УДК 004.896:658.012.011.56
DOI https://doi.org/10.32782/3041-2080/2025-5-7

АВТОМАТИЗОВАНА СИСТЕМА ВІДБОРУ ОЗНАК ДЛЯ КОМП'ЮТЕРНО-ІНТЕГРОВАНОГО ПРОГНОЗУВАННЯ УСПІШНОСТІ B2B-ЗАМОВЛЕНЬ З ВИКОРИСТАННЯМ XGBOOST

Мірошниченко Сергій Олександрович,
студент групи 174-24-1дф
ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА»
ORCID ID: 0009-0007-4868-3006

Мірошниченко Вікторія Ігорівна,
кандидат технічних наук, доцент,
доцентка кафедри автоматизації, електро- та робототехнічних систем
ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА»
ORCID ID: 0000-0002-5956-7867

Койфман Олексій Олександрович,
кандидат технічних наук, доцент,
завідувач кафедри автоматизації, електро- та робототехнічних систем
ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА»
ORCID ID: 0000-0003-2075-7417

Вовна Олександр Володимирович,
доктор технічних наук, професор,
професор кафедри автоматизації, електро- та робототехнічних систем
ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА»
ORCID ID: 0000-0003-4433-7097

Інтенсивний розвиток сектору корпоративної електронної торгівлі створює потребу в автоматизованих системах прогнозування для оптимізації керування ланцюгами постачання та мінімізації фінансових ризиків. Точне прогнозування успішності B2B-замовлень є важливим для ефективного керування ресурсами та оптимізації виробничих процесів підприємства. Сучасні ERP-системи накопичують великі обсяги даних про поведінку клієнтів, що створює можливості для застосування методів машинного навчання у процесах прийняття рішень.

Мета дослідження полягає у розробленні оптимальної стратегії автоматизованого відбору ознак для прогнозування успішності B2B-замовлень з використанням алгоритму XGBoost в архітектурі комп'ютерно-інтегрованих систем керування підприємством.

Здійснено експериментальне дослідження ефективності шести методів відбору ознак на двох незалежних наборах даних обсягом 86 794 записи з 24 характеристиками замовлень. Порівняно прямий та зворотний відбір за важливістю XGBoost, жадібні прямий та зворотний алгоритми, рекурсивне виключення ознак та алгоритм Boruta. Автоматизована оптимізація гіперпараметрів здійснена засобами фреймворку Optuna з алгоритмом Tree-structured Parzen Estimator. Оцінювання проводилось за метрикою AUC-PR з 5-кратною крос-валідацією для забезпечення статистичної надійності результатів.

Жадібні алгоритми забезпечили найвищу ефективність класифікації: прямий відбір досягнув AUC-PR 0,97873, зворотний – 0,97785. Встановлено, що оптимальний набір ознак включає 16 характеристик. Це відповідає зменшенню розмірності на 33 %, що сприяє підвищенню прогностичної якості. Комплексний підхід забезпечив зниження помилкових класифікацій на 2,7–3,7 % порівняно з базовими налаштуваннями та дав змогу ідентифікувати дев'ять критично важливих ознак.

Отримані результати створюють методологічну основу для розроблення автоматизованих систем прогнозування в сучасних комп'ютерно-інтегрованих виробничих комплексах.

Ключові слова: *автоматизований відбір ознак, градієнтний бустинг, XGBoost, B2B-прогнозування, машинне навчання, зменшення розмірності, AUC-PR.*

Miroshnychenko Serhii, Miroshnychenko Viktoriia, Koyfman Oleksiy, Vovna Oleksandr. Automated feature selection system for computer-integrated B2B order success prediction using XGBoost

Contemporary business process automation, involving implementation of information technologies for optimizing enterprise operations, creates demand for intelligent forecasting systems to manage supply chains and minimize financial risks. Accurate prediction of corporate order success is critical for efficient resource management and production process optimization. Modern ERP systems accumulate large volumes of customer behavior data, creating opportunities for machine learning applications in decision-making processes.

This research develops an optimal automated feature selection strategy for B2B order success prediction using XGBoost within computer-integrated enterprise management systems.

An experimental study evaluated six feature selection methods on two independent datasets containing 86.794 records with 24 order characteristics. Methods compared included forward and backward XGBoost importance-based selection, greedy forward and backward algorithms, recursive feature elimination, and Boruta algorithm. Automated hyperparameter optimization was implemented using Optuna framework with Tree-structured Parzen Estimator. Evaluation employed AUC-PR metric with 5-fold stratified cross-validation for statistical reliability.

Greedy algorithms achieved highest classification efficiency: forward selection reached AUC-PR of 0.97873, backward selection achieved 0.97785. Optimal feature set size was established at 16 characteristics, representing 33 % dimensionality reduction while improving predictive quality. The comprehensive approach reduced misclassifications by 2.7–3.7 % compared to baseline configurations and identified nine critically important prediction features.

Results establish methodological foundation for developing automated forecasting systems in computer-integrated manufacturing complexes, ensuring improved accuracy while optimizing computational resources and reducing model complexity for practical implementation in enterprise environments.

Key words: *automated feature selection, XGBoost, B2B order prediction, computer-integrated systems, business process automation, machine learning optimization, hyperparameter tuning, greedy algorithms.*

Вступ. Стрімке зростання електронної комерції, зокрема у сегменті Business-to-Business (B2B), трансформує традиційні моделі взаємодії між постачальниками та клієнтами, ускладнюючи керування ланцюгами постачання й запасами [1]. Точне прогнозування для B2B є важливим для стратегічного планування, ефективного керування ресурсами та оптимізації операцій, оскільки дає змогу компаніям оперативно реагувати на динамічні ринкові тенденції [2]. Оптимізація виробничих процесів підприємства вимагає забезпечення адекватних запасів продукції, координації виробничих циклів та ефективного керування постачанням сировини. Ключовою умовою досягнення цих цілей є прогнозування обсягів продукції та сировини на складських потужностях. Згідно з джерелом [3], прогнозування складських запасів традиційно базується на затверджених замовленнях. Однак значний часовий лаг між ініціацією та затвердженням замовлень створює необхідність використання моделей машинного навчання для прогнозування ймовірності успішної реалізації замовлень на основі історичних даних. Точне прогнозування ймовірності успішного виконання замовлення дає змогу скоротити фінансові ризики, мінімізувати надлишкові запаси та підвищити рівень обслуговування клієнтів [4].

Алгоритм Extreme Gradient Boosting [5] зарекомендував себе як один з найефективніших методів для подібних завдань [6; 7], однак його продуктивність значною мірою залежить від релевантності набору ознак і налаштувань моделі [8–10]. Сучасні ERP-системи накопичують величезні обсяги даних про поведінку клієнтів, характеристики замовлень та результати бізнес-процесів. Високовимірний простір ознак призводить до «прокляття розмірності», збільшує час тренування, ускладнює інтерпретацію та підвищує ризик перенавчання. Це створює необхідність відбору найбільш інформативних ознак [11; 12].

Вибір оптимального методу відбору ознак визначає ефективність моделей бінарної класифікації [13], особливо в контексті високовимірних даних B2B-замовлень. Фундаментальна проблема відбору ознак полягає в тому, що кількість можливих підмножин ознак зростає експоненціально з кількістю доступних ознак [14; 15]. Недоцільність та виключну ресурсоємність повного перебору ознак підкреслено в роботах [16–18]. Chen та колеги підкреслюють, що «визначення ідеальної підмножини ознак зі списку можливостей є комбінаторною проблемою, яка не може бути вирішена за високої розмірності без використання специфічних припущень або компромісів» [19]. Ця експоненціальна складність робить необхідним використання евристичних алгоритмів для знаходження субоптимальних, але практично прийнятних рішень. Автори [20–22] демонструють, що різні підходи до відбору ознак мають специфічні переваги залежно від характеристик даних та цільових метрик оптимізації.

Метою дослідження є визначення оптимальної стратегії відбору ознак для прогнозування успішності B2B-замовлень з використанням XGBoost в поєднанні з автоматизованою оптимізацією гіперпараметрів.

Для досягнення поставленої мети визначено такі завдання дослідження:

– порівняти ефективність шести різних методів відбору ознак;

– визначити оптимальний розмір набору ознак для досліджуваних даних;

– ідентифікувати найбільш інформативні характеристики B2B-замовлень;

– розробити практичні рекомендації для впровадження у виробничі системи.

Методи та методики дослідження. Дослідження базується на двох незалежних наборах історичних даних про успішність B2B-замовлень. Для забезпечення репрезентативності та збалансованості аналізу розміри обох датасетів було стандартизовано до 86 794 записів, що містять інформацію про замовлення з 24 характеристиками (ознаками). Цільова змінна (is_successful) характеризує успішність замовлення: дорівнює 1 у випадку схваленого та успішного виконаного, або 0 в іншому випадку. Розподіл цільових класів: 54 763 успішних замовлення (клас 1) та 32 031 неуспішне замовлення (клас 0), що відповідає співвідношенню приблизно 63:37.

Досліджувані дані мають таку структуру:

1) характеристики замовлення: кількість повідомлень у замовленні (order_messages), сума замовлення (order_amount), кількість змін у замовленні (order_changes), кількість позицій (order_lines_count), джерело замовлення (source), менеджер, що обробляє замовлення (salesperson), знижка в замовленні (discount_total);

2) темпоральні ознаки: вік замовлення від початку діяльності компанії в місяцях (create_date_months), квартал, місяць, день тижня та година доби, коли зроблено замовлення (quarter, month, day_of_week, hour_of_day);

3) агреговані показники клієнта: відсоток успішних замовлень у клієнта (partner_success_rate), загальна кількість замовлень клієнта (partner_total_orders), строк співпраці з клієнтом в днях (partner_order_age_days), загальна кількість повідомлень клієнта (partner_total_messages), середня сума в замовленнях клієнта (partner_avg_amount), середня кількість змін у замовленнях у клієнта (partner_avg_changes), середні значення сум, повідомлень, змін для успішних та неуспішних замовлень клієнта (partner_success_avg_amount, partner_fail_avg_amount, partner_success_avg_messages, partner_fail_avg_messages, partner_success_avg_changes, partner_fail_avg_changes).

Відсутні дані оброблено шляхом заповнення медіаною для забезпечення стійкості до екстремальних значень [23]. Негативні значення замінено на нуль у випадках, коли від'ємні величини не мають фізичної чи економічної інтерпретації [24; 25]. Викиди виявлено за критерієм 1,5 міжквартильного розмаху (IQR) [26; 27] та оброблено методом вінсоризації [28] для збереження всіх спостережень за мінімізації впливу відхилень. Числові ознаки нормалізовано за допомогою Robust Scaler, що, як показано в джерелі [29], підвищує стабільність і точність моделей порівняно зі стандартним масштабуванням. Циклічні часові індикатори трансформовано синусоїдальними та косинусоїдальними перетвореннями [30], категоріальні змінні закодовано методом OHE Extended Compact [31]. Дисбаланс класів компенсовано параметром scale_pos_weight [32]. Дані розподілено на навчальну та тестову вибірки у співвідношенні 70:30 із стратифікованим розбиттям для збереження пропорції цільової змінної [33]. Генератор псевдовипадкових чисел зафіксовано параметром random_state = 42 для забезпечення повторюваності результатів [34].

