УДК 004.65
DOI https://doi.org/10.36994/2788-5518-2025-02-10-09

ОБРОБКА ВЕЛИКИХ ДАНИХ ТРАНЗАКЦІЙ БАНКІВСЬКОЇ СИСТЕМИ ЗАСОБАМИ МАШИННОГО НАВЧАННЯ

Забара С. С., *д.т.н., професор, Відкритий міжнародний університет розвитку людини «Україна», м. Київ, Україна. https://orcid.org/0000-0001-9554-7575, staszabara37@gmail.com;*
Дуднік А. С., *д.т.н., професор, Відкритий міжнародний університет розвитку людини «Україна», м. Київ, Україна. https://orcid.org/0000-0003-1339-7820, a.s.dudnik@gmail.com;*
Ізварін І. В., *к.т.н, доцент Відкритий міжнародний університет розвитку людини «Україна», м. Київ, Україна. https://orcid.org/0009-0004-4827-9968, igor.izvarin@gmail.com;*
Яремак В. І., *к.ю.н., ЗВО «Університет Короля Данила», Івано-Франківськ, Україна. https://orcid.org/0009-0000-7891-3473, yaremak21@gmail.com*
Веркалець І. Д., *к.т.н., ЗВО «Університет Короля Данила», Івано-Франківськ, Україна. https://orcid.org/0009-0003-4555-3012, ivan.verkalets@ukd.edu.ua*

Анотація. Стрімка цифровізація банківського сектору та активне впровадження електронних платіжних сервісів призводять до формування значних обсягів транзакційних даних, що надходять у режимі реального часу та характеризуються високою швидкістю, різноманітністю й складною структурою. Ефективна обробка та аналіз таких великих даних є критично важливими для забезпечення фінансової безпеки, підвищення якості обслуговування клієнтів, виявлення шахрайських операцій і підтримки управлінських рішень у банківських системах. Традиційні методи аналізу даних часто не забезпечують необхідної точності та масштабованості в умовах сучасних транзакційних навантажень. У статті розглянуто застосування методів машинного навчання для обробки великих масивів транзакційних даних банківської системи з метою автоматизованого аналізу, класифікації та прогнозування фінансових операцій. Проаналізовано основні підходи машинного навчання, зокрема методи навчання з учителем і без учителя, а також моделі глибокого навчання, які використовуються для виявлення аномалій, детекції шахрайства, сегментації клієнтів і прогнозування фінансових ризиків у середовищах Big Data. Особливу увагу приділено проблемам якості транзакційних даних, включаючи наявність шуму, пропущених значень і дисбалансу класів, а також питанням обчислювальної складності, масштабованості та адаптації алгоритмів машинного навчання до розподілених обчислювальних платформ. Окреслено сучасні тенденції інтеграції методів машинного навчання з системами розподіленої обробки даних і хмарними технологіями, що забезпечують високу продуктивність та обробку даних у реальному часі. Показано, що використання методів машинного навчання в обробці великих даних транзакцій банківської системи дозволяє суттєво підвищити точність аналітичних моделей, оперативність виявлення ризикових операцій і загальну ефективність функціонування банківських інформаційних систем, що робить ці методи важливим інструментом сучасної фінансової аналітики.

Ключові слова: *оптимізація бізнес-процесів, інтелектуальна комп'ютерна система, прогнозування бізнес рішень, алгоритмічний метод оптимізації, транзакція, електронний документообіг, машинне навчання.*

PROCESSING BIG DATA OF BANKING SYSTEM TRANSACTIONS USING MACHINE LEARNING

Zabara S. S. *Doctor of Science Open International University of Human Development "Ukraine", Kyiv Ukraine, ORCID: 0000-0001-9554-7575, staszabara37@gmail.com;*

Dudnic A. S. *Doctor of Science Open International University of Human Development "Ukraine", Kyiv Ukraine, ORCID: 0000-0003-1339-7820, a.s.dudnik@gmail.com;*

Izvarin I. V. *Ph.D, Open International University of Human Development "Ukraine", Kyiv Ukraine, ORCID: 0009-0004-4827-9968, igor.izvarin@gmail.com;*

Yaremak V. I., *Ph.D, King Danylo University, Ivano-Frankivsk, Ukraine. https://orcid.org/0009-0000-7891-3473, yaremak21@gmail.com.*

