

## АВТОМАТИЗАЦІЯ, КОМП'ЮТЕРНО-ІНТЕГРОВАНІ ТЕХНОЛОГІЇ ТА РОБОТОТЕХНІКА

УДК 004.852:004.67

DOI <https://doi.org/10.32782/3041-2080/2026-7-4>

### ВПЛИВ СТРАТЕГІЇ ФОРМУВАННЯ НАВЧАЛЬНОГО ВІКНА НА ЯКІСТЬ ПЕРЕНАВЧАННЯ МОДЕЛІ XGBOOST В УМОВАХ КОНЦЕПТУАЛЬНОГО ДРЕЙФУ В2В-ТРАНЗАКЦІЙ

**Мірошниченко Сергій Олександрович,**

аспірант, старший викладач кафедри автоматизації, електро- та робототехнічних систем

ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА»

ORCID ID: 0009-0007-4868-3006

Стаття містить порівняльне дослідження одинадцяти стратегій формування навчального вікна для алгоритму XGBoost у задачі бінарної класифікації успішності В2В-замовлень в умовах концептуального дрейфу. Дослідження виконано на деперсоніфікованому масиві 86 786 транзакцій промислової ERP-системи за період липень 2017 – листопад 2024 р. Оцінювання проводилося за преквенційним протоколом із щомісячним перенавчанням на 76-місячному горизонті. Порівнювалися чотири категорії стратегій: кумулятивна, фракційні (1/2, 1/3, 1/4), часові (3, 4, 6, 12 місяців) і лічильникові (3 000, 5 000, 10 000 записів). В результаті стратегія «часове вікно 6 місяців» показала найвище значення показнику ROC-AUC = 0,9440, найнижче LogLoss = 0,2732 та мінімальний коефіцієнт варіації (CV = 0,0269). Найгірший результат серед часових конфігурацій показала стратегія «Часове вікно 12 місяців» (ROC-AUC = 0,9302). Кумулятивна стратегія, попри доступ до понад 80 тис. записів на пізніх ітераціях, систематично поступалася більш оптимальним за розміром вікнам внаслідок ефекту «катастрофічного запам'ятовування» (Catastrophic Retention) – надмірного збереження нерелевантних застарілих патернів, симетричного до «катастрофічного забування» в нейронних мережах. Статистичну значущість відмінностей підтверджено критерієм Фрідмана ( $\chi^2 = 99,89$ ;  $p < 10^{-16}$ ;  $W = 0,131$ ) та тестами Вілкоксона з поправкою Бонферроні (17 із 55 пар значущі при  $\alpha = 0,05$ ). Застосування стратегії, що відповідає часовому вікну 6 місяців (Time 6M) замість кумулятивної скорочує обчислювальні витрати вдвічі, реалізуючи принципи Green AI в промислових MLOps-конвеєрах.

**Ключові слова:** машинне навчання, XGBoost, концептуальний дрейф, навчальне вікно, перенавчання моделей, В2В-прогнозування, преквенційна оцінка, ERP-системи, катастрофічне запам'ятовування.

#### **Miroshnychenko Serhii. Impact of training window formation strategy on XGBoost model retraining quality under concept drift in B2B transactions**

This paper presents a systematic comparative study of eleven training window strategies for the XGBoost algorithm in a binary classification task of B2B order success prediction under concept drift. The study is conducted on a depersonalized dataset of 86,786 transactions from a real industrial ERP system spanning from July 2017 to November 2024. Evaluation follows a prequential protocol with monthly model retraining over a 76-month streaming horizon. Four strategy categories are compared: cumulative (full history), fractional (1/2, 1/3, 1/4 of history), temporal (12, 6, 4, 3 months), and count-based (10,000; 5,000; 3,000 samples). The “6-month time window” strategy achieved the highest global ROC-AUC of 0.9440 and the lowest LogLoss of 0.2732, while simultaneously exhibiting the lowest coefficient of variation (CV – 0.0269). The “12-month time window” strategy showed the worst result among the time configurations (ROC-AUC = 0.9302). The cumulative strategy, despite access to over 80,000 records in later iterations, systematically lost to more optimally sized windows due to the author-proposed “Catastrophic Retention” effect — the excessive preservation of irrelevant, outdated patterns, symmetric to catastrophic forgetting in neural networks. Statistical significance of differences is confirmed by the Friedman test ( $\chi^2 = 99,89$ ;  $p < 10^{-16}$ ;  $W = 0,131$ ) and pairwise Wilcoxon signed-rank tests with Bonferroni correction (17 of 55 pairs significant at  $\alpha=0.05$ ). Replacing the cumulative strategy with a “6-month time window” strategy reduces retraining computational costs by half, implementing Green AI principles in industrial MLOps pipelines.

**Key words:** machine learning, XGBoost, concept drift, training window, model retraining, B2B forecasting, prequential evaluation, ERP systems, order classification, non-stationary environment, catastrophic retention.

**Вступ.** Розвиток корпоративних інформаційних систем характеризується переходом від ретроспективного обліку до проактивного управління операційними процесами. Традиційні системи планування ресурсів підприємства (ERP), що побудовані на архітектурі реляційних баз даних, забезпечували переважно статичну звітність на основі кількісних метрик минулих періодів [1; 2], що суттєво обмежувало їхню адаптивність до динамічного бізнес-середовища [3]. Інтеграція алгоритмів машинного навчання в ERP-системи сформувала концепцію інтелектуальних ERP (i-ERP): накопичені масиви транзакційних даних трансформуються в інструменти предиктивної аналітики та підтримки рішень у режимі реального часу [4; 5]. ML-моделі виявляють приховані патерни, прогнозують попит, оптимізують ланцюги постачання і передбачають фінансові ризики [5; 6].