Дослідження реалізовано з використанням комплексного підходу, який заснований на принципі послідовного покращення якості моделі через три основні етапи.

На першому етапі проведено базове тестування моделі XGBoost зі стандартними параметрами на повному наборі ознак для встановлення початкового рівня якості.

Другий етап присвячений оптимізації гіперпараметрів за допомогою фреймворку Optuna з алгоритмом Tree-structured Parzen Estimator [35], який показав високу якість оптимізації гіперпараметрів XGBoost [36; 37], та був визначений найбільш ефективним [38]. Простір пошуку включав дев'ять ключових гіперпараметрів XGBoost [39]: кількість дерев (100–1 000), швидкість навчання (0,01–0,35), максимальна глибина (3–12), мінімальна вага дочірнього вузла (1–10), gamma-параметр (0–2,0), параметри субвибірки (0,6–1,0) та регуляризації (L1: 0–1,5; L2: 0,1–3,0). Цільова функція базувалася на 5-кратній крос-валідації [40], з метрикою Area Under Precision-Recall Curve (AUC-PR), що зарекомендувала себе як найбільш придатна для оцінювання ефективності бінарної класифікації у контексті незбалансованих даних у багатьох дослідженнях [41–45]. Оптимізація проводилася протягом 100 ітерацій.

Третій етап включає реалізацію та порівняння шести різних методів відбору ознак з використанням оптимізованих гіперпараметрів, визначених на попередньому етапі.

Для розв'язання поставленої задачі було розроблено програмний комплекс, який реалізує

повний цикл порівняння шести підходів до відбору ознак.

1) Прямий відбір за важливістю XGBoost. Метод, що використовує внутрішні метрики важливості, визначені XGBoost, для ранжування ознак [5]. Алгоритм методу послідовно додає ознаки від найважливішої до найменш важливої на основі показників gain, weight, cover [46]. Актуальні дослідження доводять, що застосування методів відбору ознак, які ґрунтуються на внутрішніх показниках важливості, значно покращує продуктивність моделей машинного навчання [47–50]. Для мереж «розумної» енергетики та смартґридів такий підхід знизив час тренування майже на третину без втрати якості [51]. Попри популярність цього методу завдяки його інтеграції у сам алгоритм XGBoost, його обмеженням є схильність до переоцінювання ознак з великою дисперсією або високою кореляцією. У дослідженні [52] було показано, що поєднання XGBoost з Recursive Feature Elimination та Boruta дає змогу досягти стабільнішої продуктивності, ніж під час застосування лише вагової важливості.

2) Зворотний відбір за важливістю XGBoost. Протилежний попередньому підхід, що починається з повного набору ознак та ітеративно видаляє найменш важливі. Метод ефективно зменшує упередженість традиційних метрик важливості XGBoost, особливо у випадках з корельованими предикторами [53].

3) Жадібний прямий відбір (Greedy Forward Selection, GFS). Обгортковий метод, що починається з порожнього набору та додає ознаку з максимальним покращенням цільової метрики на кожному кроці. Теоретичне обґрунтування конвергенції представлено у класичних роботах [54; 55]. Сучасні варіації методу включають паралельний алгоритм для великих даних [56], метаевристичний підхід для аналізу тексту [57], спеціалізований алгоритм для систем ранжування [58] та модифікації для задач з обмеженим бюджетом [21]. Гібридний підхід з PSO демонструє покращення точності та зменшення підмножини ознак [59]. Основне обмеження полягає у локальній природі пошуку [60].

4) Жадібний зворотний відбір (Greedy Backward Elimination, GBE). Стартує з повного набору та видаляє ознаку з найменшим впливом на продуктивність [61]. Ефективний для оптимізації моделей з високою початковою розмірністю, як продемонстровано для прогнозування деформації незв’язаних агрегатів [62].

5) Рекурсивне виключення ознак (Recursive Feature Elimination, RFE). Обгортковий алгоритм, що поєднує зворотне видалення з повторним перенавчанням моделі [63]. Систематизований для геномної класифікації метод демонструє автоматичне усунення надлишковості та забезпечення компактних піднаборів ознак [64]. У проведеному дослідженні використовувалася реалізація цього алгоритму з бібліотеки scikit-learn [65]. Порівняльні дослідження [19; 20; 66] демонструють конкурентну ефективність RFE відносно інших методів, тоді як RFE з перехресною валідацією забезпечує стійкість до перенавчання у системах виявлення вторгнень [67].

6) Алгоритм Boruta. Обгортковий метод для виявлення всіх релевантних ознак через порівняння з тіньовими копіями, згенерованими випадковою пермутацією [68]. У представленому дослідженні використана Python-реалізація, сумісна з ensemble-методами scikit-learn [69]. Ефективність цього алгоритму підтверджена дослідженнями у медичній діагностиці [66], класифікації даних [19; 20; 70], прогнозування діабету [71] та радіоміки [72], у фінансах для P2P-кредитування [73], кредитного скорингу [74] із забезпеченням 3–4 % приросту точності за скорочення часу інференсу [75].

Для кожного методу формується послідовність підмножин ознак, для яких тренується модель XGBoost з оптимізованими гіперпараметрами та обчислюється значення метрики AUC-PR на тестовій вибірці. Кожен метод тестується на послідовності наборів ознак різного розміру, що дає змогу побудувати криві залежності якості від кількості ознак та визначити оптимальну конфігурацію для кожної стратегії.

Результати. Результати систематичного оцінювання ефективності різних етапів оптимізації показали поступове покращення якості моделей, що є наслідком застосування комплексного підходу до налаштування параметрів та відбору ознак. Базові моделі за замовчуванням досягли значень AUC-PR, рівних 0,97499 для першого датасету та 0,97451 для другого датасету, демонструючи високу інформативність вхідних даних, ефективність процедур попередньої обробки та значну прогностичну здатність XGBoost у базовій конфігурації.

Оптимізація гіперпараметрів за допомогою фреймворку Optuna забезпечила покращення показників на 0,33 % для першого датасету та на 0,25 % для другого датасету (табл. 1).

Наявність різниці між оптимальними значеннями параметрів свідчить про відмінності у структурі та характеристиках датасетів, що потребує індивідуального підходу до налаштування моделей.

Для обох датасетів побудовано криві залежності AUC-PR від кількості ознак для всіх методів

Таблиця 1

Результати оптимізації гіперпараметрів XGBoost

Параметр	Датасет 1	Датасет 2
Початкове AUC-PR	0,97499	0,97451
Покращення AUC-PR	0, 97830 (+0,33 %)	0, 97704 (+0,25 %)
Кількість дерев	750	800
Швидкість навчання	0,02	0,1
Максимальна глибина	12	7
Мінімальна вага дочірнього вузла	2	3
Gamma	0,7	0,2
Subsample	0,97	1,0
Colsample_bytree	0,69	0,7
Reg_alpha	0,9	0,5
Reg_lambda	0,35	2,0

у діапазоні від 12 до 24 ознак (рис. 1). Встановлено, що зниження кількості ознак менше 12 призводить до погіршення якості моделі незалежно від методу відбору.

Аналіз результатів, зокрема максимальних значень AUC-PR для кожного методу та датасету (табл. 2), показує, що найвищі значення цих параметрів досягаються під час використання саме жадібних методів відбору ознак за однакової оптимальної кількості ознак (16 шт.), проте оптимальні типи жадібного алгоритму різняться для обох датасетів. Зокрема, максимальний результат для датасету 1 забезпечує GFS (0,97873), тоді як для датасету 2 найкращим є GBE (0,97785). Це свідчить про чутливість оптимальної стратегії до особливостей вибірки, навіть за схожої структури даних.

Таблиця 2

Максимальні значення AUC-PR (*n*) для кожного методу та датасету (*n* – кількість ознак в оптимальному наборі)

Метод	AUC-PR (n)	
	Датасет 1:	Датасет 2:
Прямий XGBoost	0,97841 (19)	0,97740 (24)
Зворотній XGBoost	0,97849 (20)	0,97767 (14)
Прямий Greedy	0,97873 (16)	0,97764 (22)
Зворотній Greedy	0,97870 (23)	0,97785 (16)
RFE	0,97841 (19)	0,97752 (16)
Boruta	0,97830 (22)	0,97712 (19)

За результатами дослідження частоти включення ознак у найкращі набори, тобто такі, що забезпечують максимальні значення AUC-PR, виявлено набір ключових характеристик замовлень, які стабільно використовуються всіма методами для обох датасетів (рис. 2). До них належать наступні ознаки: order_messages, create_date_months, order_amount, order_changes, order_lines_count, source, partner_success_avg_amount, salesperson, partner_success_avg_messages. Як