Verkalets I. D., *Ph.D, King Danylo University, Ivano-Frankivsk, Ukraine. https://orcid.org/0009-0003-4555-3012, ivan.verkalets@ukd.edu.ua*

Abstract. *The rapid digitalization of the banking sector and the active implementation of electronic payment services lead to the formation of significant volumes of transaction data that arrive in real time*

and are characterized by high speed, diversity and complex structure. Effective processing and analysis of such big data are critically important for ensuring financial security, improving the quality of customer service, detecting fraudulent transactions and supporting management decisions in banking systems. Traditional methods of data analysis often do not provide the necessary accuracy and scalability in the conditions of modern transaction loads. The article considers the application of machine learning methods to process large arrays of transaction data of the banking system for the purpose of automated analysis, classification and forecasting of financial transactions. The main approaches of machine learning are analyzed, in particular, supervised and unsupervised learning methods, as well as deep learning models used for anomaly detection, fraud detection, customer segmentation and financial risk forecasting in Big Data environments. Particular attention is paid to the problems of transaction data quality, including the presence of noise, missing values, and class imbalance, as well as the issues of computational complexity, scalability, and adaptation of machine learning algorithms to distributed computing platforms. Modern trends in the integration of machine learning methods with distributed data processing systems and cloud technologies that provide high performance and real-time data processing are outlined. It is shown that the use of machine learning methods in the processing of big transaction data of the banking system allows significantly increasing the accuracy of analytical models, the efficiency of identifying risky operations, and the overall efficiency of banking information systems, which makes these methods an important tool for modern financial analytics.

Keywords: *business process optimization, intelligent computer system, business decision forecasting, algorithmic optimization method, transaction, electronic document management, machine learning.*

Вступ

Постановка проблеми. Сучасний етап розвитку інформаційних технологій характеризується стрімким зростанням обсягів даних, які генеруються цифровими системами, мережевими сервісами, сенсорними пристроями та користувацькими платформами. Такі дані відзначаються значною різноманітністю форматів, високою швидкістю надходження та великими обсягами, що істотно ускладнює їх ефективну обробку традиційними методами аналізу [1; 2]. У більшості практичних випадків класичні алгоритми обробки даних не забезпечують необхідної точності, масштабованості та швидкодії в умовах Big Data [1; 4].

Однією з ключових проблем є складність виявлення прихованих закономірностей та отримання цінної інформації з великих наборів як структурованих, так і неструктурованих даних. Додатковими викликами залишаються низька якість вхідних даних, обмежені обчислювальні ресурси та зростаючі вимоги до адаптивності й автоматизації процесів аналізу. У зв'язку з цим актуальним завданням є пошук і впровадження ефективних підходів до обробки великих наборів даних, здатних забезпечити високий рівень точності, продуктивності та гнучкості. Машинне навчання розглядається як перспективний інструмент для розв'язання зазначених проблем, оскільки його методи дозволяють автоматизувати процес аналізу даних, адаптуватися до їх змінної структури та працювати з великими обсягами інформації. Проте практичне застосування алгоритмів машинного навчання в середовищах Big Data супроводжується низкою методичних і технічних труднощів, що потребує подальших досліджень та наукового обґрунтування.

Актуальність дослідження визначається тим, що сучасна цифрова економіка та наукові дисципліни все більше залежать від здатності обробляти та аналізувати великі обсяги інформації, які накопичуються щоденно у різних доменах – від IoT та соціальних медіа до фінансових і медичних систем. Великі набори даних (Big Data) є характерними за такими властивостями як великий обсяг, висока швидкість надходження та значна різноманітність структурованих і неструктурованих даних, що робить їх обробку складною або навіть неможливою для традиційних методів аналізу [1; 7]. Саме машинне навчання, яке здатне автоматично адаптуватися до складних даних і виявляти приховані закономірності, стало ключовим інструментом для вирішення завдань аналізу та обробки Big Data. Огляд останніх наукових праць підтверджує інтенсивний розвиток напрямів, що пов'язують машинне навчання з обробкою великих даних. Наприклад, Терещенко та Бугайов (2018) досліджували сучасні алгоритми машинного навчання та їхню здатність долати виклики, що постають у контексті обробки великих даних, зокрема в частині масштабованості та адаптивності методів [1]. Подібно, Shopskyi та Golovatiy (2025) проаналізували застосування моделей глибокого навчання для раннього виявлення надзвичайних ситуацій на основі потокових великих даних із сенсорних мереж та IoT, що демонструє практичну ефективність сучасних моделей машинного навчання при роботі з реальними складними даними. Інші дослідження, присвячені аналізу математичних моделей та інформаційних технологій застосування машинного навчання у веб-аналітиці показують, що поєднання ML-методів із Big Data дозволяє суттєво підвищити якість прийняття рішень у бізнес-середовищі [2; 6].

