteria for choosing optimal analytical solutions are determined depending on the data volume, required accuracy, and the user's technical background. The research results confirm that the deep integration of AI into analytical processes is crucial for discovering hidden patterns and making management decisions in the context of digitalization.

Keywords: artificial intelligence, intelligent data analysis, Data Mining, machine learning, neural networks, Scikit-learn, TensorFlow, AutoML, digitalization.

Вступ

Одним із найбільш розвинених напрямів сучасної науки та техніки є штучний інтелект (ШІ). Інформаційні завдання, які раніше вважались можливими лише людиною: розпізнавання та розуміння мови, передбачення дій, прийняття рішень – сьогодні виконується різними інструментами ШІ.

Фундаментом для трансформації у промисловості, медицині та фінансовому секторі є стрімкий розвиток штучного інтелекту. В умовах постійного зростання об'ємів інформації особливої актуальності набуває інтелектуальний аналіз даних (Data Mining), який дозволяє виявляти приховані закономірності та формувати нові знання на основі великих масивів даних. Особливої актуальності набуває інтелектуальний аналіз даних (ІАД), що базується на машинному навчанні та статистичному аналізі, оскільки він забезпечує перехід від накопичення даних до формування нових знань. Постає необхідність пошуку найбільш оптимальних рішень з огляду на їх відкритість, зручність інтеграції та масштабованість.

Метою роботи є дослідження та комплексний аналіз сучасних інструментів штучного інтелекту, що використовуються для задач Data Mining, а також виявлення їхніх переваг і недоліків для подальшого застосування у фаховій діяльності.

Об'єктом дослідження є процес застосування технологій штучного інтелекту в аналітичних системах.

Предметом дослідження є інструментарій штучного інтелекту для інтелектуального аналізу даних (зокрема фреймворки Scikit-learn, TensorFlow та платформи AutoML).

Результати дослідження

1. Алгоритми інтелектуального аналізу даних та їх інтеграція з ШІ.

Ключовою технологічною компонентою штучного інтелекту, що забезпечує навчання системи на основі досвіду вважається інтелектуальний аналіз даних (ІАД). ШІ додає процесам аналітики

адаптивності та самонавчання, перетворюючи ІАД на комплексну систему отримання знань. Сучасні алгоритми, що застосовуються в цій галузі, поділяються на кілька основних класів:

- алгоритми класифікації: дерева рішень (Decision Trees), метод k-найближчих сусідів (k-NN), наївний байєсівський класифікатор та метод опорних векторів (SVM). Найвищу точність забезпечують штучні нейронні мережі. Ці методи є важливими для медичної діагностики та виявлення шахрайських операцій;

- алгоритми кластеризації: методи k-середніх (k-Means), ієрархічна кластеризація, DBSCAN для виявлення аномалій та нейронні карти Кохонена (SOM) – групують об'єкти за схожістю без попередніх категорій;

- асоціативний аналіз: алгоритми Apriori, FP-Growth та Eclat – виявлення закономірностей у великих масивах;

- регресійний аналіз: класична лінійна та логістична регресія, градієнтний бустинг (XGBoost, LightGBM) та глибокі нейронні мережі – моделювання залежностей між змінними;

- виявлення аномалій та ансамблеві методи: методи Isolation Forest або автоенкодера, Random Forest, CatBoost – виявлення нетипової поведінки та досягнення стійкості до перенавчання.

Напрямок Deep Learning використовує багатопланові нейронні мережі для автоматичного виділення ознак. Актуальними трендами розвитку галузі є впровадження систем AutoML, які автоматизують вибір алгоритмів, та напрямок Explainable AI (XAI), що забезпечує інтерпретованість результатів і підвищує довіру користувачів. Реалізація цих підходів здійснюється через потужні програмні засоби: бібліотеки Scikit-learn, TensorFlow, PyTorch, а також платформи RapidMiner та IBM SPSS Modeler.

2. Аналіз програмних інструментів інтелектуального аналізу даних

Сучасні алгоритми Data Mining реалізуються за допомогою програмних засобів, які варіюються від гнучких бібліотек програмування до автоматизованих хмарних платформ:

- бібліотека Scikit-learn – це відкрите програмне забезпечення на базі Python, що стало стандартом для реалізації класичних алгоритмів аналізу даних. Бібліотека побудована поверх NumPy та SciPy, що забезпечує високу швидкість обчислень завдяки реалізації алгоритмів на C/C++. Простий у створенні складних моделей кількома рядками коду, стабільний в експериментах та має наявність докладної документації. Відсутня підтримка складних нейронних мереж, обмежений при роботі з великими об'ємами даних через необхідність завантаження всього набору в оперативну пам'ять, а також виконання обчислень виключно на CPU;

- фреймворк TensorFlow – це інструмент, який є еталоном для глибокого навчання та побудови складних нейромереж. Робота фреймворка базується на обчислювальних графах, де вузли виконують математичні операції, а ребра представляють тензори (масиви даних). У нього висока швидкодія, масштабованість від мобільних пристроїв до серверів та наявна потужна екосистема (Keras API, TensorBoard, TF Lite). Проте також складний низькорівневий API, високий поріг входження для початківців та значні вимоги до обчислювальних ресурсів;