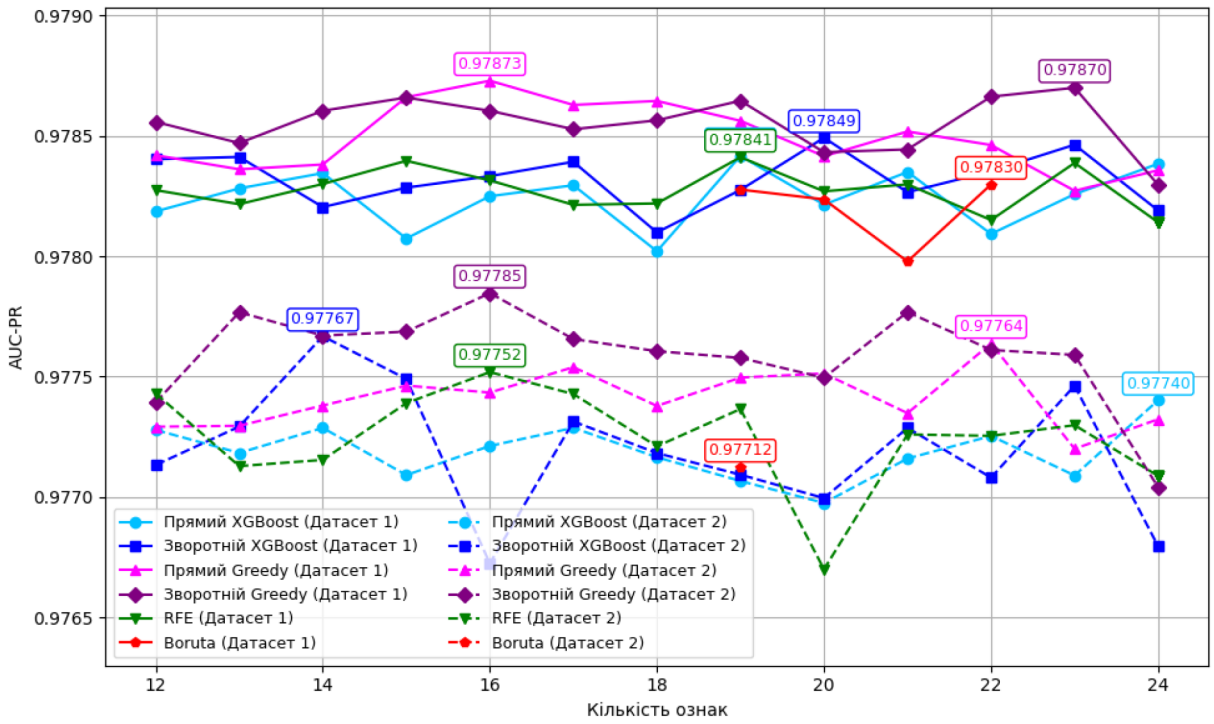

Рис. 1. Порівняння AUC-PR кривих для обох датасетів

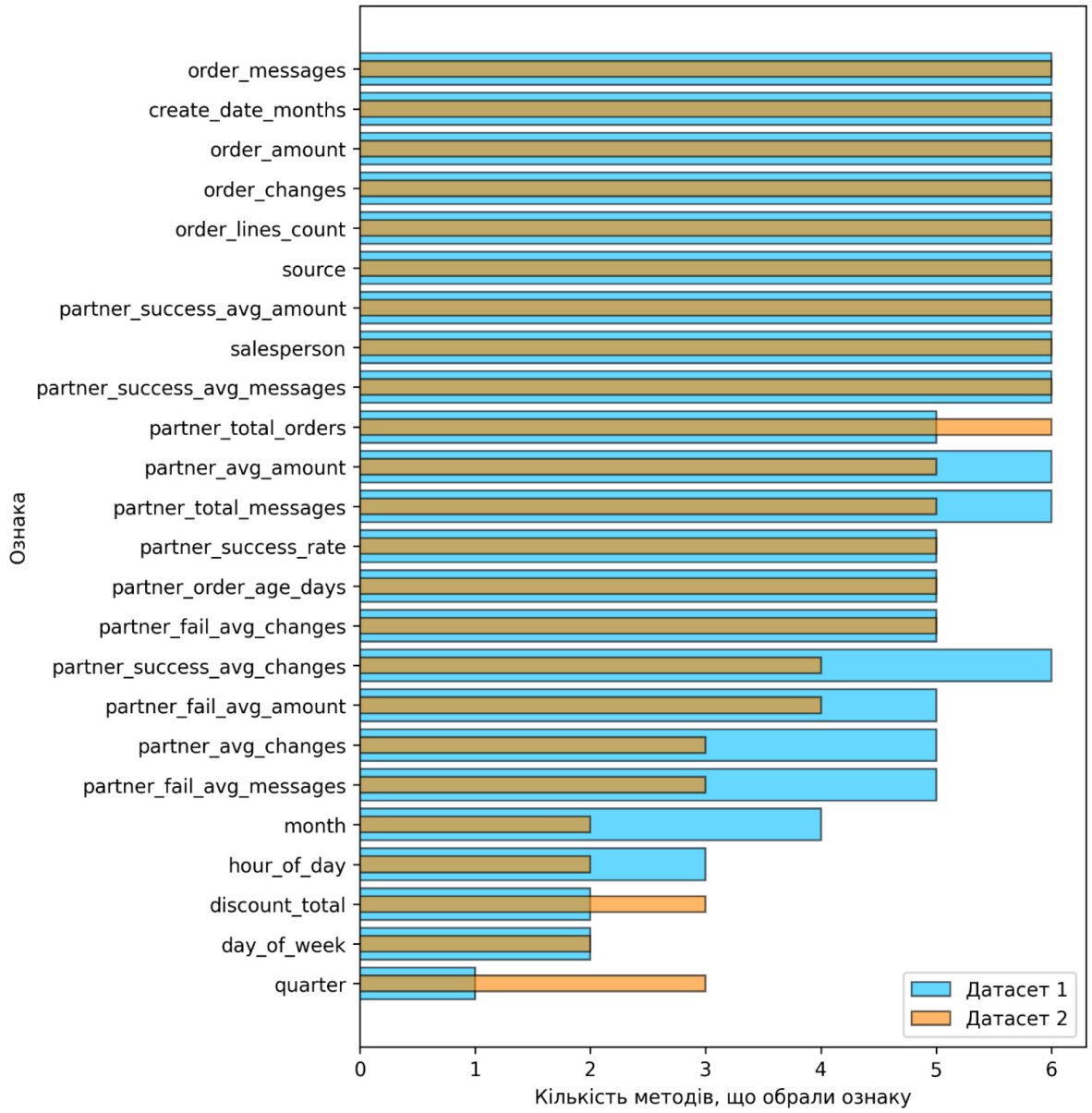

Рис. 2. Частота включення методами ознак у найкращі набори (по методах)

можна бачити на рис. 2, ці дев'ять ознак демонструють максимальну частоту включення усіма методами та високу прогностичну цінність незалежно від специфіки конкретного датасету.

Теплова карта (рис. 3) демонструє, що більшість ключових ознак стабільно включається у найкращі набори для обох датасетів (рис. 3, зелений). Проте для окремих ознак спостерігаються відмінності: деякі ознаки вибираються лише для одного з датасетів (рис. 3, блакитний або помаранчевий).

Аналіз частоти використання ознак (рис. 3) показує, що більшість характеристик (17 з 24) включається у найкращі набори принаймні чотирма з шести методів, підтверджуючи стабільність ключових ознак для прогнозування успішності B2B-замовлень.

До найкращого набору ознак, який забезпечив максимальне значення цільової метрики для обох досліджених наборів даних, належать такі: order_messages, order_amount, order_changes, partner_success_rate, partner_success_avg_amount, partner_fail_avg_amount, partner_total_messages, partner_success_avg_messages, order_lines_count, discount_total, salesperson, source, create_date_months, hour_of_day. Водночас до зазначеного набору для датасету 1 додатково віднесені partner_success_avg_changes та partner_avg_amount, до набору для датасету 2 – partner_total_orders та quarter.

Результати порівняльного аналізу демонструють високий ступінь збіжності оптимальних ознак між досліджуваними наборами даних. З 16 ознак в кожному наборі 14 (87,5 %)

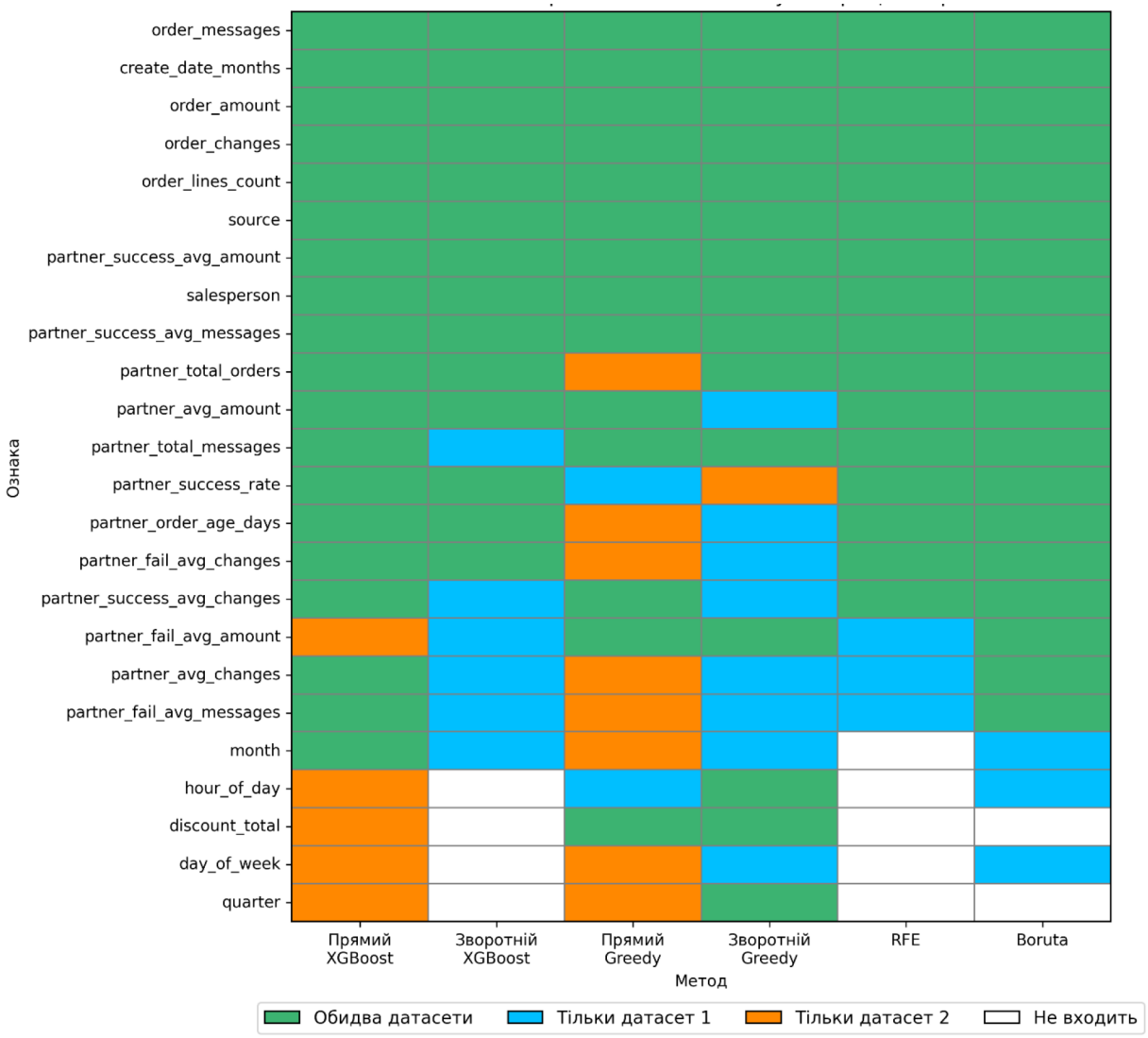

Рис. 3. Теплова карта включення ознак у найкращі набори

є спільними для обох конфігурацій, що вказує на стабільність найважливіших предикторів, схожість структури бізнес-процесів та надійність методів відбору ознак. Більшість впливовіших параметрів виявилася пов'язаною з тривалістю, якістю та інтенсивністю взаємодії з клієнтом. Ці результати узгоджуються з бізнес-логікою B2B-транзакцій, де довіра, історія співпраці та інтенсивність комунікації є ключовими факторами успіху, що підтверджується дослідженнями у сфері B2B електронної комерції [76].