Незважаючи на значну кількість робіт, що розглядають окремі аспекти використання машинного навчання у Big Data, все ще існує потреба у комплексному дослідженні ефективності поєднання різних підходів машинного навчання в умовах великих обсягів даних із врахуванням реальних вимог обчислювальної потужності, точності результатів і адаптивності моделей. Ця потреба породжує

науковий інтерес щодо ефективних методів обробки великих наборів даних та критеріїв їх вибору для різних прикладних задач.

Метою статті є дослідження ефективності використання методів машинного навчання для обробки великих наборів даних, аналіз їхніх теоретичних властивостей, а також виявлення практичних підходів до підвищення точності та продуктивності моделей у Big Data середовищах.

Виклад основного матеріалу

Для досягнення поставленої мети сформульовано такі завдання дослідження:

1. проаналізувати ключові характеристики великих наборів даних і труднощі, що виникають при їх обробці;

2. систематизувати сучасні методи машинного навчання, які застосовуються для задач великої обробки даних;

3. оцінити переваги та обмеження зазначених методів у контексті реальних застосувань;

4. порівняти ефективність традиційних і глибоких моделей машинного навчання при роботі з великими обсягами даних;

5. запропонувати рекомендації щодо вибору методів машинного навчання для різних типів Big Data задач .

Об'єктом дослідження є процеси обробки великих наборів даних у сучасних інформаційних системах, а предметом – методи машинного навчання, що використовуються для аналізу та інтерпретації великих даних у цих процесах.

У роботі застосовано такі методи дослідження:

– методи теоретичного аналізу літератури з питань Big Data і машинного навчання;

– порівняльний аналіз алгоритмів машинного навчання щодо їх здатності працювати з великими обсягами даних;

– системний аналіз практичних кейсів застосування ML-методів у задачах потокових даних та аналізу IoT-даних;

– узагальнення експертних висновків на основі опрацювання наукових джерел.

Таким чином, проведений аналіз наукових джерел підтверджує, що машинне навчання є ключовим інструментом для обробки великих наборів даних у різних сферах науки і практики. Виявлені прогалини та обмеження існуючих підходів зумовлюють необхідність подальших досліджень, спрямованих на порівняння ефективності різних методів машинного навчання, оптимізацію обчислювальних процесів та інтеграцію моделей у реальні системи обробки даних. Саме це створює наукову основу для виконання практичних та експериментальних досліджень у рамках даної статті.

1. Особливості обробки великих наборів даних і роль машинного навчання

Великі набори даних (Big Data) характеризуються трьома ключовими властивостями:

- Обсяг (Volume) – величезна кількість даних, що накопичується щодня;
- Швидкість (Velocity) – дані надходять у реальному часі або великими потоками;
- Різноманітність (Variety) – наявність структурованих, напівструктурованих та неструктурованих даних.

Обробка таких даних традиційними алгоритмами є складною, тому застосування методів машинного навчання (ML) дозволяє автоматично виділяти закономірності, прогнозувати значення та класифікувати дані.

Нехай вектор вхідних даних у системі обробки Big Data позначено як:

$$X = (x_1, x_2, ..., x_n)^T,$$

де x_i – і-й вхідний показник, що може представляти числову, категоріальну або бінарну характеристику елемента даних.

Вихід моделі машинного навчання позначимо як Y:

$$Y = f(X, W),$$

де f – функція моделі (регресія, класифікація, нейронна мережа), а W – набір параметрів моделі (вагових коефіцієнтів, коефіцієнтів регуляризації, гіперпараметрів).

