



DOI 10.28925/2663-4023.2025.31.1022

УДК 004.7

Лисенко Ірина Анатоліївна

кандидат технічних наук, старший викладач кафедри кібербезпеки та програмного забезпечення
Центрально український національний технічний університет, Кропивницький, Україна

m

i

Минайленко Роман Миколайович

кандидат технічних наук, доцент кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна

n

m

a

x

i

Смірнов Сергій Анатолійович

кандидат технічних наук, доцент, доцент кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна

Буравченко Костянтин Олегович

кандидат технічних наук, доцент кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна

Якименко Наталія Миколаївна

кандидат фізико-математичних наук, доцент, доцент кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна

Смірнов Олексій Анатолійович

доктор технічних наук, професор, завідувач кафедри кібербезпеки та програмного забезпечення
Центральноукраїнський національний технічний університет, Кропивницький, Україна

dr.smirnovoa@gmail.com

ДОСЛІДЖЕННЯ ІНСТРУМЕНТІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

Анотація. У статті проведено комплексне дослідження сучасних інструментів штучного інтелекту (ШІ), призначених для інтелектуального аналізу даних (Data Mining), їх обчислювальних можливостей та сфер практичного застосування. Розглянуто ключові алгоритми аналізу даних, зокрема методи машинного навчання, глибинного навчання, кластеризації, класифікації, регресійного моделювання та нейронних мереж. Особливу увагу приділено інструментам Scikit-learn, TensorFlow, RapidMiner, Google AutoML, Power BI та іншим платформам та фреймворкам, що забезпечують автоматизацію процесів обробки, аналізу та візуалізації великих об'ємів даних. У роботі здійснено порівняльний аналіз переваг і недоліків найпоширеніших інструментів ШІ з точки зору точності моделей, обчислювальної ефективності, простоти інтеграції, масштабованості, підтримки AutoML та можливостей роботи з неструктурованими даними. Наведено приклади застосування інструментів ШІ у промисловості, економіці, медицині, фінансовому секторі та інших сферах діяльності людини. Враховано тенденції розвитку інтелектуального аналізу даних з огляду на зростання ролі хмарних платформ, автоматизованих систем створення моделей, мультимодальних моделей ШІ та інтеграції з корпоративними аналітичними системами. Результати



дослідження дозволяють визначити найбільш оптимальні підходи та інструменти для вирішення прикладних завдань Data Mining, забезпечуючи вибір технологій відповідно до вимог точності, продуктивності та відкритості. Отримані висновки можуть бути використані в наукових дослідженнях, бізнес-аналітиці, цифровій трансформації підприємств та проектуванні інтелектуальних систем підтримки прийняття рішень.

Ключові слова: штучний інтелект, інтелектуальний аналіз даних, машинне навчання, нейронні мережі, алгоритми аналізу даних, Data Mining, TensorFlow, Scikit-learn, RapidMiner,

ВСТУП

Постановка завдання дослідження

Зважаючи на те, що штучний інтелект (ШІ) сьогодні є одним із найбільш швидко розвинених напрямів сучасної науки і техніки, його методи та інструменти застосовуються для розв'язання завдань, які ще кілька десятиліть тому вважалися виключно можливими лише людиною: розпізнавання образів, розуміння людської або тваринної мови, прогнозування, прийняття рішень у невизначених умовах. Інтелектуальні системи стають основою цифрової трансформації у всіх сферах – від промисловості до освіти, від медицини до фінансового сектору.

В умовах зростання об'ємів інформації та ускладнення структур даних особливої актуальності набуває інтелектуальний аналіз даних (ІАД) (Data Mining), що базується на методах машинного навчання, нейронних мереж, статистичного аналізу та візуалізації результатів. Саме ІАД забезпечує перехід від простого накопичення даних до формування нових знань, виявлення прихованих закономірностей та побудови моделей поведінки об'єктів чи процесів.

У сучасних умовах ефективність процесів Data Mining значною мірою покращилась за рахунок великого набору інструментів штучного інтелекту, які дозволяють не лише будувати моделі, а й автоматизувати їх налаштування, навчання та оцінювання якості. Розвиток таких інструментів сприяє переходу від спеціалізованих досліджень до масового використання аналітичних сервісів у бізнесі, державному управлінні, освіті та науці.

Застосування інструментів ШІ для ІАД відкриває широкі можливості з підвищення точності прогнозів завдяки глибокому навчанню (Deep Learning), автоматичного виявлення закономірностей у потоках даних у режимі реального часу, інтеграції з великими об'ємами даних (Big Data) та хмарними аналітичними платформами, створення інтерпретованих моделей, придатних для пояснення користувачем результатів (напрямок Explainable AI).

Постановка проблеми. Водночас із розвитком ШІ та активним впровадженням його інструментів в процеси аналізу даних постають нові виклики: етичні та правові аспекти використання штучного інтелекту, прозорість алгоритмів, захист персональних даних, необхідність стандартизації форматів і процедур аналітики.

Особливу роль у цьому контексті відіграють інструменти відкритого коду (open source), які формують середовище колективного розвитку і сприяють швидкому поширенню нових розробок. До них належать бібліотеки Scikit-learn, TensorFlow, PyTorch, а також аналітичні платформи KNIME та RapidMiner. Досить активно розвиваються комерційні системи (IBM SPSS Modeler, Google AutoML, Microsoft Power



ВІ), які інтегруються у корпоративні середовища та забезпечують автоматизацію бізнес-аналітики.

Таким чином постає завдання пошуку найбільш оптимальних рішень відповідно до наступних критеріїв: зручність використання, здатність інтегруватися з іншими бібліотеками та платформами, відкритість, можливість використання безкоштовної версії у повному чи обмеженому варіанті. Дослідження існуючих інструментів відповідно до заданих умов дає можливість їх повної оцінки для обрання більш оптимального варіанту.