Стратегічне значення у цьому контексті набуває задача предиктивного скорингу B2B-транзакцій, тобто оцінки ймовірності успішного закриття комерційних угод. Точність таких прогнозів безпосередньо впливає на оптимізацію ресурсів відділів продажу, планування виробничих потужностей і управління фінансовими потоками [4]. Практичне розгортання прогностичних моделей в ERP-середовищі неминуче стикається з проблемою деградації якості прогнозування внаслідок нестаціонарності середовища.

Деградація ML-моделей зумовлена концептуальним дрейфом – зміною прихованих статистичних залежностей між предикторами та цільовою змінною з плином часу [7; 8]. Формально, якщо розподіл даних у момент  $t_0$  відрізняється від розподілу в момент  $t_1$ , то дрейф зафіксовано. Декомпозиція Баєса  $P(X, y) = P(y|X) \cdot P(X)$  дозволяє виокремити три джерела нестаціонарності: коваріатний зсув (зміна розподілу  $P(X)$  при незмінній функції  $P(y|X)$ , що проявляється, наприклад, як зростання середніх сум замовлень унаслідок інфляції); зсув апіорної ймовірності (зміна розподілу класів  $P(y)$ , характерна для кризових періодів зі стрибкоподібним зростанням частки скасованих замовлень); справжній концептуальний дрейф (еволюція апостеріорного розподілу  $P(y|X)$ , тобто зміна правил прийняття рішень унаслідок трансформації ринкової кон'юнктури). Остання форма є найскладнішою для виявлення та адаптації [7; 9; 10].

За характером часових змін концептуальний дрейф поділяється на чотири типи: раптовий (геополітичні шоки), поступовий (паралельне співіснування старого та нового концептів), інкрементний (повільна безперервна

трансформація) та рекурентний (повернення до попередніх концептів, характерне для сезонних і циклічних збурень). Макроекономічні збурення, сезонність і безперервна зміна поведінки контрагентів формують у B2B-секторі нестаціонарне середовище, за умов якого статично навчена модель неминуче втрачає прогностичну валідність [7; 11; 12].

Дослідження у сфері потокового навчання традиційно зосереджуються на статистичних детекторах дрейфу: DDM, EDDM та ADWIN [13–15]. Проте застосування цих алгоритмів до табличних ERP-даних виявляє суттєві обмеження: висока зашумленість і дисбаланс класів B2B-транзакцій провокують надмірну кількість хибних спрацювань [16]. Зазначені детектори розроблялися для інкрементних алгоритмів (дерева Хефдінга, наївний Баєс), тоді як індустріальним стандартом для табличних даних є ансамблеві пакетні моделі, зокрема XGBoost, для яких безперервне перенавчання обчислювально нерентабельне [11; 17]. Стандартом MLOps-архітектури залишається регулярне пакетне перенавчання на оновленому масиві даних [18; 19].

Ключовою архітектурною проблемою пакетного перенавчання є визначення оптимального обсягу навчальної вибірки. Виокремлюють чотири категорії стратегій: кумулятивні (повна накопичена історія), фракційні (фіксована частка від поточного обсягу), часові ковзні (жорсткий календарний горизонт) і лічильникові (фіксована кількість найновіших записів). Кумулятивна стратегія зберігає максимальну статистичну різноманітність, проте включає застарілі патерни, що конфліктують із поточним розподілом. Фракційні стратегії автоматично масштабуються з ростом даних, але не враховують часову структуру. Часові вікна забезпечують актуальність ціною можливої втрати обсягу під час сезонних спадів. Лічильникові вікна фіксують кількість записів, ігноруючи часову динаміку [20–22].

В емпіричній літературі порівняльний аналіз стратегій вікна залишається недостатньо формалізованим. В роботі [23] продемонстровано перевагу помірних ковзних вікон над великими статичними блоками у задачах пакетної класифікації в умовах дрейфу, однак без кількісної верифікації на реальних B2B-даних. Автори [24] досліджували адаптацію кредитного скорингу на основі градієнтного бустингу без систематичного порівняння кількох стратегій навчального вікна. В дослідженні [25] автори систематизують підходи до перенавчання в контексті загальної теорії дрейфу без емпіричного

порівняння більш ніж трьох конфігурацій. У суміжних областях – прогнозуванні дефектів програмного забезпечення [26] та фінансовому скорингу – підтверджено перевагу ковзних вікон над кумулятивними підходами та критичну залежність XGBoost від горизонту навчальних даних [27; 28]. Оскільки більшість попередніх досліджень обмежується двома-трьома конфігураціями, комплексне оцінювання широкого спектру стратегій на реальних нестационарних B2B-транзакціях залишалося незавершеним.

**Мета дослідження** — систематичне порівняльне оцінювання одинадцяти стратегій формування навчального вікна для XGBoost у задачі прогнозування успішності B2B-замовлень із застосуванням преквенційного протоколу, статистичних критеріїв значущості та аналізу компромісу між прогностичною якістю, стабільністю й обчислювальною вартістю.