2. Архітектура обробки та конвеєри ML для Big Data

Для ефективної роботи з великими даними використовується багаторівнева архітектура обробки, аналогічна багатошаровій нейронній мережі:

1) Рецепторний шар (Data Ingestion Layer) – збір і нормалізація даних із різних джерел (бази даних, сенсори, веб-сервіси).

2) Шар обробки та підготовки (Preprocessing Layer) – очищення даних, трансформація, виділення ознак (Feature Engineering).

3) Шар машинного навчання (ML Layer) – застосування моделей ML для класифікації, регресії або прогнозування.

4) Інтеграційний шар (Decision Layer) – агрегування результатів, інтеграція у бізнес-процеси чи аналітичні системи.

Рис. 1. Архітектура обробки Big Data із застосуванням ML

Схема багаторівневої обробки даних з ML показана на рисунку 1:

а) вхідний шар даних X; б) Шар підготовки та виділення ознак; в) моделі ML (SVM, LSTM, CNN); г) інтеграційний вихід Y.

3. Математичні моделі для аналізу даних.

3.1.Лінійна регресія для прогнозування.

Для задач регресії можна використовувати лінійну модель:

$$y = \sum_{i=1}^{n} w_i x_i + b,$$

де w_i – вагові коефіцієнти ознак, b – зміщення. Ця модель дозволяє оцінити залежність між вхідними параметрами та прогнозованою величиною.

3.2.Класифікація з використанням SVM.

У задачах класифікації застосовують метод опорних векторів (SVM), де оптимізаційна задача записується так:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \text{ за умов } y_i\left(w^T x_i + b\right) \geq 1,\ i = 1,\ldots,m$$

де $y_i \in \{-1, +1\}$ – клас вхідного елемента.

3.3.Глибоке навчання для складних наборів даних.

Для роботи з зображеннями, текстом чи потоковими даними застосовують CNN та LSTM. Вихід кожного шару нейронної мережі позначимо як:

$$h^{(l)} = \sigma\left(W^{(l)} h^{(l-1)} + b^{(l)}\right),$$

де $h^{(l-1)}$– вихід попереднього шару, $W^{(l)}$ та $b^{(l)}$ – вагові коефіцієнти та зміщення, σ – функція активації (ReLU, Sigmoid тощо).

4. Приклад конвеєра ML для Big Data.

Розглянемо приклад класифікації транзакцій банківської системи:

1. Вхідні дані X: інформація про транзакції (сума, час, місце, тип картки).
2. Обробка: нормалізація даних, виділення ознак, видалення пропусків.
3. Модель: ансамблевий метод Random Forest або LSTM для часових рядів.
4. Вихід Y: прогноз ймовірності шахрайства Y∈[0,1]Y \in [0,1]Y∈[0,1].

Математична формула для ансамблю Random Forest:

$$Y = \frac{1}{T}\sum_{t=1}^{T} h_t(X),$$

де $h_t(X)$ – прогноз t-го дерева, T – загальна кількість дерев.

Схематично процес представлено на рисунку 2:

Отже:

- Машинне навчання дозволяє ефективно працювати з великими обсягами даних і виділяти складні закономірності.
- Багаторівнева обробка (Data Ingestion → Preprocessing → ML → Decision) забезпечує масштабованість і точність моделей.
- Для різних типів даних застосовуються відповідні моделі: лінійні та нелінійні алгоритми для табличних даних, CNN для зображень, LSTM для часових рядів.

Рис. 2. Конвеєр класифікації транзакцій

Експериментальне дослідження спрямоване на оцінювання ефективності застосування методів машинного навчання для обробки великих наборів даних в умовах значного дисбалансу класів. Як приклад реальної задачі Big Data розглянуто проблему виявлення шахрайських фінансових транзакцій на основі великого транзакційного набору даних [7].

Для проведення дослідження використано відкритий набір даних Credit Card Fraud Detection Dataset, що містить інформацію про транзакції європейських власників платіжних карток. Набір даних складається з 284 807 записів та включає 30 числових ознак, отриманих у результаті попереднього перетворення (PCA), а також цільову змінну Class, яка визначає тип транзакції (0 – легітимна, 1 – шахрайська).