Аналіз останніх досліджень і публікацій. В роботах [1, 2, 19] розглядаються основні алгоритми інтелектуального аналізу даних, їх можливості та сфери застосування. Автори розкривають основи та підґрунтя застосування штучного інтелекту для Data Mining. На думку авторів, це високотехнічні та дуже перспективні напрямки розвитку та вдосконалення класичних алгоритмів аналізу та можливостей штучного інтелекту.

У роботах [27-29, 32-37] автори приділяють окрему увагу напрямкам застосування алгоритмів машинного навчання для розв'язання практичних задач. Сучасні алгоритми штучного інтелекту дають досить широкий спектр інструментів в умовах їх застосування до конкретних умов аналізу даних та прогнозування. В результаті огляд практичних задач та можливостей їх розв'язання вказує на те, що інтеграція класичних алгоритмів аналізу даних із алгоритмами штучного інтелекту дає гарні результати під час аналізу та моделювання.

Метою статті є дослідження та комплексний аналіз сучасних інструментів штучного інтелекту, що використовуються для інтелектуального аналізу даних, виявлення їхніх переваг і недоліків, а також формування узагальненої класифікації за типом, архітектурою та сферою застосування.

Об'єктом дослідження є процес застосування штучного інтелекту.

Предметом дослідження є визначення інструментів штучного інтелекту для інтелектуального аналізу даних.

ТЕОРЕТИЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

Алгоритми інтелектуального аналізу даних та їх інтеграція з ШІ

Інтелектуальний аналіз даних (ІАД) поєднує широкий спектр методів і алгоритмів, спрямованих на виявлення прихованих закономірностей, класифікацію об'єктів, прогнозування та побудову моделей поведінки. У контексті штучного інтелекту ІАД виступає як ключова технологічна компонента, що реалізує навчання системи на основі досвіду, тобто даних.

Штучний інтелект забезпечує адаптивність, самонавчання та евристичну оптимізацію аналітичних процесів, роблячи ІАД не просто інструментом обробки інформації, а комплексною системою отримання знань. Алгоритми, які застосовуються в цій галузі, умовно поділяються на кілька основних класів.

1. Алгоритми класифікації

Мета класифікації полягає у віднесенні об'єктів до певних наперед визначених категорій. До основних алгоритмів цього типу належать:



– Древа рішень (Decision Trees) – інтерпретовані моделі, які створюють правила у вигляді послідовності логічних розгалужень. Вони зручні для побудови прозорих систем підтримки прийняття рішень.

– Метод k -найближчих сусідів (k -NN) – простий, але ефективний алгоритм для класифікації об'єктів за відстанню у багатовимірному просторі ознак.

– Наївний баєсівський класифікатор (Naive Bayes) – застосовується для текстової класифікації, аналізу спаму та емоцій у повідомленнях.

– Метод опорних векторів (SVM) – використовується для пошуку оптимальної межі між класами в даних великої розмірності.

– Нейронні мережі (Artificial Neural Networks) – основа сучасних моделей глибокого навчання, які забезпечують найвищу точність при роботі з великими масивами даних.

У застосуваннях Data Mining класифікаційні алгоритми використовуються для виявлення шахрайських операцій, прогнозування банкрутства, медичної діагностики, сортування клієнтів у CRM-системах [3].

2. Алгоритми кластеризації

На відміну від класифікації, кластеризація не потребує попереднього визначення категорій. Її мета – групувати об'єкти за ступенем схожості. Найпоширеніші алгоритми:

– k -середніх (k -Means) – простий і швидкий алгоритм для створення кластерів за центрами тяжіння;

– Ієрархічна кластеризація – створює дерево об'єднання об'єктів, що зручно для візуалізації;

– DBSCAN – алгоритм кластеризації за щільністю, ефективний для виявлення шумів і аномалій;

– Self-Organizing Maps (SOM) – нейронні карти Кохонена, що дозволяють відобразити багатовимірні дані у двовимірному просторі.

Кластеризація використовується для сегментації клієнтів, аналізу соціальних мереж, групування текстів та зображень [4].

3. Алгоритми асоціативного аналізу

Асоціативні правила дозволяють виявити закономірності типу «якщо ... то ...» у великих наборах транзакційних даних. Класичний приклад – алгоритм Apriori, який виявляє взаємозв'язки між товарами у торгових мережах. Більш сучасні методи це FP-Growth та Eclat, які забезпечують високу продуктивність при обробці мільйонів записів.

Асоціативний аналіз активно використовується у маркетингу, логістиці, рекомендаційних системах [5].

4. Алгоритми регресійного аналізу

Регресія спрямована на моделювання залежності між змінними.

Класичні методи – лінійна та логістична регресія залишаються базовими для більшості прикладних завдань. У складніших випадках застосовують регресійні дерева, градієнтний бустинг (XGBoost, LightGBM) та глибокі нейронні мережі, які можуть апроксимувати складні нелінійні залежності [6, 7].

Регресійні алгоритми застосовуються в прогнозуванні попиту, фінансовому аналізі, прогнозуванні кліматичних змін тощо.

5. Алгоритми виявлення аномалій

Окрему групу становлять алгоритми для виявлення відхилень або нетипової поведінки. До них належать: Isolation Forest, One-Class SVM, Autoencoder Networks (автоенкодера для пошуку аномалій у часових рядах).



Такі методи широко використовуються у фінансовому моніторингу, кібербезпеці, технічній діагностиці.

6. Інтелектуальні ансамблеві методи

Ансамблеві методи – це комбінації кількох моделей для досягнення вищої точності. Найвідоміші з них: Random Forest, Bagging, Boosting (XGBoost, CatBoost,

Вони забезпечують стійкість до перенавчання, високу здатність до узагальнення і зручність у застосуванні.

Глибоке навчання (Deep Learning) – це напрям, який базується на багатошарових нейронних мережах, що автоматично виділяють ознаки з даних.