- платформа RapidMiner – це платформа, яка орієнтована на повний цикл Data Mining без необхідності написання програмного коду. Архітектура побудована на принципі візуального моделювання, де процеси представлені у вигляді дерева операцій. Вона має інтуїтивно зрозумілий графічний інтерфейс, можливість автоматичного підбору алгоритмів та візуалізація всієї логіки моделі у вигляді схем. Проте споживає багато оперативної пам'яті;

- Google AutoML – це система хмарних сервісів для автоматизації побудови моделей ШІ. Використовує технологію Neural Architecture Search (NAS) для автоматичного пошуку найбільш точної архітектури нейронної мережі. Не потребує навичок програмування, забезпечує високу точність моделей та легке розгортання через REST API. Висока вартість використання, «закритість» алгоритмів та залежність від стабільного інтернет-з'єднання з хмарою Google Cloud;

- Microsoft Power BI – платформа для навчання моделей, є критично важливою для візуалізації результатів Data Mining. Вона підтримує понад 120 джерел даних та має вбудовані функції прогнозування в реальному часі. В ній обмежена робота з великими графами та складна мова запитів DAX для новачків.

Вибір конкретного інструменту для інтелектуального аналізу даних залежить від балансу між необхідною гнучкістю налаштувань, обсягом обчислювальних ресурсів та рівнем технічної підготовки користувача.

3. Узагальнена класифікація інструментів ШІ для Data Mining.

Сучасний інструментарій штучного інтелекту, що застосовується для задач Data Mining, можна

класифікувати за кількома ключовими ознаками, що визначають ефективність їхнього впровадження у конкретних сферах діяльності. За рівнем автоматизації та типом взаємодії з користувачем виокремлюють низькорівневі бібліотеки, як-от Scikit-learn та TensorFlow, що потребують глибоких навичок програмування, а також візуальні платформи на кшталт RapidMiner, які дозволяють будувати моделі без написання коду. Системи автоматизованого машинного навчання (AutoML), що самостійно підбирають оптимальну архітектуру та параметри нейронних мереж виділяються окремо.

З обчислювальної архітектури, інструменти поділяються на класичні рішення для роботи з табличними даними на центральних процесорах (CPU) та спеціалізовані фреймворки для глибокого навчання, що використовують потужності графічних прискорювачів (GPU) для обробки зображень та звуку. Важливим критерієм є також тип доступу до програмного забезпечення: відкриті бібліотеки (open source) забезпечують прозорість алгоритмів та активний розвиток спільнотою, тоді як комерційні корпоративні системи пропонують високу стабільність, підтримку та легку інтеграцію в бізнес-інфраструктуру.

За функціональним призначенням ці засоби охоплюють шлях від дослідницьких середовищ для створення нових моделей до потужних аналітичних платформ, як Microsoft Power BI, що фокусуються на візуалізації результатів та підтримці прийняття рішень у реальному часі. Така багаторівнева класифікація дозволяє фахівцям обирати технологічний стек, який найкраще відповідає вимогам точності, масштабованості та наявних ресурсів конкретного проєкту.

Висновки

У роботі було проведено комплексне дослідження, яке підтверджує, що в умовах сучасної цифровізації ефективність процесів інтелектуального аналізу даних безпосередньо залежить від глибокої інтеграції з методами штучного інтелекту. Використання машинного навчання, нейронних мереж та ансамблевих моделей дозволяє трансформувати просте накопичення інформації у процес формування нових знань та виявлення прихованих закономірностей, що є критично важливим для прийняття управлінських рішень.

Аналіз програмного інструментарію показав, що вибір конкретних засобів має базуватися на специфіці поставлених задач та наявних ресурсах. Для наукових досліджень і створення унікальних архітектур найбільш доцільним є використання бібліотек із відкритим кодом, таких як Scikit-learn та TensorFlow, які пропонують максимальну гнучкість і прозорість алгоритмів. Водночас для швидкого впровадження аналітики в бізнес-процеси без необхідності глибокого програмування оптимальними рішеннями виступають платформи RapidMiner та Google AutoML.

СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

1. Дослідження інструментів штучного інтелекту для інтелектуального аналізу даних / І. Лисенко, Р. Минайленко, С. Смірнов та ін. [Електронний ресурс]. – Режим доступу: <https://csecurity.kubg.edu.ua/index.php/journal/article/view/1022> (дата звернення: 02.03.2026).
2. Scikit-learn: Machine Learning in Python [Електронний ресурс]. – Режим доступу: <https://scikit-learn.org/stable/> (дата звернення: 02.03.2026).
3. RapidMiner Documentation [Електронний ресурс]. – Режим доступу: <https://docs.rapidminer.com/> (дата звернення: 02.03.2026).
4. KNIME Documentation [Електронний ресурс]. – Режим доступу: <https://docs.knime.com/ap/latest/> (дата звернення: 02.03.2026).
5. TensorFlow Developers Guide [Електронний ресурс]. – Режим доступу: <https://www.tensorflow.org/guide> (дата звернення: 02.03.2026).

Анастасія Костянтинівна Кірчанова – студентка групи 1КІТС-24б, факультет менеджменту та інформаційної безпеки, Вінницький національний технічний університет, м. Вінниця, e-mail: n.kircman@gmail.com;

Anastasia K. Kirchanova – student of group 1KITS-24b, Faculty of Management and Information Security, Vinnytsia National Technical University, Vinnytsia, e-mail: n.kircman@gmail.com.