Порівняльний аналіз стратегій моделювання на обох датасетах підтвердив ефективність запропонованого комплексного підходу, проте з різними оптимальними конфігураціями (табл. 3).

Модель для датасету 1 виявляє вищу чутливість до оптимізації гіперпараметрів з покращенням AUC-PR на 0,33 % порівняно з приростом лише 0,25 % для датасету 2. Водночас процедура відбору ознак забезпечує покращення цільової метрики на 0,08 % для датасету 2 та на 0,04 % для датасету 1 відповідно. Це підтверджує гіпотезу про взаємозв'язок між оптимальними параметрами моделі та набором ознак [77–79], а також підкреслює важливість врахування специфіки конкретного датасету. Також отримані результати відображають вплив автоматизованого відбору ознак на якість моделей градієнтного бустингу для B2B-прогнозування. Це узгоджується з наявними дослідженнями, які підкреслюють важливість відбору ознак для підвищення точності, ефективності та інтерпретованості моделей машинного навчання [1; 80–82].

Оптимальні стратегії відбору ознак виявилися специфічними для кожного датасету: прямий жадібний алгоритм забезпечив найкращі результати для датасету 1, тоді як для датасету 2 оптимальним став зворотний жадібний підхід. Це підкреслює, що оптимальний метод відбору ознак може відрізнятися залежно від

Таблиця 3

Метрики тестування моделей

Датасет	Стратегія	Accuracy	Precision	Recall	F1-міра	ROC AUC	AUC-PR
1	Усі ознаки (стандартні налаштування)	0,92903	0,92178	0,96981	0,94519	0,96631	0,97499
	Усі ознаки + Optuna	0,93110	0,93102	0,96208	0,94630	0,969744	0,97830
	Найкращі ознаки (GFS, 16) + Optuna	0,93168	0,93149	0,96251	0,94674	0,97016	0,97873
2	Усі ознаки (стандартні налаштування)	0,92934	0,92123	0,97103	0,94548	0,96667	0,97451
	Усі ознаки + Optuna	0,93053	0,92269	0,97127	0,94636	0,96863	0,97704
	Найкращі ознаки (GBE, 16) + Optuna	0,93126	0,92292	0,97224	0,94694	0,96912	0,97785

характеристик конкретного набору B2B-даних [83; 84]. Слід також відзначити, що обидва алгоритми відбору конвергували до однакової кількості ознак (16), що може свідчити про існування оптимального балансу між інформативністю та складністю моделі для цієї предметної сфери.

Матриці помилок (рис. 4) показують суттєве зниження кількості помилкових оцінювань за одночасного зниження розмірності даних на 33 %.

Для першого датасету кількість помилок зменшилася з 1 848 до 1 779 (покращення на 3,7 %), для другого – з 1 840 до 1 790 помилок (покращення на 2,7 %).

Отримані результати мають важливі практичні наслідки для розроблення систем прогнозування у B2B-сфері. Зменшення набору ознак на 33 %, що призводить до полегшення аналізу та пояснення результатів бізнес-користувачам [85–87], за одночасного зниження помилкових оцінок на 2,7–3,7 % робить запропонований підхід привабливим для практичного впровадження.

Висновки. Проведене дослідження на двох незалежних датасетах демонструє значну

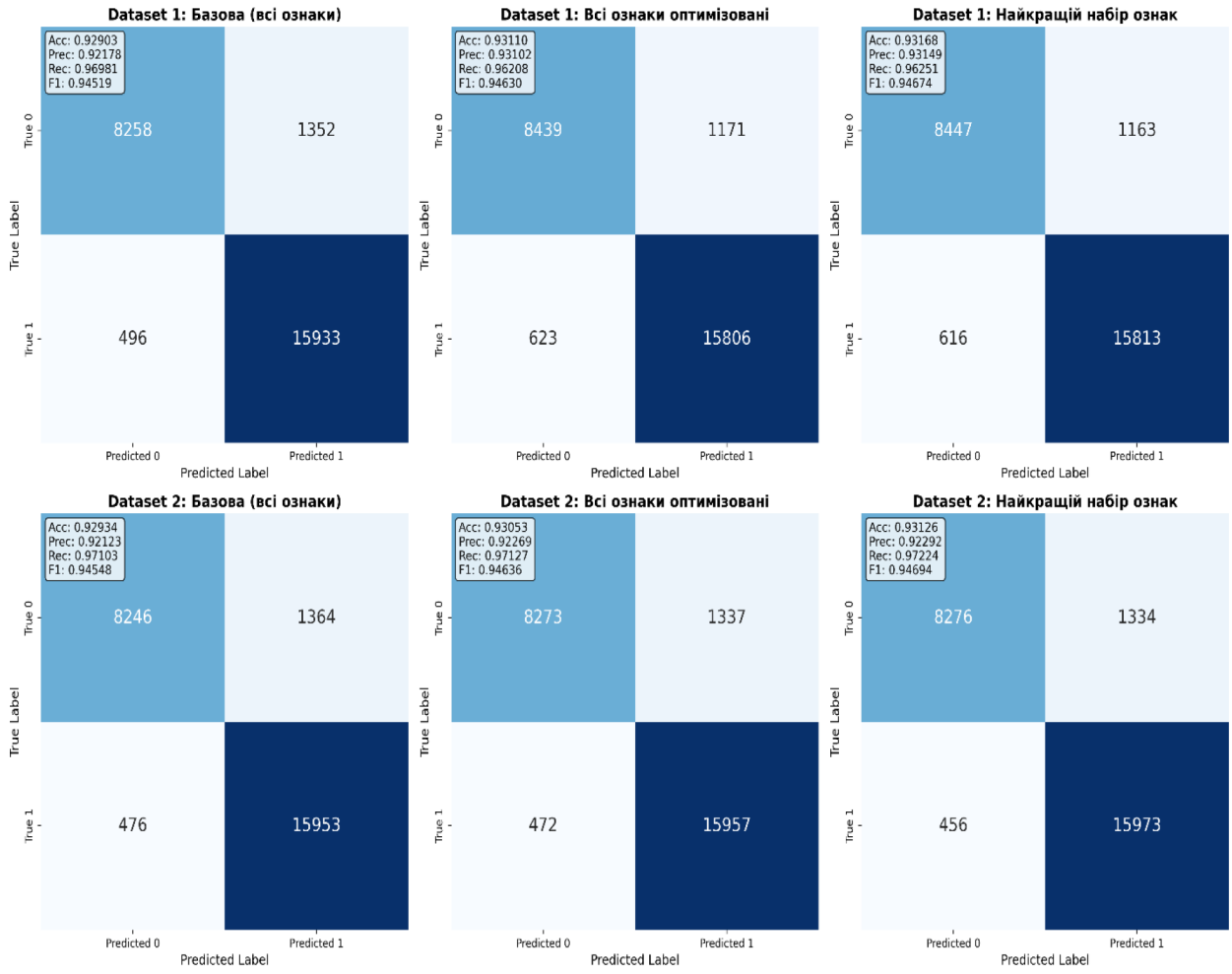

Рис. 4. Матриці помилок для датасетів

ефективність комплексного підходу до оптимізації моделей машинного навчання для прогнозування успішності B2B-замовлень.

Жадібні методи показали себе як найбільш результативні на обох досліджуваних датасетах, проте з відмінностями в оптимальних стратегіях. Дослідження виявило, що оптимальна стратегія відбору ознак залежить від специфічних характеристик датасету. Для першого датасету жадібний прямий відбір забезпечив найкращий результат (AUC-PR = 0,97873), тоді як для другого датасету оптимальним виявився жадібний зворотний алгоритм (AUC-PR = 0,97785). Це підкреслює важливість тестування різних підходів навіть для структурно схожих даних.

Оптимізація гіперпараметрів за допомогою алгоритму Optuna забезпечила суттєве покращення якості моделей, проте з різною ефективністю для кожного датасету. Для першого датасету виявлено більшу чутливість до оптимізації гіперпараметрів (покращення на 0,33 %), тоді як для другого більший ефект від процедури відбору ознак (покращення на 0,08 % проти 0,04 %). Загальний синергетичний ефект поєднання відбору ознак та оптимізації гіперпараметрів забезпечив покращення на 0,37 % для першого датасету та 0,33 % для другого за зменшення розмірності на 33 %. Шістнадцять ознак виявилися оптимальним розміром набору для обох датасетів, незважаючи на різні алгоритми відбору. Це свідчить про існування фундаментального балансу між інформативністю та складністю моделі для предметної сфери B2B-класифікації. Практична ефективність запропонованого підходу підтверджується зниженням кількості помилкових класифікацій: з 1 848 до 1 779 помилок (3,7 %) для першого датасету та з 1 840 до 1 790 помилок (2,7 %) для другого.

Визначено дев'ять найбільш інформативних ознак, які стабільно були вибрані всіма методами для обох датасетів: order_messages, create_date_months, order_amount, order_changes, order_lines_count, source, partner_success_avg_amount, salesperson, partner_success_avg_messages. Загалом 14 з 16 ознак збігаються між оптимальними наборами датасетів, що вказує на універсальність ключових характеристик бізнес-процесів.

Рекомендації для практичного впровадження включають використання жадібних методів вибору ознак з обов'язковим порівнянням прямого та зворотного підходів для вибору найкращого за показниками цільової метрики.

Впровадження автоматизованої оптимізації гіперпараметрів у виробничі системи має бути адаптивним до специфіки конкретних датасетів.

Особливу увагу слід приділити покращенню якості збирання та оброблення дев'яти виявлених найважливіших ознак, а також регулярній переоцінці релевантності ознак під час зміни бізнес-процесів.

Обмеження дослідження включають використання обмеженої кількості датасетів з однієї предметної сфери, оцінювання ефективності тільки для XGBoost-алгоритму, а також дослідження шести методів відбору ознак. Виявлена залежність оптимальної стратегії від характеристик датасету може зменшити узагальнюваність результатів для інших типів B2B-даних.

Подальші дослідження повинні включати тестування на більшій кількості різноманітних датасетів з B2B-платформ для підтвердження універсальності виявлених закономірностей, порівняння з іншими алгоритмами машинного навчання, дослідження додаткових методів відбору ознак, включно з гібридними підходами, а також розроблення автоматизованих систем вибору оптимальної стратегії на основі характеристик датасету.

Отримані результати створюють підґрунтя для розвитку більш ефективних та адаптивних систем прогнозування у сфері B2B електронної комерції та демонструють важливість комплексного підходу до оптимізації моделей машинного навчання з урахуванням специфіки конкретних даних.