Основні архітектури: CNN (Convolutional Neural Networks) – обробка зображень, відео, RNN (Recurrent Neural Networks) – робота з часовими рядами, мовою, Transformers (BERT, GPT) – універсальні моделі природної мови, перекладу, генерації текстів [9, 10].

Сучасним трендом є використання AutoML (Automated Machine Learning) систем, які автоматично обирають алгоритм і параметри моделі без участі аналітика. Також набуває популярності напрям Explainable AI (XAI) пояснюваного штучного інтелекту, який дозволяє інтерпретувати результати та підвищує довіру користувачів.

Реалізація цих алгоритмів можлива завдяки численным програмним засобам: Scikit-learn – класичні методи класифікації, регресії, кластеризації; TensorFlow, середовища для створення робочих процесів без програмування; IBM SPSS Modeler – автоматизований аналітичний комплекс для бізнесу [11, 12].

Важливо зазначити, що штучний інтелект не лише використовує ці алгоритми, а й постійно їх вдосконалює. Завдяки мета-навчанню (meta-learning) системи здатні навчатися вибору найкращих методів для конкретних складних задач. У поєднанні з хмарними обчисленнями та квантовими підходами ШІ відкриває перспективи створення аналітичних систем, що здатні працювати в реальному часі та самостійно навчатись.

Сучасні інструменти штучного інтелекту для Data Mining

Враховуючи сучасний стан розвитку технологій машинного навчання, хмарних обчислень та Big Data, що сприяв появі значної кількості інструментів штучного інтелекту, маємо суттєве розширення можливостей інтелектуального аналізу даних. Ці інструменти можна поділити на три основні види: бібліотеки програмування, візуальні аналітичні платформи та хмарні сервіси автоматизованого машинного навчання. Розглянемо кожен вид окремо більш детально. До найбільш поширених бібліотек штучного інтелекту та машинного навчання належать: Scikit-learn, TensorFlow, PyTorch,

– універсальна Python-бібліотека, що працює із класичними алгоритмами машинного навчання таких як класифікація, регресія, кластеризація, зниження розмірності даних. Переваги цієї бібліотеки у простоті синтаксису, інтеграції з *NumPy*, *Pandas* і *Matplotlib*, що робить її зручним навчальним інструментом від початківців і дослідників [13].

– фреймворк глибокого навчання від Google, який підтримує обчислення на GPU/TPU, також створення складних нейронних архітектур і розгортання моделей у хмарному середовищі. Він має добре розвинену екосистему (TensorFlow Extended, TensorBoard, TensorFlow Lite) для практичного впровадження моделей та тестування [14].

– розроблений Meta (Facebook) фреймворк для створення гнучких нейронних мереж. Його головна перевага це динамічні обчислювальні графи, що дозволяють



експериментувати з архітектурою у процесі виконання коду. PyTorch широко використовується у наукових та практичних дослідженнях, зокрема в NLP і Computer

– високорівнева надбудова над TensorFlow, яка забезпечує спрощене створення нейронних мереж завдяки інтуїтивному API. Часто використовується для швидкого створення прототипів систем [16].

– алгоритми градієнтного бустингу, що дають високу точність при роботі з табличними даними і широко використовуються у змаганнях *Kaggle* [17].

Візуальні аналітичні платформи Data Mining це: RapidMiner, KNIME, Orange, IBM SPSS Modeler. Вони створені для користувачів, які не мають великого досвіду у програмуванні, такі середовища мають графічний інтерфейс і дозволяють створювати моделі без написання безпосередньо коду:

– RapidMiner це платформа для інтелектуального аналізу даних з інтуїтивно зрозумілим інтерфейсом відображення та побудови процесів. Підтримує процедури імпорту даних, підготовки та навчання моделей, а також візуалізацію результатів. Маючи розширення для роботи з Big Data і глибоким навчанням, RapidMiner часто використовується у бізнес-аналітиці [18].

– KNIME (Konstanz Information Miner) це відкрита платформа з модульною архітектурою, яка дозволяє створювати складні робочі потоки всередині моделей систем. KNIME підтримує інтеграцію з Python, R, Weka, TensorFlow і базами даних, а також є можливість візуального налагодження процесу обробки даних та високий рівень масштабованості [19].

– Orange Data Mining це проста у використанні система для навчання початківців та швидкого експериментування з різними видами даних та моделями. Забезпечує інтерактивну візуалізацію кластеризації, дерев рішень та нейронних мереж.

– IBM SPSS Modeler це комерційна платформа з розвиненим модулем автоматизованого моделювання, яка потужні засоби прогнозування, аналізу часових рядів та сегментації клієнтів. SPSS Modeler використовується у фінансових установах, маркетингових компаніях і державних структурах для моделювання та прогнозування

З розширенням та повсякчасним використанням хмарних технологій у більшості сфер наукових та практичних досліджень з'явилися та активно розвиваються AI-сервіси нового покоління, які дозволяють автоматично створювати, тестувати й розгортати моделі без написання коду. До них відносяться: Google AutoML, Microsoft Azure Machine Learning, Amazon SageMaker, DataRobot. Вони підтримують функції автоматичного вибору алгоритму аналізу, налаштування різноманітних параметрів та оцінка результатів. Також вони надають можливості інтегрувати результати в інші хмарні сервіси компанії та забезпечують гарну масштабованість і безпеку.

Очевидно, що результати аналізу даних мають бути представлені зрозуміло, докладно та наочно. Для цього використовують такі системи, як Tableau, Microsoft Power BI та Google Data Studio. Кожна з них має схожий функціонал але все таки відрізняється деякими особливостями. Так Tableau забезпечує інтерактивну візуалізацію великих обсягів даних, побудову дашбордів, виявлення трендів серед покупців, Power навчання, а от Google Data Studio є безкоштовною хмарною альтернативою для створення інтерактивних звітів.

Оскільки створення моделі та проведення всебічного ґрунтовного інтелектуального аналізу даних це процес неперервний і складний тож він вимагає на



кожному етапі різних інструментів обробки. Таким чином візуальні інструменти часто поєднуються з аналітичними платформами, забезпечуючи єдиний цикл аналітики – від збору даних, структуризації та обробки до візуалізації результатів та прийняття рішення.

РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ

З метою всебічного огляду наявних інструментів штучного інтелекту, які використовуються для інтелектуального аналізу даних далі буде порівняно окремі рішення, що мають досить вузький спектр застосувань або навпаки надають широкі можливості до використання. Більшість цих інструментів мають безкоштовний доступ, працюють достатньо ефективно та не вимагають глибоких знань з програмування. При цьому мають місце і закриті системи з платним доступом, різними ступенями інтеграції до середовищ кампаній та гарним функціоналом.

Scikit-learn це одна з найпопулярніших бібліотек машинного навчання у Python, яка стала фактично стандартом для реалізації класичних алгоритмів Data Mining. Це відкрите програмне забезпечення з широкою спільнотою розробників, що забезпечує стабільність, повторюваність експериментів і простоту інтеграції з іншими інструментами середовища Python.

Scikit-learn побудований поверх фундаментальних бібліотек Python: NumPy для роботи з багатовимірними масивами, SciPy для математичних операцій та статистики, Усі алгоритми в бібліотеці реалізовані за єдиним державним стандартом API. Також він містить практично всі класичні алгоритми інтелектуального аналізу даних: метод опорних векторів, метод дерева рішень, метод логістичної регресії, методи кластеризації та асоціативного аналізу обробки текстів.

Простота використання цього інструменту полягає в тому, що він дозволяє створювати досить складні моделі за допомогою декількох рядків програмного коду. Незважаючи на це, є можливість будувати якісні та стабільні моделі навіть при роботі з неструктурованими даними. Для забезпечення високої продуктивності більшість алгоритмів реалізовано на C/C++, а через відкритий код та широку спільноту користувачів має дуже великий і докладний пакет документації, різноманітні розширення та добудови, підтримує як суто наукові так і прикладні дослідження.

Попри велику популярність серед користувачів дана бібліотека має ряд недоліків. А саме: не підтримує навчання складних нейронних мереж, оскільки потребує завантаження всього набору даних в пам'ять то має обмеження по роботі з великими об'ємами, всі обчислення відбуваються тільки на CPU, не підходить для обробки зображень і текстів у їхній оригінальній формі. Частково ці недоліки компенсуються комбінаціями з Pandas, NumPy та TensorFlow/PyTorch Scikit-learn, що в результаті забезпечує повний цикл Data Mining [21, 22].

Один з найпотужніших та найвідоміших фреймворків для глибокого навчання створений Google Brain це TensorFlow. Якщо казати про стандарт у комерційному використанні нейронних мереж, обробки та аналізу зображень, текстів, звукових файлів, роботи з великими об'ємами даними та побудови складних моделей штучного інтелекту то це все про нього. TensorFlow дає можливість забезпечити масштабованість,



швидкодію та розгортання моделей на різних платформах від мобільних пристроїв до великих серверів [23].

Робота TensorFlow організована за принципом побудови обчислювальних графів, у яких є вузли (nodes) – це математичні операції, ребра (edges) – це тензори (багатовимірні масиви даних). Сама назва TensorFlow відображає буквально ідею «потоків тензорів» через граф обчислень.

TensorFlow – це велика «екосистема» різноманітних додатків та бібліотек, що включає: TensorFlow Core (базовий рівень для створення вибіркового моделі із максимальним доступом), Keras API (високорівневий інтерфейс, який дозволяє створювати моделі лише декількома рядками коду), TensorBoard (система візуалізації), TensorFlow Lite (додаток для мобільних пристроїв), TensorFlow Extended (комерційне застосування аналізу та створення моделей), TensorFlow Serving (система для створення і аналізу моделей у вигляді сервісів). Також TensorFlow підтримує всі основні архітектури глибокого навчання.

Таким чином серед переваг TensorFlow можна виділити масштабованість, високу продуктивність, дуже розгалужену екосистему різноманітних додатків та сервісів, підтримку комерційного використання та універсальність. Серед недоліків варто визначити: порівняно складний низькорівневий API, набагато вищу складність для початківців, необхідність використання значних обчислювальних ресурсів для достатнього об'єму обчислень, меншу гнучкість при застосуванні у наукових дослідженнях.

На даний час TensorFlow – це один із найпотужніших інструментів для створення систем штучного інтелекту на рівні корпорацій, дослідницьких лабораторій і хмарних платформ.

Однією з найпотужніших платформ інтелектуального аналізу даних, орієнтована на користувачів, які бажають виконувати повний цикл процесу Data Mining без програмування є RapidMiner [24]. Ця платформа використовується у бізнес-аналітиці, комерційних та виробничих дослідженнях, маркетингу, фінансах та освіті.

Архітектура та концепції RapidMiner побудовані за принципом візуального моделювання, коли користувач створює моделі із готових блоків. Для того щоб аналіз був інтуїтивно зрозумілим кожен з процесів представлено у вигляді дерева операцій з'єднаних між собою потоками даних.

Серед функціональних можливостей представлені всі етапи життєвого циклу аналізу даних, а саме: підготовка та очищення даних, моделювання з використанням десятків інтуїтивно зрозумілих алгоритмів, які не потребують написання програмного коду, автоматичний підбір алгоритмів, оцінювання моделей за допомогою вбудованих інструментів, розгортання моделей досить широкого спектру. Весь наведений функціонал дає можливість створити за необхідністю досить великі аналітичні проєкти.

Однією з основних переваг RapidMiner варто зазначити, що всю логіку моделей можна побачити у вигляді графічної схеми. Ця можливість значно розширює можливості візуалізації процесу та представлення результатів аналізу. Також серед важливих переваг для початківці – повна відсутність потреби у програмуванні і написанні коду. Завдяки RapidMiner Server аналітичну модель можливо представити у вигляді постійно працюючого сервісу.