Для досягнення мети визначено наступні завдання:

1. Реалізувати 76-місячну преквенційну симуляцію з щомісячним перенавчанням для чотирьох категорій стратегій;
2. Оцінити глобальні та щомісячні метрики якості класифікації;
3. Верифікувати статистичну значущість відмінностей між стратегіями;
4. Проаналізувати компроміс між якістю, стабільністю та обчислювальними витратами.

#### Методи та методики дослідження.

Емпіричну базу дослідження становлять 86 786 деперсонфікованих B2B-транзакцій реальної промислової ERP-системи за період із 29 липня 2017 р. по 26 листопада 2024 р. Транзакції, датовані після 1 грудня 2024 р., вилучено як неповні. Цільова змінна `is_successful` є бінарним індикатором: клас «1» відповідає угоді,

успішно закритій із дотриманням зобов'язань і плановою маржинальністю; клас «0» – скасування, критичному відстроченню або збитковому закриттю.

Динаміка щомісячного обсягу транзакцій та середнього відсотка успішних угод ілюструє структурні збурення: виражену волатильність обсягів замовлень у 2020–2021 роках і різке синхронне падіння кількості транзакцій та частки їх успішності на початку 2022 року (рис. 1). Така нелінійна поведінка часових рядів із формуванням нових патернів стабілізації підтверджує наявність інтенсивного концептуального дрейфу та методично обґрунтовує відмову від статичних моделей на користь регулярного преквенційного перенавчання.

Річна структура транзакцій демонструє ефект зсуву апіорної ймовірності: у 2020–2021 роках частка успішних угод знизилася до 53,1% та 50,7% відповідно, тоді як у 2017–2018 та 2023–2024 роках вона перевищувала 70%. Ця виражена нестабільність балансу класів пояснює вразливість кумулятивної стратегії навчального вікна, яка інтегрує суперечливі історичні пропорції в єдиний простір навчання.

Простір ознак налічує 35 предикторів п'яти аналітичних категорій: фінансові метрики (сума замовлення, середня ціна позиції, середній чек контрагента для загальних, успішних і неуспішних угод), кількісні метрики (кількість рядків, унікальних продуктів, загальний обсяг позицій), комунікаційні метрики (інтенсивність CRM-взаємодії, кількість цінових ітерацій та змін у замовленні на рівні угоди й контрагента), часові предиктори (година, день тижня, місяць, рік, ознаки вихідного дня, початку/кінця місяця, інтервал від попереднього замовлення) і калькулятивні метрики (відношення суми замовлення

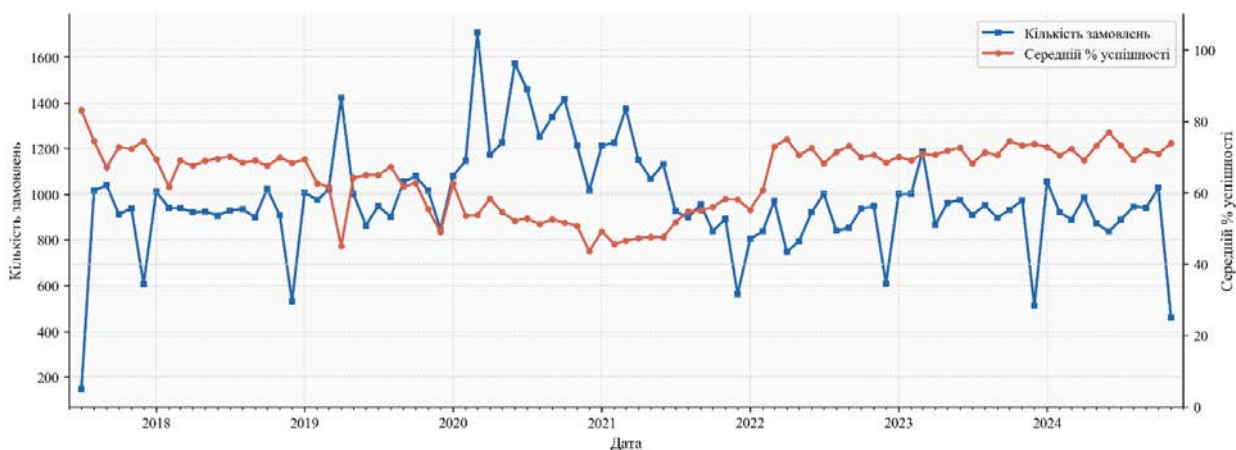


Рис. 1. Динаміка щомісячного обсягу B2B-замовлень та частки успішних угод (липень 2017 – листопад 2024 рр.)

до середнього чеку контрагента, вартість на рядок, коефіцієнт комунікаційної інтенсивності, питома вартість продукту відносно середнього чеку). Пропущені значення інтервальних метрик імпутовано значенням «-1»; для фіксації першої взаємодії контрагента введено бінарний прапорець `is_first_order`.

Стандартна k-кратна перехресна перевірка є методологічно некоректною для часово залежних нестационарних даних: випадкове перемішування руйнує хронологічну послідовність і спричиняє витік інформації. Дослідження реалізовано за преквенційним протоколом (Predict-then-Learn) – стандартним підходом для оцінювання моделей в умовах нестационарності [29–31].