ЛІТЕРАТУРА:

1. De T. S., Singh P., Patel A.A Machine learning and Empirical Bayesian Approach for Predictive Buying in B2B E-commerce. *ICMLSC 2024: 2024 The 8th International Conference on Machine Learning and Soft Computing*, Singapore Singapore. New York, NY, USA, 2024. URL: https://doi.org/10.1145/3647750.3647754

2. Li K. A Sales Prediction Method Based on XGBoost Algorithm Model. *BCP Business & Management*. 2023. Vol. 36. P. 367–371. URL: https://doi.org/10.54691/bcpbm.v36i.3487

3. Miroshnychenko S. O. Improvement of the existing functionality for forecasting finished goods inventory at the enterprise warehouses based on the stock forecasted report ERP ODOO. *MININGMETALTECH 2024 – THE MINING AND METALS SECTOR: INTEGRATION OF BUSINESS, TECHNOLOGY AND EDUCATION. Volume 2*. 2024. P. 48–51. URL: https://doi.org/10.30525/978-9934-26-506-8-133

4. Heino A. New Product Demand Forecasting in Retail: Applying Machine Learning Techniques to Forecast Demand for New Product Purchasing Decisions: extended abstract of Master of Science Thesis. 2021. 59 p.

URL: https://www.semanticscholar.org/paper/Applying-Machine-Learning-Techniques-to-Forecast-Heino/d17599a4ebc651fd29997c4f2f5f10e0b905e9a4.

5. Chen T., Guestrin C. XGBoost. *KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco California USA. New York, NY, USA, 2016. URL: https://doi.org/10.1145/2939672.2939785

6. A gradient boosting classifier for purchase intention prediction of online shoppers / Abdullah-All-Tanvir et al. *Heliyon*. 2023. Vol. 9, no. 4. P. e15163. URL: https://doi.org/10.1016/j.heliyon.2023.e15163

7. Leevy J. L., Hancock J., Khoshgoftaar T. M. Medicare fraud detection: a comparative study on the effectiveness of one-class and binary classification models. *International Journal of Computers and Applications*. 2024. P. 1–6. URL: https://doi.org/10.1080/1206212x.2024.2431874

8. A framework for feature selection through boosting / A. Alsahaf et al. *Expert Systems with Applications*. 2021. P. 115895. URL: https://doi.org/10.1016/j.eswa.2021.115895

9. Attack Classification using Machine Learning on UNSW-NB 15 dataset using XGBoost Feature Selection & Ablation Analysis / N. Pansari et al. *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*, Pune, India, 5–7 April 2024. 2024. URL: https://doi.org/10.1109/i2ct61223.2024.10543523

10. Dou Y., Meng W. Predicting Cancer Classification Based on the Improved XGBoost Algorithms. *2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC)*, Qingdao, China, 13–15 December 2024. 2024. P. 1013–1016. URL: https://doi.org/10.1109/icftic64248.2024.10913271

11. A Review of Feature Selection Methods for Machine Learning-Based Disease Risk Prediction / N. Pudjihartono et al. *Frontiers in Bioinformatics*. 2022. Vol. 2. URL: https://doi.org/10.3389/fbinf.2022.927312

12. Cheng X. A Comprehensive Study of Feature Selection Techniques in Machine Learning Models. *Insights in Computer, Signals and Systems*. 2024. Vol. 1, no. 1. P. 65–78. URL: https://doi.org/10.70088/xpf2b276

13. Noroozi Z., Orooji A., Erfannia L. Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Scientific Reports*. 2023. Vol. 13, no. 1. URL: https://doi.org/10.1038/s41598-023-49962-w

14. Welch W. J. Algorithmic complexity: three NP-hard problems in computational statistics. *Journal of Statistical Computation and Simulation*. 1982. Vol. 15, no. 1. P. 17–25. URL: https://doi.org/10.1080/00949658208810560

15. Fong S., Wong R., Vasilakos A. Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data. *IEEE Transactions on Services Computing*. 2015. P. 1. URL: https://doi.org/10.1109/tsc.2015.2439695

16. Blum A. L., Langley P. Selection of relevant features and examples in machine learning. *Artificial Intelligence*. 1997. Vol. 97, no. 1–2. P. 245–271. URL: https://doi.org/10.1016/s0004-3702(97)00063-5

17. Dash M., Liu H. Feature Selection for Classification. *Intelligent Data Analysis*. 1997. Vol. 1, no. 3. P. 131–156. URL: https://doi.org/10.3233/ida-1997-1302

18. Bertsimas D., King A., Mazumder R. Best subset selection via a modern optimization lens. *The Annals of Statistics*. 2016. Vol. 44, no. 2. P. 813–852. URL: https://doi.org/10.1214/15-aos1388

19. Selecting critical features for data classification based on machine learning methods / R.-C. Chen et al. *Journal of Big Data*. 2020. Vol. 7, no. 1. URL: https://doi.org/10.1186/s40537-020-00327-4 (дата звернення: 11.08.2025).

20. Degenhardt F., Seifert S., Szymczak S. Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics*. 2017. Vol. 20, no. 2. P. 492–503. URL: https://doi.org/10.1093/bib/bbx124

21. Cost-Constrained feature selection in binary classification: adaptations for greedy forward selection and genetic algorithms / R. Jagdhuber et al. *BMC Bioinformatics*. 2020. Vol. 21, no. 1. URL: https://doi.org/10.1186/s12859-020-3361-9

22. Saeys Y., Inza I., Larranaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics*. 2007. Vol. 23, no. 19. P. 2507–2517. URL: https://doi.org/10.1093/bioinformatics/btm344

23. Koca Y. B., Aktepe E. Evaluation of Missing Data Imputation Methods and PCA Techniques for Machine Learning Models in Breast Cancer Diagnosis Using WBCD. *Türk Doğa ve Fen Dergisi*. 2024. URL: https://doi.org/10.46810/tdfd.1460871

24. Arimie C. O., Biu E. O., Ijomah M. A. Outlier Detection and Effects on Modeling. *OALib*. 2020. Vol. 07, no. 09. P. 1–30. URL: https://doi.org/10.4236/oalib.1106619

25. Shved A. V., Davydenko Y. O. Outlier Detection Technique for Heterogeneous Data Using Trimmed-Mean Robust Estimators. *Radio Electronics, Computer Science, Control*. 2022. No. 3. P. 50. URL: https://doi.org/10.15588/1607-3274-2022-3-5

26. Jones P. R. A note on detecting statistical outliers in psychophysical data. *Attention, Perception, & Psychophysics*. 2019. Vol. 81, no. 5. P. 1189–1196. URL: https://doi.org/10.3758/s13414-019-01726-3

27. Haanappel C. P., Voor in ‘t holt A. F. Using the interquartile range in infection prevention and control research. *Infection Prevention in Practice*. 2024. P. 100337. URL: https://doi.org/10.1016/j.infpip.2024.100337

28. MSCI Global Investable Market Value and Growth Indexes Methodology. *Bringing clarity to investment decisions MSCI*. URL: https://www.msci.com/eqb/methodology/meth_docs/MSCI_GIMIVGMethod_Feb2021.pdf

29. Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance / M. M. Ahsan et al. *Technologies*. 2021. Vol. 9, no. 3. P. 52. URL: https://doi.org/10.3390/technologies9030052

30. CyclicalFeatures – 1.8.3. *Feature-engine – 1.8.3*. URL: https://feature-engine.trainindata.com/en/1.8.x/user_guide/creation/CyclicalFeatures.html

31. Categorical Features Transformation with Compact One-Hot Encoder for Fraud Detection in Distributed Environment / I. Ul Haq et al. *Communications in Computer and Information Science*. Singapore, 2019. P. 69–80. URL: https://doi.org/10.1007/978-981-13-6661-1_6

32. Notes on Parameter Tuning – xgboost 3.0.4 documentation. *XGBoost Documentation – xgboost 3.0.4 documentation*. URL: https://xgboost.readthedocs.io/en/stable/tutorials/param_tuning.html#handle-imbalanced-dataset

33. Särndal C.-E., Swensson B., Wretman J. Model Assisted Survey Sampling (Springer Series in Statistics). Springer, 2003. 694 p.

34. Shenouda J., Bajwa W. U. A Guide to Computational Reproducibility in Signal Processing and Machine Learning [Tips & Tricks]. *IEEE Signal Processing Magazine*. 2023. Vol. 40, no. 2. P. 141–151. URL: https://doi.org/10.1109/msp.2022.3217659

35. Florek P., Zagda´nski A. Benchmarking state-of-the-art gradient boosting algorithms for classification. *Computer Science. Machine Learning*. 2023. P. 1–9. URL: https://doi.org/10.48550/arXiv.2305.17094

36. GK., C. AD. S., Kumar R N. Enhanced DNS Cybersecurity: Optimization of XGBoost Using Automated Hyperparameter Tuning via Optuna. *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, Tumkuru, India, 4–5 December 2024. 2024. P. 1–7. URL: https://doi.org/10.1109/icmnwc63764.2024.10872404

37. Comparative analysis of SWAT and SWAT coupled with XGBoost model using Optuna hyperparameter optimization for nutrient simulation: A case study in the Upper Nan River basin, Thailand / C. Pinichka et al. *Journal of Environmental Management*. 2025. Vol. 388. P. 126053. URL: https://doi.org/10.1016/j.jenvman.2025.126053

38. Мірошниченко С. О. Оптимізація гіперпараметрів XGBOOST для інтелектуальних систем прогнозування B2B-замовлень. *Науковий Журнал Метінвест Політехніки. Серія: Технічні науки*. 2025. № 4.