Недоліки та обмеження RapidMiner це: умовно безкоштовна версія (містить багато обмежень в роботі), високі вимоги до ресурсів, зокрема до оперативної пам'яті, менша гнучкість порівняно з Python-бібліотеками.

Google AutoML – це комплекс хмарних сервісів Google Cloud, призначений для автоматизації машинного навчання та побудови моделей ШІ без необхідності програмування або глибоких технічних та аналітичних знань.

Основна ідея Google AutoML – максимальна автоматизація створення моделей, а також використання технології Neural Architecture Search (NAS), яка забезпечує автоматичний пошук архітектур нейронних мереж з найвищою точністю [25].

Екосистема продуктів Google AutoML складається з кількох автоматизованих сервісів: AutoML Vision (створення моделей для аналізу зображень), AutoML Natural табличних даних), AutoML Video Intelligence (відеоаналіз, розпізнавання об'єктів та подій у відеопотоці), AutoML Translation (модуль автоматичного перекладу, який дозволяє створювати моделі перекладу для тематичних текстів).

AutoML виконує автоматично такі завдання, як обробка даних, вибір моделі, оптимізація гіперпараметрів та оцінка моделі, які у традиційному ML-середовищі потребують ручної роботи. Також існує можливість автоматичної генерації докладного звіту про якість моделі, що підходить для наукових досліджень та корпоративного аналізу.

Тож серед переваг Google AutoML слід зазначити відсутність потреби у програмуванні, автоматизована система створення високоточних моделей, масштабованість, можливість інтеграції із Google Cloud, а також легке розгортання моделі коли одним кліком модель можна перетворити на REST API та інтегрувати у будь-який застосунок.

Попри значні можливості, AutoML має також ряд недоліків: закритість алгоритмів, обмежена гнучкість, висока вартість та необхідність стабільного інтернет-з'єднання та акаунту Google Cloud. Не дивлячись на це Google AutoML залишається одним з найзручніших інструментів для бізнесу та організацій, які не мають великих команд Data

Power BI від компанії Microsoft – це потужна платформа бізнес-аналітики та візуалізації даних. Вона забезпечує повний цикл аналізу даних: від роботи безпосередньо з джерелами, очищення та моделювання – до створення інтерактивних застосунків, звітів і автоматизованої аналітики. Завдяки інтеграції з іншими продуктами Microsoft (Excel, Azure, SQL Server, Power Apps) Power BI є одним з найпопулярніших інструментів корпоративної аналітики і дедалі частіше використовується у Data Mining для візуалізації результатів моделей аналізу даних [26].

Архітектура та компоненти Power BI складається з кількох елементів, що формують єдину систему: Power BI Desktop (снвне середовище розробки аналітичних матеріалів), Power BI Service (Cloud) (автоматична система для звітів та спільної роботи), Power BI Mobile (мобільний додаток для перегляду файлів та застосунків), Power Query (механізм ETL-трансформацій для роботи з різними наборами даних), Power Pivot/Data Model (система безпосередньо моделювання даних).

Power BI підтримує понад 120 джерел, що дозволяє використовувати його як універсальний інструмент для інтеграції різних типів даних. Хоча Power BI не є



класичним середовищем машинного навчання, він широко застосовується у Data Mining як інструмент візуалізації результатів, а також має власні ML-функції. Вбудовані алгоритми прогнозування дають можливість проводити прогнозні обчислення в реальному часі, розраховувати сезонність та виявляти відхилення.

Серед переваг Power BI варто зазначити інтерактивність, що дозволяє одразу аналізувати поведінку моделі або сегменти даних, високу продуктивність, інтеграція з таблицями Excel, можливість автоматичного оновлення даних в режимі реального часу та відносна доступність.

Недоліками Power BI залишаються обмежена візуалізація великих графів, обмеження по кількості рядків у безкоштовній версії, складність DAX для новачків, а також те, що він не є повноцінною платформою машинного навчання.

Узагальнена класифікація інструментів ШІ для Data Mining

Користуючись отриманими результатами аналізу найбільш використовуваних додатків штучного інтелекту для інтелектуального аналізу даних складемо порівняльну таблицю, що містить їх класифікацію, переваги, недоліки, ступінь відкритості середовища [27, 29, 30].

Таблиця 1

Порівняльна таблиця засобів ШІ для ІАД

Назва	Тип	Відкритість (так/ні/частково)	Основний функціонал	Сфери застосування
	Бібліотека ML	так	Класифікація, кластеризація, регресія	Освіта, наука
	Фреймворк DL	так	Нейронні мережі, CNN,	ШІ-дослідження, виробництво, бізнес
	Платформа	частково	Візуальне моделювання,	Бізнес-аналітика
	Хмарний сервіс	ні	Автоматичне навчання моделей	Швидке прототипування
	BI-система	ні	Візуалізація, звітність	Корпоративна аналітика

Таким чином слід зазначити, що кожен з інструментів штучного інтелекту, що використовується для інтелектуального аналізу даних має свій функціонал, можливості та відповідно сферу використання, які необхідно враховувати при виборі інструментарію для аналізу даних [30]. Також важливо орієнтуватись на відкритість або часткову відкритість системи, що може давати переваги для більш досвідчених користувачів у питаннях доопрацювання системи чи більш складного налаштування моделей, коли користувач має навички написання програмного коду.

Кожен з перелічених засобів має свої специфічні недоліки, що в певних ситуаціях можуть впливати на ефективність та швидкість обробки інформації, створення та аналіз моделей, їх візуалізацію та прозорість.

Серед основних недоліків часто зустрічається закритість систем, висока вартість, недостатня кількість документації чи підтримки. Також зустрічається обмежена функціональність для специфічних задач та високі вимоги до масштабування.

Вибір інструменту повинен залежати від типу джерела даних для аналізу, вимог до необхідності масштабування чи постійного оновлення, а також від фінансових можливостей замовників.