Протокол охоплює дві фази. На фазі ініціалізації перші 12 місяців даних (29.07.2017–29.07.2018; 11 157 записів; частка позитивного класу 69,76%) використано для початкового навчання базової моделі  $M_0$ . Упродовж потокової фази (76 місяців-ітерацій; серпень 2018 – листопад 2024; 75 546 записів; частка позитивного класу 62,09%) щомісяця виконується чотири кроки: модель  $M_t$  генерує прогнози для транзакцій поточного місяця  $D_t$  до перенавчання; обчислюються повні метрики класифікації з фіксацією ROC-AUC та LogLoss; записи місяця  $t$  додаються до загального сховища; нова навчальна вибірка  $W_t$  формується відповідно до поточної стратегії вікна і модель  $M_{t+1}$  повністю перенавчається на  $W_t$ . Такий порядок унеможливує витік даних, оскільки модель завжди оцінюється на ще непобачених прикладах.

Зниження частки успішних угод із 69,8% на фазі ініціалізації до 62,1% у потоковій фазі кількісно підтверджує зсув апріорної ймовірності (рис. 3). Ця структурна розбіжність обґрунтовує застосування динамічного балансування класів під час потокового перенавчання – зокрема, ітеративного перерахунку гіперпараметра `scale_pos_weight` для кожного нового зрізу навчального вікна. Протокол генерує часовий ряд метрик довжиною 76 точок на кожен стратегію, що забезпечує репрезентативний статистичний аналіз [31].

Порівнювалися одинадцять стратегій формування навчального вікна, об'єднаних у чотири категорії (табл. 1).

Кумулятивна стратегія (Full History) використовує всю доступну історію: обсяг вікна монотонно зростає від ~12 000 записів на початку потокової фази до ~86 700 наприкінці. Фракційні стратегії (Frac 1/2, Frac 1/3, Frac 1/4) відбирають фіксовану частку від найновіших записів, зберігаючи постійне співвідношення «свіжих» і «давніх» даних. Часові стратегії (Time 12M, Time 6M, Time 4M, Time 3M) обмежують вибірку календарним горизонтом: Time 6M містила від ~5 000 до ~8 400 записів залежно від сезонної активності. Лічильникові стратегії (Count 10K, Count 5K, Count 3K) фіксують абсолютну кількість найновіших записів, що забезпечує сталий обсяг, проте змінює часову глибину відповідно до інтенсивності транзакцій.

Як механізм предиктивного моделювання обрано XGBoost – індустріальний стандарт класифікації на фінансових і операційних табличних даних [32]. Алгоритм демонструє

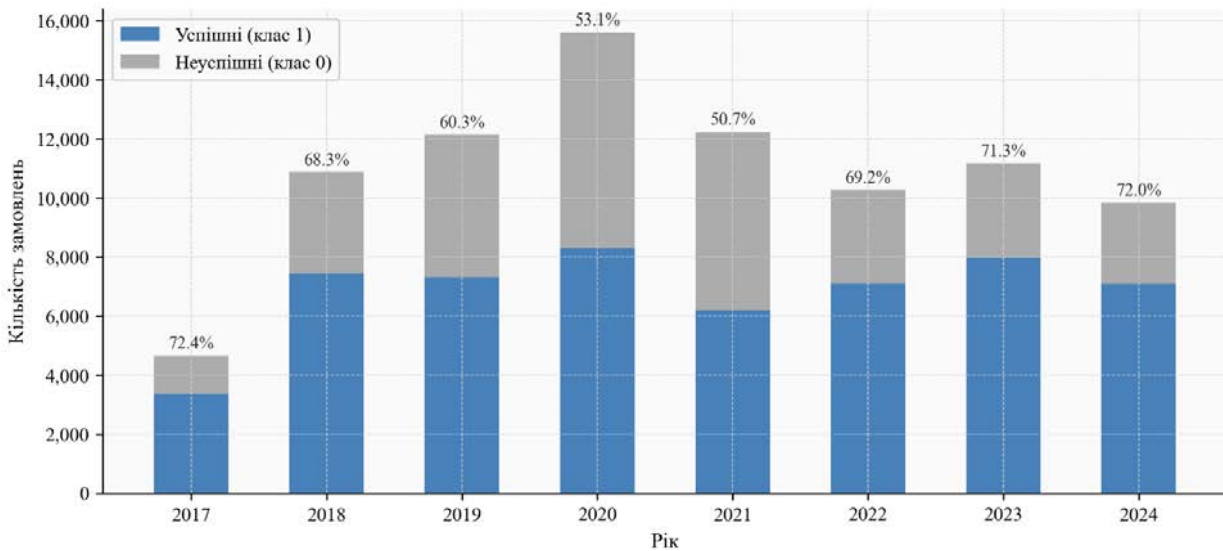


Рис. 2. Річна динаміка загального обсягу та структури B2B-замовлень за класами успішності (2017–2024 рр.)

стійкість до нелінійних залежностей, автоматичну обробку пропущених значень і стійкість до викидів. Об'єктивна функція  $L$  на ітерації  $t$  поєднує функцію втрат  $l$  та регуляризаційний член  $\Omega$ :

$$L^{(t)} = \sum_i l(y_i, \hat{y}_i^{(t)}) + \Omega(f_t)$$

В умовах концептуального дрейфу суперечність між старою та новою інформацією у навчальній вибірці спричиняє розмиття меж прийняття рішень, що робить алгоритм надзвичайно чутливим до обсягу навчального вікна [7; 27].

В усіх стратегіях використано єдину конфігурацію для забезпечення коректності порівняння (табл. 2).

Параметр *scale\_pos\_weight* обчислювався динамічно для кожного циклу перенавчання як

відношення негативних до позитивних прикладів у поточному навчальному вікні.