39. XGBoost Parameters – xgboost 3.1.0-dev documentation. *XGBoost Documentation – xgboost 3.0.4 documentation*. URL: https://xgboost.readthedocs.io/en/latest/parameter.html

40. Heart Disease Prediction Evaluation of Machine Learning Models with PSO-Optimized K-Fold Cross-Validation / N. Gupta et al. *2024 IEEE Region 10 Symposium (TENSYMP)*, New Delhi, India, 27–29 September 2024. 2024. P. 1–6. URL: https://doi.org/10.1109/tensymp61132.2024.10752215

41. Sofaer H. R., Hoeting J. A., Jarnevich C. S. The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*>. 2019. Vol. 10, no. 4. P. 565–577. URL: https://doi.org/10.1111/2041-210x.13140

42. Efficient Prequential AUC-PR Computation / D. L. Pereira Gomes et al. *2023 International Conference on Machine Learning and Applications (ICMLA)*, Jacksonville, FL, USA, 15–17 December 2023. 2023. URL: https://doi.org/10.1109/icmla58977.2023.00335

43. Khan S. A., Ali Rana Z. Evaluating Performance of Software Defect Prediction Models Using Area Under Precision-Recall Curve (AUC-PR). *2019 2nd International Conference on Advancements in Computational Sciences (ICACS)*, Lahore, Pakistan, 18–20 February 2019. 2019. URL: https://doi.org/10.23919/icacs.2019.8689135

44. The receiver operating characteristic curve accurately assesses imbalanced datasets / E. Richardson et al. *Patterns*. 2024. P. 100994. URL: https://doi.org/10.1016/j.patter.2024.100994

45. Poisot T. Guidelines for the prediction of species interactions through binary classification. *Methods in Ecology and Evolution*. 2023. URL: https://doi.org/10.1111/2041-210x.14071

46. Python API Reference – XGBoost 3.0.4 documentation. *XGBoost Documentation – XGBoost 3.0.4 documentation*. URL: https://xgboost.readthedocs.io/en/stable/python/python_api.html

47. Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods / H. Wang et al. *Journal of Big Data*. 2024. Vol. 11, no. 1. URL: https://doi.org/10.1186/s40537-024-00905-w

48. Gan L. XGBoost-Based E-Commerce Customer Loss Prediction. *Computational Intelligence and Neuroscience*. 2022. Vol. 2022. P. 1–10. URL: https://doi.org/10.1155/2022/1858300

49. Tool wear prediction based on XGBoost feature selection combined with PSO-BP network / Z. Lin et al. *Scientific Reports*. 2025. Vol. 15, no. 1. URL: https://doi.org/10.1038/s41598-025-85694-9

50. Optimizing cancer classification: a hybrid RDO-XGBoost approach for feature selection and predictive insights / A. Yaqoob et al. *Cancer Immunology, Immunotherapy*. 2024. Vol. 73, no. 12. URL: https://doi.org/10.1007/s00262-024-03843-x

51. XGBoost-SFS and Double Nested Stacking Ensemble Model for Photovoltaic Power Forecasting under Variable Weather Conditions / B. Zhou et al. *Sustainability*. 2023. Vol. 15, no. 17. P. 13146. URL: https://doi.org/10.3390/su151713146

52. Demir S., Sahin E. K. An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost. *Neural Computing and Applications*. 2022. URL: https://doi.org/10.1007/s00521-022-07856-4

53. Atias D., Obolski U. Addressing bias in feature importances derived from XGBoost. Response to Br J Anaesth 2025. *British Journal of Anaesthesia*. 2025. URL: https://doi.org/10.1016/j.bja.2025.01.022

54. Miller A. J. The Convergence of Efroymson's Stepwise Regression Algorithm. *The American Statistician*. 1996. Vol. 50, no. 2. P. 180. URL: https://doi.org/10.2307/2684436

55. Mansfield E. R., Webster J. T., Gunst R. F. An Analytic Variable Selection Technique for Principal Component Regression. *Applied Statistics*. 1977. Vol. 26, no. 1. P. 34. URL: https://doi.org/10.2307/2346865

56. A greedy feature selection algorithm for Big Data of high dimensionality / I. Tsamardinos et al. *Machine Learning*. 2018. Vol. 108, no. 2. P. 149–202. URL: https://doi.org/10.1007/s10994-018-5748-7

57. Gokalp O., Tasci E., Ugur A. A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Systems with Applications*. 2020. Vol. 146. P. 113176. URL: https://doi.org/10.1016/j.eswa.2020.113176

58. Greedy feature selection for ranking / H. Lai et al. *2011 15th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, Laussane, Switzerland, 8–10 June 2011. 2011. URL: https://doi.org/10.1109/cscwd.2011.5960053

59. Gabbi Reddy K., Mishra D. An effective initialization for Fuzzy PSO with Greedy Forward Selection in feature selection. *International Journal of Data Science and Analytics*. 2025. URL: https://doi.org/10.1007/s41060-024-00712-9

60. Soper D. S. Greed Is Good: Rapid Hyperparameter Optimization and Model Selection Using Greedy k-Fold Cross Validation. *Electronics*. 2021. Vol. 10, no. 16. P. 1973. URL: https://doi.org/10.3390/electronics10161973

61. Draper N. R. Applied regression analysis. New York: Wiley, 1966. 407 p.

62. Sequential Backward Feature Selection for Optimizing Permanent Strain Model of Unbound Aggregates / S. O. Aregbesola et al. *Case Studies in Construction Materials*. 2023. P. e02554. URL: https://doi.org/10.1016/j.cscm.2023.e02554

63. Priyatno A. M., Widiyaningtyas T. A Systematic Literature Review: Recursive Feature Elimination Algorithms. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*. 2024. Vol. 9, no. 2. P. 196–207. URL: https://doi.org/10.33480/jitk.v9i2.5015

64. Gene Selection for Cancer Classification using Support Vector Machines / I. Guyon et al. *Machine Learning*. 2002. Vol. 46, no. 1/3. P. 389–422. URL: https://doi.org/10.1023/a:1012487302797

65. RFE. *scikit-learn*. URL: https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

66. Öznacar T., Güler T. Prediction of Early Diagnosis in Ovarian Cancer Patients Using Machine Learning Approaches with Boruta and Advanced Feature Selection. *Life*. 2025. Vol. 15, no. 4. P. 594. URL: https://doi.org/10.3390/life15040594

67. Awad M., Fraihat S. Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. *Journal of Sensor and Actuator Networks*. 2023. Vol. 12, no. 5. P. 67. URL: https://doi.org/10.3390/jsan12050067

68. Kursa M. B., Rudnicki W. R. Feature Selection with the Boruta Package. *Journal of Statistical Software*. 2010. Vol. 36, no. 11. URL: https://doi.org/10.18637/jss.v036.i11

69. Boruta 0.4.3. *PyPI The Python Package Index*. URL: https://pypi.org/project/Boruta/

70. Tang R., Zhang X. CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification. *2020 5th IEEE International Conference on Big Data Analytics (ICBDA)*, Xiamen, China, 8–11 May 2020. 2020. URL: https://doi.org/10.1109/icbda49040.2020.9101199

71. Zhou H., Xin Y., Li S. A diabetes prediction model based on Boruta feature selection and ensemble learning. *BMC Bioinformatics*. 2023. Vol. 24, no. 1. URL: https://doi.org/10.1186/s12859-023-05300-5

72. Computed tomography radiomics for the prediction of thymic epithelial tumor histology, TNM stage and myasthenia gravis / C. Blüthgen et al. *PLOS ONE*. 2021. Vol. 16, no. 12. P. e0261401. URL: https://doi.org/10.1371/journal.pone.0261401

73. Anusha Hegde H., Bhowmik B. Feature Selection for Peer-to-Peer Lending Default Risk Using Boruta and mRMR Approach. *2023 IEEE 20th India Council International Conference (INDICON)*, Hyderabad, India, 14–17 December 2023. 2023. URL: https://doi.org/10.1109/indicon59947.2023.10440917

74. XGBoost-B-GHM: An Ensemble Model with Feature Selection and GHM Loss Function Optimization for Credit Scoring / Y. Xia et al. *Systems*. 2024. Vol. 12, no. 7. P. 254. URL: https://doi.org/10.3390/systems12070254

75. Utilising Explainable Techniques for Quality Prediction in a Complex Textiles Manufacturing Use Case / B. Forsberg et al. *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)*, Bari, Italy, 28 August – 1 September 2024. 2024. P. 245–251. URL: https://doi.org/10.1109/case59546.2024.10711357

76. Мірошниченко С. О., Вовна О. В. Багатофакторний статистичний аналіз успішності замовлень як інструмент оптимізації автоматизованих систем керування ресурсами. *Вісник Приазовського державного технічного університету. Серія: Технічні науки*. 2025. № 50. С. 182–190. URL: https://doi.org/10.31498/2225-6733.50.2025.336372

77. Multi-objective hyperparameter tuning and feature selection using filter ensembles / M. Binder et al. *GECCO '20: Genetic and Evolutionary Computation Conference*, Cancún Mexico. New York, NY, USA, 2020. URL: https://doi.org/10.1145/3377930.3389815

78. Lorraine J., Duvenaud D. Stochastic Hyperparameter Optimization through Hypernetworks. *Computer Science. Machine Learning*. 2018. P. 1–9. URL: https://doi.org/10.48550/arXiv.1802.09419

79. Efficient and Joint Hyperparameter and Architecture Search for Collaborative Filtering / Y. Wen et al. *KDD '23: The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Long Beach CA USA. New York, NY, USA, 2023. URL: https://doi.org/10.1145/3580305.3599322

80. Gradient boosted feature selection / Z. Xu et al. *KDD '14: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York New York USA. New York, NY, USA, 2014. URL: https://doi.org/10.1145/2623330.2623635

81. Shakshi, Kulhare R. Gradient Boosting Feature Selection with Machine Learning Classifier for Prediction of Intrusion Image. *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Tashkent, Uzbekistan, 13–15 November 2024. 2024. P. 1214–1219. URL: https://doi.org/10.1109/ictacs62700.2024.10840506

82. Optimizing Software Defect Prediction Models: Integrating Hybrid Grey Wolf and Particle Swarm Optimization for Enhanced Feature Selection with Popular Gradient Boosting Algorithm / Angga Maulana Akbar et al. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*. 2024. Vol. 6, no. 2. P. 169–181. URL: https://doi.org/10.35882/jeeemi.v6i2.388

83. Performance Analysis of Feature Selection Methods in Software Defect Prediction: A Search Method Approach / A. O. Balogun et al. *Applied Sciences*. 2019. Vol. 9, no. 13. P. 2764. URL: https://doi.org/10.3390/app9132764

84. Impact of Feature Selection Methods on the Predictive Performance of Software Defect Prediction Models: An Extensive Empirical Study / A. O. Balogun et al. *Symmetry*. 2020. Vol. 12, no. 7. P. 1147. URL: https://doi.org/10.3390/sym12071147

85. Manipulating and Measuring Model Interpretability / F. Poursabzi-Sangdeh et al. *CHI '21: CHI Conference on Human Factors in Computing Systems*, Yokohama Japan. New York, NY, USA, 2021. URL: https://doi.org/10.1145/3411764.3445315

86. Multiobjective Evolutionary Feature Selection for Fuzzy Classification / F. Jimenez et al. *IEEE Transactions on Fuzzy Systems*. 2019. Vol. 27, no. 5. P. 1085–1099. URL: https://doi.org/10.1109/tfuzz.2019.2892363

87. Exploring local interpretability in dimensionality reduction: Analysis and use cases / N. Mylonas et al. *Expert Systems with Applications*. 2024. P. 124074. URL: https://doi.org/10.1016/j.eswa.2024.124074

REFERENCES:

1. De, T. S., Singh, P., & Patel, A. (2024). A Machine learning and Empirical Bayesian Approach for Predictive Buying in B2B E-commerce. In *ICMLSC 2024: 2024 The 8th International Conference on Machine Learning and Soft Computing*. ACM. https://doi.org/10.1145/3647750.3647754

2. Li, K. (2023). A sales prediction method based on XGBoost algorithm model. *BCP Business & Management*, *36*, 367–371. https://doi.org/10.54691/bcpbm.v36i.3487