ВИСНОВКИ ТА ПЕРСПЕКТИВИ ПОДАЛЬШИХ ДОСЛІДЖЕНЬ

В роботі було досліджено засоби штучного інтелекту для інтелектуального аналізу даних та можливості їх використання для розв'язання проблем, що виникають при роботі з різними джерелами даних. Було визначено проблеми, які зазвичай стоять перед дослідниками при роботі з різними типами даних, їх різноманітним способом представлення та виділення для створення відповідної моделі. Було виконано огляд основних алгоритмів інтелектуального аналізу даних та їх інтеграції із сучасними платформами та застосунками. Також досліджено основні найпопулярніші інструменти штучного інтелекту для Data Mining, виявлено їх переваги та недоліки, ступінь відкритості, сфери застосування та можливості на різних етапах моделювання.

Вдалий вибір інструментарію для аналітики визначає успіх роботи організації. Галузі освіти та науки, комунікації, фінансові та корпоративні дослідження, покладаються на засоби аналізу даних для роботи з великими об'ємами інформації, для моделювання, класифікації та прогнозування.

Кожен із інструментів штучного інтелекту, що розглянуто, представляють гарні та реальні рішення, мають унікальні функції, сильні та слабкі сторони та здатні розв'язувати комплексні задачі. Оскільки ШІ продовжує активно розвиватись, ці інструменти, безумовно, є невід'ємною частиною аналітики та моделювання даних в організаціях, на виробництві та в бізнесі.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Kharkevych, M. Yu. (2023). *Artificial intelligence: Theory and practice*. Lira. (In Ukrainian)
2. Kuzmenko, I. V. (2022). *Intelligent data analysis*. KhNURE. (In Ukrainian)
3. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning*. Springer. <https://hastie.su.domains/ElemStatLearn/>
4. Bishop, C. (2023). *Pattern recognition and machine learning*. Springer. <https://link.springer.com/book/10.1007/978-0-387-31073-2>
5. Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques*. Elsevier. <https://www.sciencedirect.com/book/9780128146262>
6. Géron, A. (2023). *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow*. O'Reilly. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>
7. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2021). *Data mining: Practical machine learning tools & techniques*. Morgan Kaufmann. <https://www.elsevier.com/books/data-mining/>
8. IBM. (n.d.). *IBM SPSS Modeler documentation*. <https://www.ibm.com/docs/en/spss-modeler>
9. RapidMiner. (n.d.). *RapidMiner documentation*. <https://docs.rapidminer.com>
10. KNIME. (n.d.). *KNIME documentation*. <https://docs.knime.com>
11. TensorFlow. (n.d.). *TensorFlow developers guide*. <https://www.tensorflow.org/guide>
12. PyTorch. (n.d.). *PyTorch documentation*. <https://pytorch.org/docs/stable/>
13. Microsoft. (n.d.). *Power BI documentation*. <https://learn.microsoft.com/en-us/power-bi/>
14. Chen, Z., & Chan, J. (2024). *AI tools for data analytics*. IEEE. <https://ieeexplore.ieee.org>
15. Elsevier. (2023). *Deep learning architectures*. <https://www.sciencedirect.com/book/9780128197400>
16. Sydorenko, T. V. (2024). *Development of intelligent systems in industry*. <https://eastjournal.org>
17. OpenAI. (2025). *OpenAI research reports*. <https://openai.com/research>
18. Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach*. Pearson. <https://aima.cs.berkeley.edu>
19. Chollet, F. (2021). *Deep learning with Python*. Manning. <https://www.manning.com/books/deep-learning-with-python>
20. Jordon, J. (2023). *Machine learning for healthcare*. Cambridge University Press. <https://www.cambridge.org/9781108953988>



21. Provost, F., & Fawcett, T. (2021). *Data science for business*. O'Reilly. <https://www.oreilly.com/library/view/data-science-for/9781449374273/>
22. Raj, A. (2022). *Big data analytics*. Springer. <https://link.springer.com/book/10.1007/978-3-030-38040-7>
23. Zhang, C. (2023). *Machine learning algorithms*. CRC Press. <https://www.crcpress.com/Machine-Learning-Algorithms/9781138035454>
24. Liavynetz, H., et al. (2024). Knowledge-oriented intelligent technologies in adaptive management of the hospitality industry. *Economics and Society*. <https://economyandsociety.in.ua/index.php/journal/article/download/4775/4716> (In Ukrainian)
25. Vynnychuk, R. (2024). Intelligent data analysis (data mining) in the food industry. *Graal of Science*. <https://www.researchgate.net/publication/384673396> (In Ukrainian)
26. Application of artificial intelligence and data-mining methods in communication processes. (2023). *Scientific Bulletin of Karazin KhNU*. <https://ekhnuir.karazin.ua/items/0ce249bc-45a4-4a56-9748-8cd2d066e41a> (In Ukrainian)
27. Lozovska, L. (2023). Analysis of machine learning methods in web analytics. *Economic Development*. <https://economdevelopment.in.ua/index.php/journal/article/view/848> (In Ukrainian)
28. Artificial intelligence technologies in managing the efficiency of industrial enterprises. (2024). *Information Technologies*. <https://journals.mipolytech.in.ua/index.php/tech/article/view/87> (In Ukrainian)
29. Bondarenko, Yu. (2022). Machine learning in financial monitoring. *Financial Research*. <https://frjournal.com/2022/03/bondarenko> (In Ukrainian)
30. Sydorenko, P. (2024). Intelligent decision-support systems in production. <https://engineeringjournal.org.ua> (In Ukrainian)
31. Levchenko, K. (2023). Optimization of business processes using machine learning. *Business-Inform*. <https://business-inform.net> (In Ukrainian)
32. Smirnov, O., Fedorov, E., Neskoro dieva, A., & Neskoro dieva, T. (2024). Intellectual classification method of gymnastic elements based on combinations of descriptive and generative approaches. *CEUR Workshop Proceedings*, 3664, 11–23. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85191419623>
33. Smirnov, O., Lakhno, V., Akhmetov, B., Chubaievskiy, V., Khorolska, K., & Bebesko, B. (2023). Selection of a rational composition of information protection means using a genetic algorithm. In G. Rajakumar et al. (Eds.), *Intelligent communication technologies and virtual mobile networks* (Vol. 131, pp. 21–34). Springer. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85134768958>
34. Smirnova, T. V., Konoplytska-Slobodeniuk, O. K., Buravchenko, K. O., Smirnov, S. A., Kravchuk, O. V., Kozirova, N. L., & Smirnov, O. A. (2024). Research into technologies for ensuring cybersecurity of cloud services IaaS, PaaS and SaaS. *Cybersecurity: Education, Science, Technology*, 4(24), 6–27. <https://doi.org/10.28925/2663-4023.2024.24.627>
35. Al-Mudhafar Akil Abdulhussein, M., Smirnova, T. V., Buravchenko, K. O., & Smirnov, O. A. (2023). Method for evaluating and improving the user experience of subscribers in software-defined networks based on the use of machine learning. *Modern Information Systems*, 7(2), 49–56. <https://doi.org/10.20998/2522-9052.2023.2.07>
36. Smirnov, O. A., Polihenko, O. O., Odarchenko, R. S., Tereshchenko, L. Yu., & Usik, P. S. (2020). Information technology and software for increasing the efficiency of planning the subsystem of cellular base stations. *Problems of Telecommunications*, 1(26), 83–96. <https://pt.nure.ua/articles/informacijna-tehnologiya-ta-programne-zabezpechennya-dlya-pidvishhennya-efektivnosti-planuvannya-pidsistemi-bazovih-stancij-stilnikovogo-zv-yazku/>
37. Smirnov, O. A., Smirnova, T. V., Polishchuk, L. I., Buravchenko, K. O., & Makevnnin, A. O. (2020). Research on cloud technologies as services. *Cybersecurity: Education, Science, Technology*, 3(7), 43–62. <https://www.csecurity.kubg.edu.ua/index.php/journal/article/view/132/132>