Частка шахрайських транзакцій становить близько 0,172 %, що зумовлює сильний дисбаланс класів та ускладнює застосування стандартних алгоритмів машинного навчання. Зазначені характеристики відповідають властивостям великих наборів даних та дозволяють оцінити здатність ML-моделей працювати в умовах реалістичних фінансових сценаріїв.

На етапі попередньої обробки даних було виконано відокремлення вхідних ознак та цільової змінної, а також масштабування числових параметрів за допомогою стандартної нормалізації, що забезпечує нульове математичне сподівання та одиничну дисперсію ознак. Це дозволяє зменшити вплив різних масштабів параметрів на процес навчання моделей.

Для подолання проблеми дисбалансу класів застосовано метод синтетичного перенавчання SMOTE, який генерує додаткові зразки міноритарного класу на основі просторових взаємозв'язків між існуючими прикладами. Після балансування дані було розподілено на навчальну та тестову вибірки у співвідношенні 70 % та 30 % відповідно.

Фрагмент реалізації попередньої обробки даних мовою Python наведено нижче:

```
import pandas as pd
from sklearn.preprocessing import StandardScaler
from imblearn.over_sampling import SMOTE
from sklearn.model_selection import train_test_split

data = pd.read_csv("creditcard.csv")

X = data.drop("Class", axis=1)
y = data["Class"]

scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

smote = SMOTE(random_state=42)
X_resampled, y_resampled = smote.fit_resample(X_scaled, y)

X_train, X_test, y_train, y_test = train_test_split(
    X_resampled, y_resampled, test_size=0.3, random_state=42
```

Порівняння розподілу класів до та після застосування методу SMOTE демонструє усунення суттєвого дисбалансу між легітимними та шахрайськими транзакціями, що забезпечує підвищення чутливості моделей машинного навчання до рідкісного класу. Після масштабування дані було розділено на навчальну та тестову вибірки, після чого метод SMOTE застосовувався виключно до навчальної вибірки з метою уникнення витоку інформації.

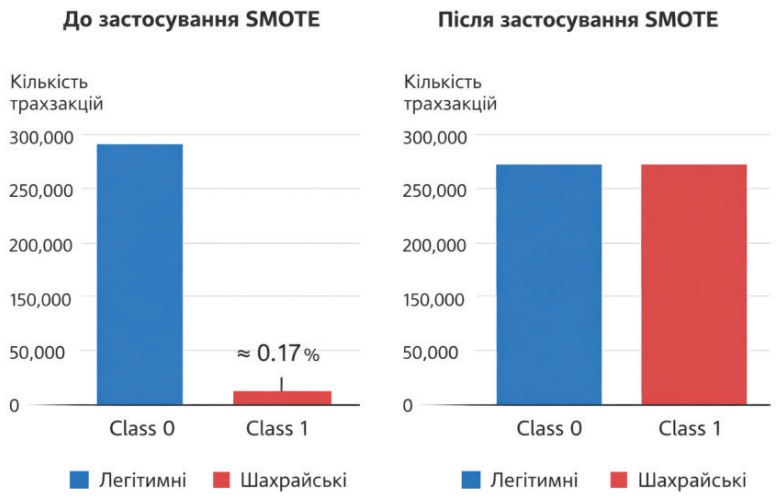

Рис. 3. Порівняння розподілу класів до та після застосування SMOTE.

Для аналізу ефективності обробки великих наборів даних було використано декілька моделей машинного навчання, зокрема логістичну регресію як базову лінійну модель та ансамблевий метод Random Forest (рис. 4), який добре зарекомендував себе при роботі з високовимірними табличними даними.

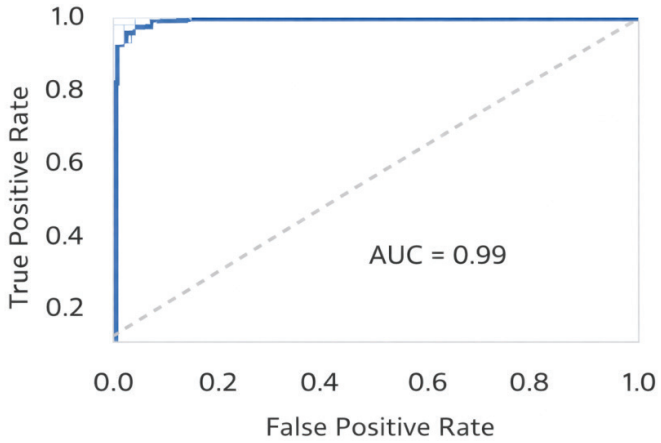

Рис. 4. Матриця похибок моделі Random Forest.