3. Miroshnychenko, S. O. (2024). Improvement of the existing functionality for forecasting finished goods inventory at the enterprise warehouses based on the stock forecasted report ERP ODOO. In *MININGMETALTECH 2024 – The mining and metals sector: Integration of business, technology and education* (Vol. 2, pp. 48–51). https://doi.org/10.30525/978-9934-26-506-8-133

4. Heino, A. (2021). *New product demand forecasting in retail: Applying machine learning techniques to forecast demand for new product purchasing decisions* [Master's thesis]. https://www.semanticscholar.org/paper/Applying-Machine-Learning-Techniques-to-Forecast-Heino/d17599a4ebc651fd29997c4f2f5f10e0b905e9a4

5. Chen, T., & Guestrin, C. (2016). XGBoost. In *KDD '16: The 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.1145/2939672.2939785

6. Abdullah-All-Tanvir, et al. (2023). A gradient boosting classifier for purchase intention prediction of online shoppers. *Heliyon*, *9*(4), e15163. https://doi.org/10.1016/j.heliyon.2023.e15163

7. Leevy, J. L., Hancock, J., & Khoshgoftaar, T. M. (2024). Medicare fraud detection: A comparative study on the effectiveness of one-class and binary classification models. *International Journal of Computers and Applications*, 1–6. https://doi.org/10.1080/1206212x.2024.2431874

8. Alsahaf, A., et al. (2021). A framework for feature selection through boosting. *Expert Systems with Applications*, 115895. https://doi.org/10.1016/j.eswa.2021.115895

9. Pansari, N., et al. (2024). Attack classification using machine learning on UNSW-NB 15 dataset using XGBoost feature selection & ablation analysis. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)*. IEEE. https://doi.org/10.1109/i2ct61223.2024.10543523

10. Dou, Y., & Meng, W. (2024). Predicting cancer classification based on the improved XGBoost algorithms. In *2024 6th International Conference on Frontier Technologies of Information and Computer (ICFTIC)* (pp. 1013–1016). IEEE. https://doi.org/10.1109/icftic64248.2024.10913271

11. Pudjihartono, N., et al. (2022). A review of feature selection methods for machine learning-based disease risk prediction. *Frontiers in Bioinformatics*, *2*. https://doi.org/10.3389/fbinf.2022.927312

12. Cheng, X. (2024). A comprehensive study of feature selection techniques in machine learning models. *Insights in Computer, Signals and Systems*, *1* (1), 65–78. https://doi.org/10.70088/xpf2b276

13. Noroozi, Z., Orooji, A., & Erfannia, L. (2023). Analyzing the impact of feature selection methods on machine learning algorithms for heart disease prediction. *Scientific Reports*, *13* (1). https://doi.org/10.1038/s41598-023-49962-w

14. Welch, W. J. (1982). Algorithmic complexity: threeNP- hard problems in computational statistics. *Journal of Statistical Computation and Simulation*, *15* (1), 17–25. https://doi.org/10.1080/00949658208810560

15. Fong, S., Wong, R., & Vasilakos, A. (2015). Accelerated PSO Swarm Search Feature Selection for Data Stream Mining Big Data. *IEEE Transactions on Services Computing*, 1. https://doi.org/10.1109/tsc.2015.2439695

16. Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, *97* (1–2), 245–271. https://doi.org/10.1016/s0004-3702(97)00063-5

17. Dash, M., & Liu, H. (1997). Feature Selection for Classification. *Intelligent Data Analysis*, *1* (3), 131–156. https://doi.org/10.3233/ida-1997-1302

18. Bertsimas, D., King, A., & Mazumder, R. (2016). Best subset selection via a modern optimization lens. *The Annals of Statistics*, *44* (2), 813–852. https://doi.org/10.1214/15-aos1388

19. Chen, R.-C., Dewi, C., Huang, S.-W., & Caraka, R. E. (2020). Selecting critical features for data classification based on machine learning methods. *Journal of Big Data*, *7* (1). https://doi.org/10.1186/s40537-020-00327-4

20. Degenhardt, F., Seifert, S., & Szymczak, S. (2017). Evaluation of variable selection methods for random forests and omics data sets. *Briefings in Bioinformatics*, *20* (2), 492–503. https://doi.org/10.1093/bib/bbx124

21. Jagdhuber, R., Lang, M., Stenzl, A., Neuhaus, J., & Rahnenführer, J. (2020). Cost-Constrained feature selection in binary classification: adaptations for greedy forward selection and genetic algorithms. *BMC Bioinformatics*, *21* (1). https://doi.org/10.1186/s12859-020-3361-9

22. Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in bioinformatics. *Bioinformatics*, *23* (19), 2507–2517. https://doi.org/10.1093/bioinformatics/btm344

23. Koca, Y. B., & Aktepe, E. (2024). Evaluation of Missing Data Imputation Methods and PCA Techniques for Machine Learning Models in Breast Cancer Diagnosis Using WBCD. *Türk Doğa ve Fen Dergisi*. https://doi.org/10.46810/tdfd.1460871

24. Arimie, C. O., Biu, E. O., & Ijomah, M. A. (2020). Outlier Detection and Effects on Modeling. *OALib*, *07* (09), 1–30. https://doi.org/10.4236/oalib.1106619

25. Shved, A. V., & Davydenko, Y. O. (2022). Outlier Detection Technique for Heterogeneous Data Using Trimmed-Mean Robust Estimators. *Radio Electronics, Computer Science, Control*, (3), 50. https://doi.org/10.15588/1607-3274-2022-3-5

26. Jones, P. R. (2019). A note on detecting statistical outliers in psychophysical data. *Attention, Perception, & Psychophysics*, *81* (5), 1189–1196. https://doi.org/10.3758/s13414-019-01726-3

27. Haanappel, C. P., & Voor in 't holt, A.F. (2024). Using the interquartile range in infection prevention and control research. *Infection Prevention in Practice*, 100337. https://doi.org/10.1016/j.infpip.2024.100337

28. *MSCI Global Investable Market Value and Growth Indexes Methodology*. (n.d.). Bringing clarity to investment decisions MSCI. https://www.msci.com/eqb/methodology/meth_docs/MSCI_GIMIVGMethod_Feb2021.pdf

29. Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies*, *9* (3), 52. https://doi.org/10.3390/technologies9030052

30. *CyclicalFeatures – 1.8.3*. (n.d.). Feature-engine – 1.8.3. https://feature-engine.trainindata.com/en/1.8.x/user_guide/creation/CyclicalFeatures.html

31. Ul Haq, I., Gondal, I., Vamplew, P., & Brown, S. (2019). Categorical Features Transformation with Compact One-Hot Encoder for Fraud Detection in Distributed Environment. In *Communications in Computer and Information Science* (pp. 69–80). Springer Singapore. https://doi.org/10.1007/978-981-13-6661-1_6

32. *Notes on Parameter Tuning – xgboost 3.0.4 documentation*. (n.d.). XGBoost Documentation – xgboost 3.0.4 documentation. https://xgboost.readthedocs.io/en/stable/tutorials/param_tuning.html#handle-imbalanced-dataset

33. Särndal, C.-E., Swensson, B., & Wretman, J. (2003). *Model Assisted Survey Sampling (Springer Series in Statistics)*. Springer.

34. Shenouda, J., & Bajwa, W. U. (2023). A Guide to Computational Reproducibility in Signal Processing and Machine Learning [Tips & Tricks]. *IEEE Signal Processing Magazine*, *40* (2), 141–151. https://doi.org/10.1109/msp.2022.3217659

35. Florek, P., & Zagda´nski, A. (2023). Benchmarking state-of-the-art gradient boosting algorithms for classification. *Computer Science. Machine Learning*, Article arXiv:2305.17094 [cs.LG]. https://doi.org/10.48550/arXiv.2305.17094

36. G, K., C. A, D. S., & Kumar R., N. (2024). Enhanced DNS Cybersecurity: Optimization of XGBoost Using Automated Hyperparameter Tuning via Optuna. In *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)* (pp. 1–7). IEEE. https://doi.org/10.1109/icmnwc63764.2024.10872404

37. Pinichka, C., Chotpantarat, S., Cho, K. H., & Siriwong, W. (2025). Comparative analysis of SWAT and SWAT coupled with XGBoost model using Optuna hyperparameter optimization for nutrient simulation: A case study in the Upper Nan River basin, Thailand. *Journal of Environmental Management*, *388*, 126053. https://doi.org/10.1016/j.jenvman.2025.126053

38. Miroshnychenko, S. O. (2025). Optymizatsiia hiperparametriv XGBOOST dlia intelektualnykh system prohnozuvannia B2B zamovlen. Naukovyi Zhurnal Metinvest Politekhniky. Seriia: Tekhnichni nauky, 3 (4).

39. *XGBoost Parameters – xgboost 3.1.0-dev documentation*. (n.d.). XGBoost Documentation – xgboost 3.0.4 documentation. https://xgboost.readthedocs.io/en/latest/parameter.html

40. Gupta, N., Jadon, K. S., Soni, P., Gupta, N., & Dhurandher, S. K. (2024). Heart Disease Prediction Evaluation of Machine Learning Models with PSO-Optimized K-Fold Cross-Validation. In *2024 IEEE Region 10 Symposium (TENSYMP)* (pp. 1–6). IEEE. https://doi.org/10.1109/tensymp61132.2024.10752215

41. Sofaer, H. R., Hoeting, J. A., & Jarnevich, C. S. (2019). The area under the precision-recall curve as a performance metric for rare binary events. *Methods in Ecology and Evolution*, *10*(4), 565–577. https://doi.org/10.1111/2041-210x.13140

42. Pereira Gomes, D. L., Grégio, A., Zanata Alves, M. A., & Lisboa de Almeida, P. R. (2023). Efficient Prequential AUC-PR Computation. In *2023 International Conference on Machine Learning and Applications (ICMLA)*. IEEE. https://doi.org/10.1109/icmla58977.2023.00335

43. Khan, S. A., & Ali Rana, Z. (2019). Evaluating Performance of Software Defect Prediction Models Using Area Under Precision-Recall Curve (AUC-PR). In *2019 2nd International Conference on Advancements in Computational Sciences (ICACS)*. IEEE. https://doi.org/10.23919/icacs.2019.8689135

44. Richardson, E., Trevizani, R., Greenbaum, J. A., Carter, H., Nielsen, M., & Peters, B. (2024). The receiver operating characteristic curve accurately assesses imbalanced datasets. *Patterns*, 100994. https://doi.org/10.1016/j.patter.2024.100994

45. Poisot, T. (2023). Guidelines for the prediction of species interactions through binary classification. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210x.14071

46. *Python API Reference – xgboost 3.0.4 documentation*. (n.d.). XGBoost Documentation – xgboost 3.0.4 documentation. https://xgboost.readthedocs.io/en/stable/python/python_api.html

47. Wang, H., Liang, Q., Hancock, J. T., & Khoshgoftaar, T. M. (2024). Feature selection strategies: a comparative analysis of SHAP-value and importance-based methods. *Journal of Big Data*, *11* (1). https://doi.org/10.1186/s40537-024-00905-w