Iryna Lysenko

Candidate of Science (Engineering), Senior Lecturer of Cybersecurity & Software Academic Department
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

ORCID: 0000-0003-4394-4960

min_max@i.ua

Roman Minailenko

Candidate of Science (Engineering), Associate Professor of Cybersecurity & Software Academic Department
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

ORCID: 0009-0000-0563-0798

aron70@ukr.net

Serhii Smirnov

Candidate of Science (Engineering), associate professor, associate professor of Cybersecurity & Software Academic Department

Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

ORCID: 0000-0002-7649-7442

smirnov.ser.81@gmail.com

Kostiantyn Buravchenko

Candidate of Science (Engineering), associate professor of Cybersecurity & Software Academic Department
Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

ORCID: 0000-0001-6195-7533

buravchenkok@gmail.com

Nataliia Yakymenko

candidate of Physical and Mathematical Sciences, associate professor, associate professor of Cybersecurity & Software Academic Department

Central Ukrainian National Technical University, Kropivnitskiy, Ukraine

ORCID: 0000-0002-4498-0093

yakimenko_n_m@ukr.net

Oleksii Smirnov

Doctor of technical sciences, professor, head of Cybersecurity & Software Academic Department

Central Ukrainian National Technical University, Kropyvnytskyi, Ukraine

ORCID: 0000-0001-9543-874X

dr.smirnovoa@gmail.com

RESEARCH OF ARTIFICIAL INTELLIGENCE TOOLS FOR INTELLIGENT DATA ANALYSIS

Abstract. The article presents a comprehensive study of modern artificial intelligence (AI) tools designed for intelligent data analysis (Data Mining), their computational capabilities, and areas of practical application. The key data analysis algorithms are examined, including machine learning methods, deep learning, clustering, classification, regression modeling, and neural networks. Special attention is given to tools such as Scikit-learn, TensorFlow, RapidMiner, Google AutoML, Power BI, and other platforms and frameworks that provide automation of data processing, analysis, and visualization of large-scale datasets. The study includes a comparative analysis of the advantages and limitations of the most widely used AI tools in terms of model accuracy, computational efficiency, ease of integration, scalability, AutoML support, and the ability to work with unstructured data. Examples of applying AI tools in industry, economics, medicine, the financial sector, and other fields of human activity are provided. The research considers current trends in the development of intelligent data analysis, taking into account the growing role of cloud platforms, automated model-building systems, multimodal AI models, and integration with corporate analytical systems. The results of the study make it possible to determine the most effective approaches and tools for solving applied Data Mining tasks, ensuring the selection of technologies according to requirements for accuracy, performance, and openness. The findings may be used in scientific research, business analytics, digital transformation of enterprises, and the design of intelligent decision-support systems.



Keywords: artificial intelligence, data mining, machine learning, neural networks, data analysis algorithms, Data Mining, TensorFlow, Scikit-learn, RapidMiner, Power BI, AutoML.