Якість класифікації оцінювалась за дев'ятьма метриками: ROC-AUC (площа під ROC-кривою), PR-AUC (площа під кривою точність–повнота), F1-міра, Ассурасу (загальна точність), Precision (точність), Recall (повнота), MCC (коефіцієнт кореляції Метьюса), Brier Score (оцінка каліброваності ймовірностей) та LogLoss (логарифмічна втрата). Для кожної метрики обчислювались глобальне агреговане значення, помісячне середнє, стандартне відхилення та коефіцієнт варіації (CV).

Для перевірки значущості відмінностей між стратегіями застосовано два непараметричні критерії. Оскільки преквенційний протокол генерує зв'язані вибірки (різні алгоритми на однакових часових блоках), параметричні тести

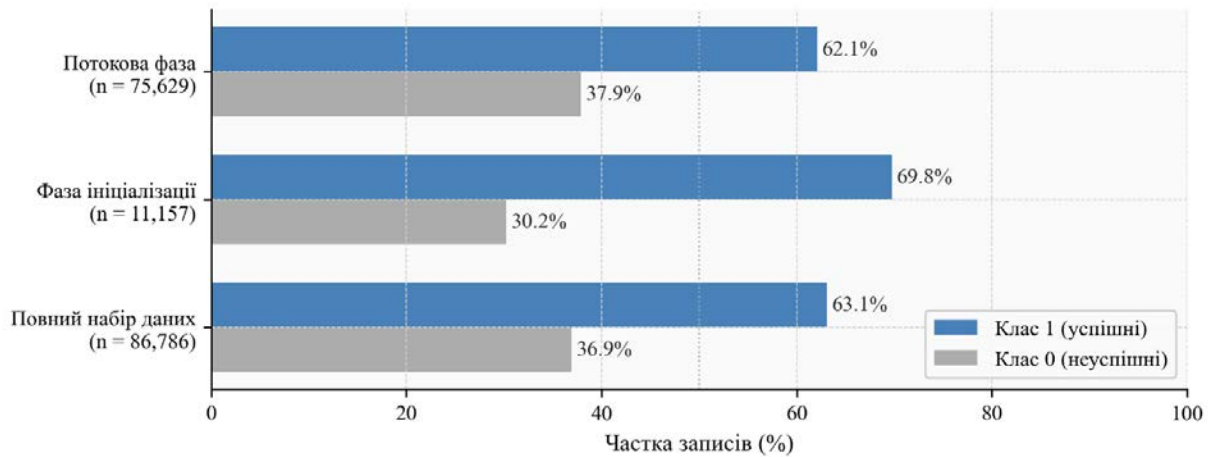


Рис. 3. Розподіл класів успішності транзакцій за фазами преквенційного симулювання

Таблиця 1

Стратегії формування навчального вікна

Категорія	Стратегія	Механізм формування	Динаміка обсягу
Кумулятивна	Full History	Увесь накопичений масив	Монотонне зростання (~12K → ~87K)
Фракційна	Frac 1/2	Останні 50% записів	Пропорційне зростання
Фракційна	Frac 1/3	Останні 33% записів	Пропорційне зростання
Фракційна	Frac 1/4	Останні 25% записів	Пропорційне зростання
Часова	Time 12M	Транзакції за останні 12 місяців	Флуктуючий
Часова	Time 6M	Транзакції за останні 6 місяців	Флуктуючий (~5K–8,4K)
Часова	Time 4M	Транзакції за останні 4 місяці	Флуктуючий
Часова	Time 3M	Транзакції за останні 3 місяці	Флуктуючий
Лічильникова	Count 10K	Останні 10 000 записів	Фіксований
Лічильникова	Count 5K	Останні 5 000 записів	Фіксований
Лічильникова	Count 3K	Останні 3 000 записів	Фіксований

Таблиця 2

Гіперпараметри XGBoost

Параметр	<i>n_estimators</i>	<i>learning_rate</i>	<i>max_depth</i>	<i>subsample</i>	<i>colsample_bytree</i>	<i>eval_metric</i>	<i>random_state</i>
Значення	200	0.05	6	0.8	0.8	logloss	42

є некоректними. Критерій Фрідмана – непараметричний аналог дисперсійного аналізу для повторних вимірювань – перевіряє нульову гіпотезу про рівність розподілів усіх стратегій. Як міра ефекту обчислено коефіцієнт узгодженості Кендалла  $W$  [33].

Попарні тести Вілкоксона зі знаком рангів проводились для всіх 55 пар стратегій із поправкою Бонферроні; як міра ефекту –  $r = Z/\sqrt{N}$ , де  $N$  – кількість місяців [34]. Додатково проведено ранжування стратегій за методикою Немені [33]: для кожного місяця стратегії ранжувалися за ROC-AUC (ранг 1 – найкращий), після чого обчислювався середній ранг за весь горизонт.

**Результати та дискусії.** За результатами 76-місячного преквенційного горизонту стратегія Time 6M досягла найвищих показників за шістьма з дев'яти метрик: ROC-AUC = 0,9440, PR-AUC = 0,9564, Accuracy = 0,8939, MCC = 0,7728, Brier = 0,0809 та LogLoss = 0,2732. Другу і третю позиції посіли Full History (ROC-AUC = 0,9435) та Frac 1/2 (0,9434), які статистично нерозрізненні за глобальним ROC-AUC. Найнижчі результати продемонстрували Time 12M (ROC-AUC = 0,9302) та Count 10K (0,9347). Абсолютний розмах між найкращою та найгіршою стратегією за ROC-AUC становить

0,0138 – помірна, але стабільна різниця на горизонті 76 місяців. За LogLoss розмах більш виражений: 0,0612 (Time 6M: 0,2732 проти Time 12M: 0,3344), що свідчить про суттєву відмінність у каліброваності ймовірнісних прогнозів.