48. Gan, L. (2022). XGBoost-Based E-Commerce Customer Loss Prediction. *Computational Intelligence and Neuroscience*, *2022*, 1–10. https://doi.org/10.1155/2022/1858300

49. Lin, Z., Fan, Y., Tan, J., Li, Z., Yang, P., Wang, H., & Duan, W. (2025). Tool wear prediction based on XGBoost feature selection combined with PSO-BP network. *Scientific Reports*, *15* (1). https://doi.org/10.1038/s41598-025-85694-9

50. Yaqoob, A., Verma, N. K., Aziz, R. M., & Shah, M. A. (2024). Optimizing cancer classification: a hybrid RDO-XGBoost approach for feature selection and predictive insights. *Cancer Immunology, Immunotherapy*, *73*(12). https://doi.org/10.1007/s00262-024-03843-x

51. Zhou, B., Chen, X., Li, G., Gu, P., Huang, J., & Yang, B. (2023). XGBoost–SFS and Double Nested Stacking Ensemble Model for Photovoltaic Power Forecasting under Variable Weather Conditions. *Sustainability*, *15*(17), 13146. https://doi.org/10.3390/su151713146

52. Demir, S., & Sahin, E. K. (2022). An investigation of feature selection methods for soil liquefaction prediction based on tree-based ensemble algorithms using AdaBoost, gradient boosting, and XGBoost. *Neural Computing and Applications*. https://doi.org/10.1007/s00521-022-07856-4

53. Atias, D., & Obolski, U. (2025). Addressing bias in feature importances derived from XGBoost. Response to Br J Anaesth 2025. *British Journal of Anaesthesia*. https://doi.org/10.1016/j.bja.2025.01.022

54. Miller, A. J. (1996). The Convergence of Efroymson's Stepwise Regression Algorithm. *The American Statistician*, *50*(2), 180. https://doi.org/10.2307/2684436

55. Mansfield, E. R., Webster, J. T., & Gunst, R. F. (1977). An Analytic Variable Selection Technique for Principal Component Regression. *Applied Statistics*, *26*(1), 34. https://doi.org/10.2307/2346865

56. Tsamardinos, I., Borboudakis, G., Katsogridakis, P., Pratikakis, P., & Christophides, V. (2018). A greedy feature selection algorithm for Big Data of high dimensionality. *Machine Learning*, *108*(2), 149–202. https://doi.org/10.1007/s10994-018-5748-7

57. Gokalp, O., Tasci, E., & Ugur, A. (2020). A novel wrapper feature selection algorithm based on iterated greedy metaheuristic for sentiment classification. *Expert Systems with Applications*, *146*, 113176. https://doi.org/10.1016/j.eswa.2020.113176

58. Lai, H., Tang, Y., Luo, H., & Pan, Y. (2011). Greedy feature selection for ranking. In *2011 15th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE. https://doi.org/10.1109/cscwd.2011.5960053

59. Gabbi Reddy, K., & Mishra, D. (2025). An effective initialization for Fuzzy PSO with Greedy Forward Selection in feature selection. *International Journal of Data Science and Analytics*. https://doi.org/10.1007/s41060-024-00712-9

60. Soper, D. S. (2021). Greed Is Good: Rapid Hyperparameter Optimization and Model Selection Using Greedy k-Fold Cross Validation. *Electronics*, *10*(16), 1973. https://doi.org/10.3390/electronics10161973

61. Draper, N. R. (1966). *Applied regression analysis*. Wiley.

62. Aregbesola, S. O., Won, J., Kim, S., & Byun, Y.-H. (2023). Sequential Backward Feature Selection for Optimizing Permanent Strain Model of Unbound Aggregates. *Case Studies in Construction Materials*, Article e02554. https://doi.org/10.1016/j.cscm.2023.e02554

63. Priyatno, A. M., & Widiyaningtyas, T. (2024). A Systematic Literature Review: Recursive Feature Elimination Algorithms. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, *9*(2), 196–207. https://doi.org/10.33480/jitk.v9i2.5015

64. Guyon, I., Weston, J., Barnhill, S., & Vapnik, V. (2002). Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning*, *46* (1/3), 389–422. https://doi.org/10.1023/a:1012487302797

65. *RFE*. (n.d.). scikit-learn. https://scikit-learn.org/stable/modules/generated/sklearn.feature_selection.RFE.html

66. Öznacar, T., & Güler, T. (2025). Prediction of Early Diagnosis in Ovarian Cancer Patients Using Machine Learning Approaches with Boruta and Advanced Feature Selection. *Life*, *15* (4), 594. https://doi.org/10.3390/life15040594

67. Awad, M., & Fraihat, S. (2023). Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems. *Journal of Sensor and Actuator Networks*, *12* (5), 67. https://doi.org/10.3390/jsan12050067

68. Kursa, M. B., & Rudnicki, W.R. (2010). Feature Selection with theBorutaPackage. *Journal of Statistical Software*, *36* (11). https://doi.org/10.18637/jss.v036.i11

69. *Boruta 0.4.3*. (n.d.). PyPI The Python Package Index. https://pypi.org/project/Boruta

70. Tang, R., & Zhang, X. (2020). CART Decision Tree Combined with Boruta Feature Selection for Medical Data Classification. In *2020 5th IEEE International Conference on Big Data Analytics (ICBDA)*. IEEE. https://doi.org/10.1109/icbda49040.2020.9101199

71. Zhou, H., Xin, Y., & Li, S. (2023). A diabetes prediction model based on Boruta feature selection and ensemble learning. *BMC Bioinformatics*, *24* (1). https://doi.org/10.1186/s12859-023-05300-5

72. Blüthgen, C., Patella, M., Euler, A., Baessler, B., Martini, K., von Spiczak, J., Schneiter, D., Opitz, I., & Frauenfelder, T. (2021). Computed tomography radiomics for the prediction of thymic epithelial tumor histology, TNM stage and myasthenia gravis. *PLOS ONE*, *16* (12), Article e0261401. https://doi.org/10.1371/journal.pone.0261401

73. Anusha Hegde, H., & Bhowmik, B. (2023). Feature Selection for Peer-to-Peer Lending Default Risk Using Boruta and mRMR Approach. In *2023 IEEE 20th India Council International Conference (INDICON)*. IEEE. https://doi.org/10.1109/indicon59947.2023.10440917

74. Xia, Y., Jiang, S., Meng, L., & Ju, X. (2024). XGBoost-B-GHM: An Ensemble Model with Feature Selection and GHM Loss Function Optimization for Credit Scoring. *Systems*, *12* (7), 254. https://doi.org/10.3390/systems12070254

75. Forsberg, B., Williams, H., MacDonald, B., Chen, T., Hamzeh, R., & Hulse, K. (2024). Utilising Explainable Techniques for Quality Prediction in a Complex Textiles Manufacturing Use Case. In *2024 IEEE 20th International Conference on Automation Science and Engineering (CASE)* (pp. 245–251). IEEE. https://doi.org/10.1109/case59546.2024.10711357

76. Miroshnychenko, S. O. (2025). Bahatofaktornyi statystychnyi analiz uspishnosti zamovlen yak instrument optymizatsii avtomatyzovanykh system keruvannia resursamy. Visnyk Pryazovskoho derzhavnoho tekhnichnoho universytetu, Seriia: Tekhnichni nauky (50), 182–190. https://doi.org/10.31498/2225-6733.50.2025.336372

77. Binder, M., Moosbauer, J., Thomas, J., & Bischl, B. (2020). Multi-objective hyperparameter tuning and feature selection using filter ensembles. In *GECCO '20: Genetic and Evolutionary Computation Conference*. ACM. https://doi.org/10.1145/3377930.3389815

78. Lorraine, J., & Duvenaud, D. (2018). Stochastic Hyperparameter Optimization through Hypernetworks. *Computer Science. Machine Learning*, Article arXiv:1802.09419 [cs.LG]. https://doi.org/10.48550/arXiv.1802.09419

79. Wen, Y., Gao, C., Yi, L., Qiu, L., Wang, Y., & Li, Y. (2023). Efficient and Joint Hyperparameter and Architecture Search for Collaborative Filtering. In *KDD '23: The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.1145/3580305.3599322

80. Xu, Z., Huang, G., Weinberger, K.Q., & Zheng, A.X. (2014). Gradient boosted feature selection. IN *KDD '14: The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM. https://doi.org/10.1145/2623330.2623635

81. Shakshi & Kulhare, R. (2024). Gradient Boosting Feature Selection with Machine Learning Classifier for Prediction of Intrusion Image. In *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)* (pp. 1214–1219). IEEE. https://doi.org/10.1109/ictacs62700.2024.10840506

82. Angga Maulana Akbar, Herteno, R., Saputro, S.W., Faisal, M.R., & Nugroho, R.A. (2024). Optimizing Software Defect Prediction Models: Integrating Hybrid Grey Wolf and Particle Swarm Optimization for Enhanced Feature Selection with Popular Gradient Boosting Algorithm. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, *6* (2), 169–181. https://doi.org/10.35882/jeeemi.v6i2.388

83. Balogun, A. O., Basri, S., Abdulkadir, S. J., & Hashim, A. S. (2019). Performance Analysis of Feature Selection Methods in Software Defect Prediction: A Search Method Approach. *Applied Sciences*, *9* (13), 2764. https://doi.org/10.3390/app9132764

84. Balogun, A. O., Basri, S., Mahamad, S., Abdulkadir, S. J., Almomani, M. A., Adeyemo, V. E., Al-Tashi, Q., Mojeed, H. A., Imam, A. A., & Bajeh, A. O. (2020). Impact of Feature Selection Methods on the Predictive Performance of Software Defect Prediction Models: An Extensive Empirical Study. *Symmetry*, *12* (7), 1147. https://doi.org/10.3390/sym12071147

85. Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., & Wallach, H. (2021). Manipulating and Measuring Model Interpretability. In *CHI '21: CHI Conference on Human Factors in Computing Systems*. ACM. https://doi.org/10.1145/3411764.3445315

86. Jimenez, F., Martinez, C., Marzano, E., Palma, J. T., Sanchez, G., & Sciavicco, G. (2019). Multiobjective Evolutionary Feature Selection for Fuzzy Classification. *IEEE Transactions on Fuzzy Systems*, *27* (5), 1085–1099. https://doi.org/10.1109/tfuzz.2019.2892363

87. Mylonas, N., Mollas, I., Bassiliades, N., & Tsoumakas, G. (2024). Exploring local interpretability in dimensionality reduction: Analysis and use cases. *Expert Systems with Applications*, 124074. https://doi.org/10.1016/j.eswa.2024.124074

Стаття надійшла: 28.08.2025
Стаття прийнята: 21.09.2025
Опублікована: 10.11.2025