REFERENCES

1. Kharkevych, M. Yu. (2023). *Artificial intelligence: Theory and practice*. Lira. (In Ukrainian)
2. Kuzmenko, I. V. (2022). *Intelligent data analysis*. KhNURE. (In Ukrainian)
3. Hastie, T., Tibshirani, R., & Friedman, J. (2021). *The elements of statistical learning*. Springer. <https://hastie.su.domains/ElemStatLearn/>
4. Bishop, C. (2023). *Pattern recognition and machine learning*. Springer. <https://link.springer.com/book/10.1007/978-0-387-31073-2>
5. Han, J., Kamber, M., & Pei, J. (2022). *Data mining: Concepts and techniques*. Elsevier. <https://www.sciencedirect.com/book/9780128146262>
6. Géron, A. (2023). *Hands-on machine learning with Scikit-Learn, Keras & TensorFlow*. O'Reilly. <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>
7. Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2021). *Data mining: Practical machine learning tools & techniques*. Morgan Kaufmann. <https://www.elsevier.com/books/data-mining/>
8. IBM. (n.d.). *IBM SPSS Modeler documentation*. <https://www.ibm.com/docs/en/spss-modeler>
9. RapidMiner. (n.d.). *RapidMiner documentation*. <https://docs.rapidminer.com>
10. KNIME. (n.d.). *KNIME documentation*. <https://docs.knime.com>
11. TensorFlow. (n.d.). *TensorFlow developers guide*. <https://www.tensorflow.org/guide>
12. PyTorch. (n.d.). *PyTorch documentation*. <https://pytorch.org/docs/stable/>
13. Microsoft. (n.d.). *Power BI documentation*. <https://learn.microsoft.com/en-us/power-bi/>
14. Chen, Z., & Chan, J. (2024). *AI tools for data analytics*. IEEE. <https://ieeexplore.ieee.org>
15. Elsevier. (2023). *Deep learning architectures*. <https://www.sciencedirect.com/book/9780128197400>
16. Sydorenko, T. V. (2024). *Development of intelligent systems in industry*. <https://eastjournal.org>
17. OpenAI. (2025). *OpenAI research reports*. <https://openai.com/research>
18. Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach*. Pearson. <https://aima.cs.berkeley.edu>
19. Chollet, F. (2021). *Deep learning with Python*. Manning. <https://www.manning.com/books/deep-learning-with-python>
20. Jordon, J. (2023). *Machine learning for healthcare*. Cambridge University Press. <https://www.cambridge.org/9781108953988>
21. Provost, F., & Fawcett, T. (2021). *Data science for business*. O'Reilly. <https://www.oreilly.com/library/view/data-science-for/9781449374273/>
22. Raj, A. (2022). *Big data analytics*. Springer. <https://link.springer.com/book/10.1007/978-3-030-38040-7>
23. Zhang, C. (2023). *Machine learning algorithms*. CRC Press. <https://www.crcpress.com/Machine-Learning-Algorithms/9781138035454>
24. Liavnyez, H., et al. (2024). Knowledge-oriented intelligent technologies in adaptive management of the hospitality industry. *Economics and Society*. <https://economyandsociety.in.ua/index.php/journal/article/download/4775/4716> (In Ukrainian)
25. Vynnychuk, R. (2024). Intelligent data analysis (data mining) in the food industry. *Graal of Science*. <https://www.researchgate.net/publication/384673396> (In Ukrainian)
26. Application of artificial intelligence and data-mining methods in communication processes. (2023). *Scientific Bulletin of Karazin KhNU*. <https://ekhnuir.karazin.ua/items/0ce249bc-45a4-4a56-9748-8cd2d066e41a> (In Ukrainian)
27. Lozovska, L. (2023). Analysis of machine learning methods in web analytics. *Economic Development*. <https://economdevelopment.in.ua/index.php/journal/article/view/848> (In Ukrainian)
28. Artificial intelligence technologies in managing the efficiency of industrial enterprises. (2024). *Information Technologies*. <https://journals.mipolytech.in.ua/index.php/tech/article/view/87> (In Ukrainian)
29. Bondarenko, Yu. (2022). Machine learning in financial monitoring. *Financial Research*. <https://frjournal.com/2022/03/bondarenko> (In Ukrainian)
30. Sydorenko, P. (2024). Intelligent decision-support systems in production. <https://engineeringjournal.org.ua> (In Ukrainian)
31. Levchenko, K. (2023). Optimization of business processes using machine learning. *Business-Inform*. <https://business-inform.net> (In Ukrainian)



32. Smirnov, O., Fedorov, E., Neskrodieva, A., & Neskrodieva, T. (2024). Intellectual classification method of gymnastic elements based on combinations of descriptive and generative approaches. *CEUR Workshop Proceedings*, 3664, 11–23. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85191419623>
33. Smirnov, O., Lakhno, V., Akhmetov, B., Chubaievskiy, V., Khorolska, K., & Bebesko, B. (2023). Selection of a rational composition of information protection means using a genetic algorithm. In G. Rajakumar et al. (Eds.), *Intelligent communication technologies and virtual mobile networks* (Vol. 131, pp. 21–34). Springer. <https://www.scopus.com/record/display.uri?eid=2-s2.0-85134768958>
34. Smirnova, T. V., Konoplytska-Slobodeniuk, O. K., Buravchenko, K. O., Smirnov, S. A., Kravchuk, O. V., Kozirova, N. L., & Smirnov, O. A. (2024). Research into technologies for ensuring cybersecurity of cloud services IaaS, PaaS and SaaS. *Cybersecurity: Education, Science, Technology*, 4(24), 6–27. <https://doi.org/10.28925/2663-4023.2024.24.627>
35. Al-Mudhafar Akil Abdulhussein, M., Smirnova, T. V., Buravchenko, K. O., & Smirnov, O. A. (2023). Method for evaluating and improving the user experience of subscribers in software-defined networks based on the use of machine learning. *Modern Information Systems*, 7(2), 49–56. <https://doi.org/10.20998/2522-9052.2023.2.07>
36. Smirnov, O. A., Polihenko, O. O., Odarchenko, R. S., Tereshchenko, L. Yu., & Usik, P. S. (2020). Information technology and software for increasing the efficiency of planning the subsystem of cellular base stations. *Problems of Telecommunications*, 1(26), 83–96. <https://pt.nure.ua/articles/informacijna-tehnologiya-ta-programne-zabezpechennya-dlya-pidvishhennya-efektivnosti-planuvannya-pidsistemi-bazovih-stancij-stilnikovogo-zv-yazku/>
37. Smirnov, O. A., Smirnova, T. V., Polishchuk, L. I., Buravchenko, K. O., & Makevnin, A. O. (2020). Research on cloud technologies as services. *Cybersecurity: Education, Science, Technology*, 3(7), 43–62. <https://www.csecurity.kubg.edu.ua/index.php/journal/article/view/132/132>