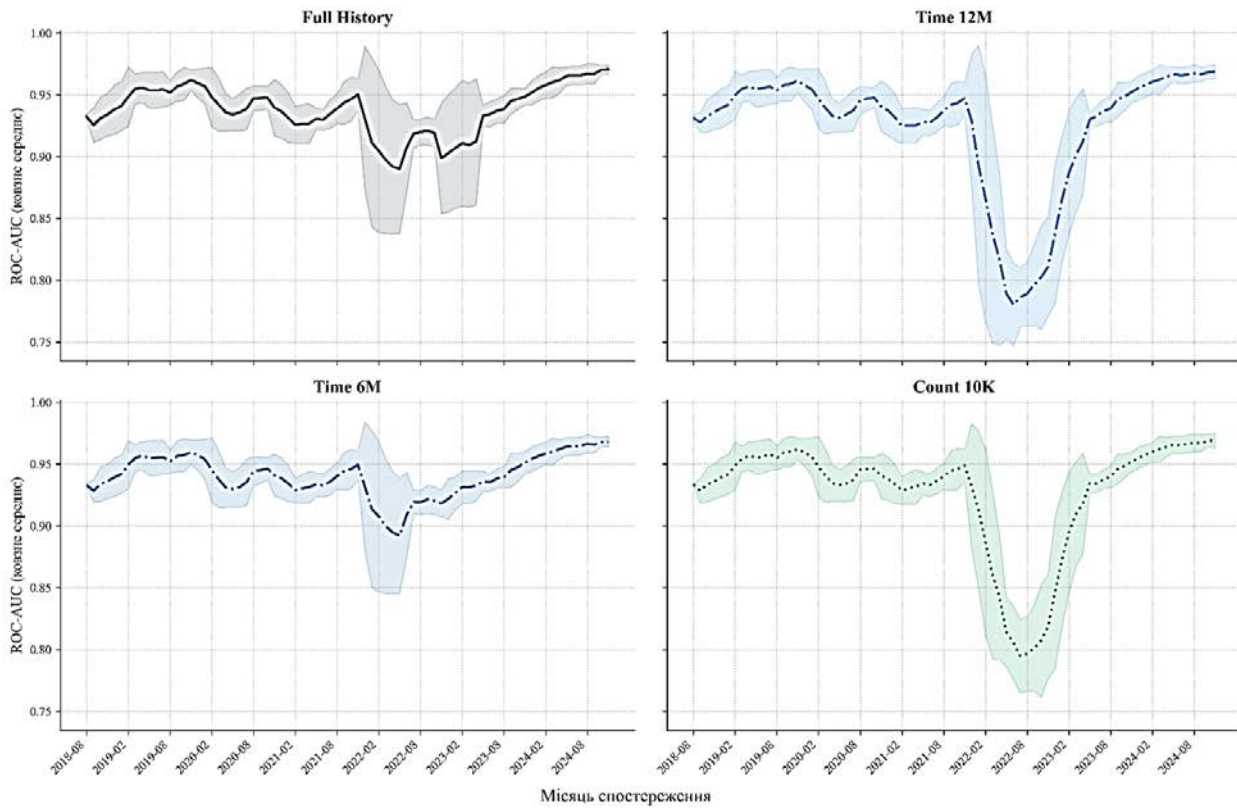
Результати спростовують парадигму «More data is better». Стратегія Full History, попри доступ до понад 80 тис. записів на пізніх ітераціях, не лідирує за жодною з ключових метрик, крім Precision = 0,8975. Frac 1/2 перевершує кумулятивну стратегію за F1 (0,9167 проти 0,9146), Accuracy (0,8937 проти 0,8919), MCC (0,7724 проти 0,7686) та LogLoss (0,2753 проти 0,2818): відкидання найстарішої половини даних не лише не погіршує дискримінаційну здатність, а й покращує каліброваність і баланс точності та повноти (рис. 4).

Аналіз варіабельності щомісячних метрик виявляє два чітко розділених кластери за стабільністю. Стратегії Time 6M, Time 4M, Time 3M, Count 5K, Count 3K та Frac 1/3 характеризуються стандартним відхиленням ROC-AUC у межах 0,025–0,030 і CV  $\approx$  0,027–0,032. Стратегії Time 12M (CV = 0,063) та Count 10K (CV = 0,056) демонструють виражену дисперсію, зумовлену їхньою вразливістю до структурних збурень.

Time 6M поєднує найвищу середню якість (mean ROC-AUC = 0,9404)



Рис. 4. Теплова карта глобальних метрик якості класифікації залежно від стратегії формування навчального вікна



**Рис. 5. Порівняння згладжених траєкторій ROC-AUC із довірчими інтервалами для ключових стратегій навчального вікна**

з найменшою дисперсією серед часових стратегій ( $\text{std} = 0,0253$ ), що визначає її оптимальність із погляду компромісу «ефективність–стабільність».