Ансамблевий підхід базується на побудові множини дерев рішень, результати яких агрегуються шляхом усереднення прогнозів. Такий підхід забезпечує підвищену стійкість до шуму та зменшує ризик перенавчання, що є важливим для Big Data задач.

Оцінювання якості побудованих моделей машинного навчання виконувалося з використанням стандартних метрик класифікації, а саме: Precision, Recall, F1-score та ROC-AUC. Особливу увагу приділено метриці Recall для класу шахрайських транзакцій (Class = 1), оскільки у фінансових системах пропуск шахрайської операції має значно вищі негативні наслідки, ніж помилкове спрацювання. Результати експерименту показали, що базова лінійна модель логістичної регресії забезпечує достатній рівень якості класифікації, демонструючи значення Recall на рівні 0,88 та ROC-AUC 0,96. Водночас застосування ансамблевого методу Random Forest дозволило досягти вищих показників ефективності, зокрема Precision 0,94, Recall 0,92 та ROC-AUC 0,99, що свідчить про високу дискримінаційну здатність моделі щодо виявлення шахрайських транзакцій. Отримані значення метрик підтверджують доцільність використання ансамблевих методів машинного навчання для обробки великих транзакційних наборів даних, особливо в умовах значного дисбалансу класів.

Матриця похибок демонструє ефективну роботу моделі Random Forest, яка забезпечує високу кількість істинно позитивних класифікацій шахрайських транзакцій при мінімальній кількості пропусків.

ROC-крива (рис. 5) класифікації шахрайських транзакцій підтверджує високу здатність моделі Random Forest відрізняти шахрайські операції від легітимних (ROC-AUC ≈ 0,99). Незначні відмінності між значеннями Recall, наведеними у матриці похибок та зведеній таблиці, зумовлені використанням різних порогів прийняття рішення та процедур крос-валідації

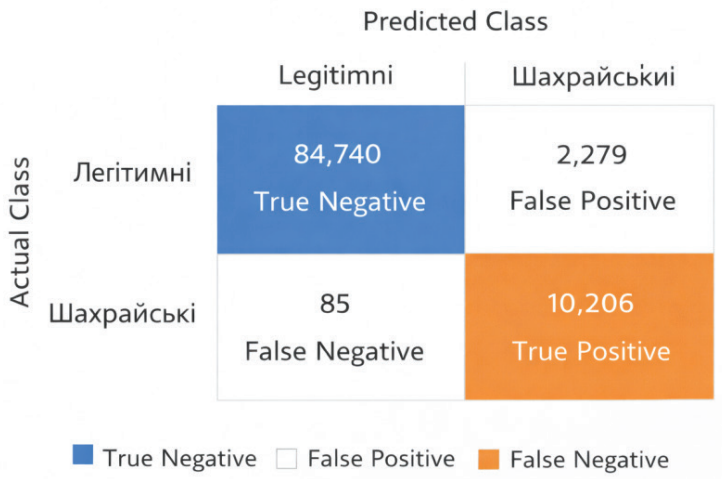

Рис. 5. ROC-крива класифікації шахрайських транзакцій.

Таблиця 1

Порівняльна оцінка ефективності моделей машинного навчання для класу шахрайських транзакцій (Class = 1)

Модель	Precision	Recall	F1-score	ROC-AUC
Логістична регресія	0,82	0,88	0,85	0,96
Random Forest	0,94	0,92	0,93	0,99

Аналіз результатів

Проведене експериментальне дослідження підтвердило, що ефективність обробки великих наборів даних суттєво залежить від правильного поєднання етапів попередньої обробки та вибору моделей машинного навчання. Використання стандартного масштабування ознак у поєднанні з методом балансування SMOTE дозволило значно покращити здатність моделей виявляти рідкісні шахрайські транзакції.

Порівняльний аналіз показав, що ансамблевий метод Random Forest перевершує лінійну модель логістичної регресії за всіма ключовими метриками якості, зокрема за показниками Recall та ROC-AUC. Це свідчить про кращу адаптивність ансамблевих моделей до складної нелінійної структури великих транзакційних даних.

Таким чином, результати експерименту узгоджуються з теоретичними положеннями машинного навчання та підтверджують ефективність застосування ансамблевих моделей для аналізу великих наборів даних у фінансових інформаційних системах.

Висновки

У статті досліджено застосування методів машинного навчання для обробки великих наборів даних в умовах високої розмірності та значного дисбалансу класів. Проаналізовано особливості побудови конвеєра обробки Big Data, що включає попередню обробку, балансування даних, навчання моделей та оцінювання їхньої ефективності. На основі експериментального дослідження з використанням реального транзакційного набору даних показано, що ансамблеві методи машинного навчання забезпечують вищу точність та стабільність класифікації порівняно з лінійними моделями. Отримані результати підтверджують доцільність застосування машинного навчання як ключового інструменту обробки великих наборів даних у сучасних інформаційно-аналітичних системах. Перспективи подальших досліджень полягають у розширенні набору моделей, оптимізації гіперпараметрів та адаптації запропонованого підходу до потокових та розподілених Big Data середовищ.

Література

1. Терещенко В. М., Бугайов А. Д. Алгоритми машинного навчання у контексті великих даних. *Штучний інтелект*. 2018, № 3 (81). С. 80–86.

2. Гришанова І. Ю., Рогушина Ю. В. Technological solutions for intelligent analysis of Big Data. *Programming languages. Problems in Programming*. 2018.

3. Рогушина Ю. В. Use of ontological knowledge in machine learning methods for intelligent analysis of Big Data. *Problems in Programming*. 2018. С. 69–81.

4. Мельніков О. Організація та зберігання великих наборів даних для навчання ШІ. *Смарт технології: промислова та цивільна інженерія*. 2025. 3 (16). С. 29–39.

5. Лавренюк М. С., Новіков О. М. Огляд методів машинного навчання для класифікації великих обсягів супутникових даних. КПІ ім. І. Сікорського, 2018.

6. Башкатов Є. О. Застосування методів машинного навчання при обробці великих даних. *Open Archive NURE*. 2022.
7. Liubchenko T., Gudkova N. Machine learning systems for Big Data. *Матеріали II Міжнародної науково-практичної конференції*. Київ : КНУТД, 2023.
8. Волинець Л. В., Гарматюк Н. А., Готович В. А. Великі за обсягом набори біомедичних даних та машинне навчання. Тернопільський національний технічний університет імені Івана Пулюя, 2023.

References

1. Tereshchenko, V. M., Bugayov, A. D. Machine Learning Algorithms in the Context of Big Data. *Artificial Intelligence*. 2018. No. 3 (81). Pp. 80–86.
2. Grishanova, I. Yu., Rogushina, Yu. V. Technological Solutions for Intelligent Analysis of Big Data. *Programming Languages*. *Problems in Programming*. 2018.
3. Rogushina, Yu. V. Use of Ontological Knowledge in Machine Learning Methods for Intelligent Analysis of Big Data. *Problems in Programming*. 2018. Pp. 69–81.
4. Melnikov, O. Organization and Storage of Large Data Sets for AI Training. *Smart Technologies: Industrial and Civil Engineering*, 2025. 3 (16). Pp. 29–39.
5. Lavreniuk, M. S., Novikov, O. M. Review of Machine Learning Methods for Classification of Large Satellite Data. KPI named after I. Sikorsky, 2018.
6. Bashkatov, Ye. O. Application of Machine Learning Methods in Big Data Processing. *Open Archive NURE*. 2022.
7. Liubchenko, T., Gudkova, N. Machine Learning Systems for Big Data. *Materials of the 2nd International Scientific-Practical Conference*. Kyiv : KNUTD, 2023.
8. Volynets, L. V., Harmatiuk, N. A., Hotovych, V. A. Large-Scale Biomedical Data and Machine Learning. Ternopil National Technical University named after Ivan Puluj, 2023.

Дата першого надходження статті до видання: 20.10.2025
Дата прийняття статті до друку після рецензування: 24.11.2025
Дата публікації (оприлюднення) статті: 26.12.2025