Під час макроекономічних шоків 2022 року стратегії з великим обсягом історичних даних (Time 12M, Count 10K) характеризуються екстремальним розширенням довірчого коридору, що підтверджує їх нестабільність у кризові періоди (рис. 5). Кумулятивна стратегія виявляє значну волатильність протягом 2022–2023 років унаслідок конфлікту акумульованих старих патернів із новими ринковими реаліями. Time 6M генерує найвужчий довірчий інтервал навіть у фазі кризового падіння метрик, що свідчить про її здатність до швидкої та структурно надійної адаптації.

При аналізі динаміки навчального вікна та обчислювальної ефективності стратегій отримані наступні результати (табл. 3).

Full History наприкінці горизонту оперувала вікном обсягом  $\sim 86\,700$  записів і потребувала 0,47 с на одне перенавчання. Time 6M за середнього вікна  $\sim 6\,000$  записів витратила 0,22 с – удвічі менше. Стратегії Count 3K і Time 3M вимагали лише 0,18–0,20 с, що становить менш ніж 40% від кумулятивного варіанту. Залежність якості від обсягу вікна є нелінійною: збільшення від 3 000 до 6 000 записів (Time 3M  $\rightarrow$  Time 6M)

дає приріст ROC-AUC на 0,0039 за мінімальних додаткових витрат (+0,04 с), тоді як збільшення від 6 000 до 51 000 (Time 6M  $\rightarrow$  Full History) не забезпечує жодного приросту якості при більш ніж подвійному зростанні часу перенавчання. Застосування Time 6M замість Full History забезпечує скорочення розміру навчальної матриці на 50-80% на пізніх ітераціях, що відповідає принципам Sustainable/Green AI і забезпечує відчутну економію ресурсів у хмарних інфраструктурах [35].

Таблиця 3

**Обчислювальна ефективність стратегій**

Стратегія	Середній обсяг вікна	Середній час перенавчання, с
Full History	51 028	0,47
Frac 1/2	25 514	0,37
Frac 1/3	17 009	0,28
Frac 1/4	12 757	0,26
Time 12M	11 967	0,26
Time 6M	5 994	0,22
Time 4M	3 999	0,20
Time 3M	3 003	0,18
Count 10K	10 000	0,32
Count 5K	5 000	0,23
Count 3K	3 000	0,20

Критерій Фрідмана підтвердив наявність статистично значущих відмінностей між стратегіями:  $\chi^2 = 99,89$  ( $p < 10^{-16}$ ). Коефіцієнт узгодженості Кендалла  $W = 0,131$  свідчить про малий-помірний ефект, узгоджений із невеликим абсолютним розмахом ROC-AUC між стратегіями.

Ранжування за методикою Нємені формує лідерську групу (позиції 1-4: Time 6M – ранг 4,90; Frac 1/2 – 4,96; Count 10K – 4,98; Full History – 5,11) з тісно згрупованими значеннями, що свідчить про конкурентоспроможність усіх чотирьох стратегій. Примітно, що Count 10K посідає третю позицію за рангом попри нижчий глобальний ROC-AUC завдяки високій медіані (0,9451), яку нівелюють глибокі провали у кризові місяці. Time 6M є єдиною стратегією, що лідирує як за глобальними метриками, так і за рейтинговим підходом, підтверджуючи найбільш збалансований характер.

З 55 попарних порівнянь Вілкоксона з поправкою Бонферроні 17 є статистично значущими ( $\alpha = 0,05$ ). Найбільш виражені відмінності зафіксовано для пар: Time 6M – Time 3M ( $p_{corr} < 10^{-7}$ ;  $r = 0,711$  – великий ефект), Time 6M – Count 3K ( $p_{corr} < 10^{-7}$ ;  $r = 0,706$  – великий ефект) та Time 4M – Time 3M ( $p_{corr} < 10^{-5}$ ;  $r = 0,617$  – великий ефект).

Відмінності між лідерами (Time 6M, Frac 1/2, Full History) є статистично незначущими ( $p_{corr} > 0,50$ ), що підтверджує їхню функціональну еквівалентність і допускає гнучкий вибір між ними залежно від обчислювальних обмежень.

Аналіз помісячної динаміки виявляє два основних кризових епізоди. У лютому-березні 2022 р. усі стратегії зазнали суттєвого зниження ROC-AUC (рис. 6). Для Full History він знизився до 0,813 (лютий) та 0,835 (березень) – відхилення на  $2,8\sigma$  від середнього. Time 6M продемонструвала аналогічне зниження (0,827 та 0,838), проте відновилася до 0,915 вже у квітні 2022 р. – на місяць раніше за кумулятивну стратегію. Count 10K та Time 12M зазнали глибшого провалу і тривалішого відновлення, що пояснює їхню вищу дисперсію. У кризовому епізоді січня 2023 р. Full History зафіксувала найнижче значення за весь горизонт (ROC-AUC = 0,809), тоді як Time 6M утримала рівень 0,914. Цей епізод ілюструє ключову вразливість кумулятивної стратегії: великий обсяг застарілих патернів уповільнює адаптацію до різких зсувів розподілу.

На відміну від нейронних мереж, де проблемою є «катастрофічне забування», тобто занадто швидке витіснення раніше засвоєних знань [36], для пакетних ансамблів дерев рішень

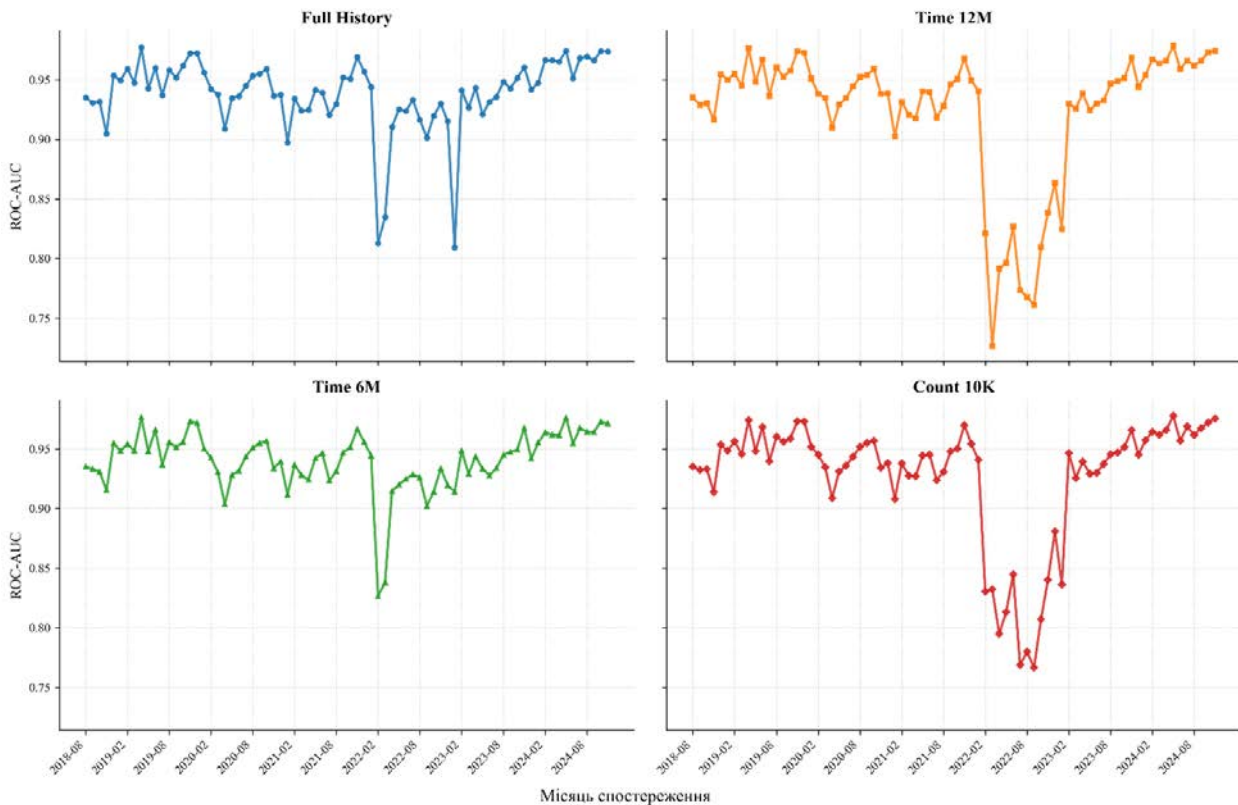


Рис. 6. Порівняння преквенційних траєкторій ROC-AUC для ключових стратегій навчального вікна в умовах кризових збурень

в умовах концептуального дрейфу спостерігається симетричний ефект надмірного збереження застарілих патернів. У рамках дилеми стабільності – пластичності [8] ці моделі тяжіють до надмірної стабільності: кумулятивні та широкі вікна примушують алгоритм зберігати вузли розгалуження, оптимізовані під нерелевантні поведінкові правила, що призводить до побудови надмірно складних «забруднених» дерев і деградації LogLoss. Пропонуємо позначати цей ефект терміном «катастрофічне запам'ятовування» (Catastrophic Retention). Вузькі ковзні вікна функціонують як механізм примусового витіснення застарілих концептів, дозволяючи XGBoost будувати лаконічні дерева, адекватні поточній ринковій реальності [26].

Оптимальність 6-місячного вікна є наслідком його структурного вирівнювання з фундаментальними ритмами B2B-сектору: корпоративне планування, перегляд бюджетів і тендерні процедури відбуваються з піврічною або кварталною частотою [22]. Шестимісячне вікно реалізує подвійний компроміс: обсяг у 5 000–8 400 транзакцій гарантує математичну збіжність градієнтного бустингу, одночасно виключаючи дані попереднього піврічного циклу, сформовані за іншої кон'юнктури. Вужчий горизонт (Time 3M) означає інформаційний дефіцит, ширший (Time 12M) — поглинання нерелевантних макропатернів. Лічильникові вікна при зростанні інтенсивності замовлень можуть стискатися в часі до 2–3 тижнів, спричиняючи гіперчутливість моделі та втрату здатності до генералізації; часове вікно Time 6M натомість природно масштабує абсолютний обсяг вибірки відповідно до бізнес-активності [22].

З ширшої перспективи, отримані результати спростовують парадигму «Data is an asset» стосовно предиктивного моделювання в нестационарних умовах: застарілі дані перетворюються з активу на технічний борг, що руйнує прогностичну здатність моделі [7]. Для промислових MLOps-конвеєрів механізм відсікання навчальної вибірки є фундаментальною архітектурною вимогою [19, 26]. Архітектура регулярного пакетного перенавчання з 6-місячним горизонтом є більш надійним і ресурсоефективним рішенням для ERP-систем порівняно зі складними статистичними детекторами дрейфу (ADWIN, DDM та аналогами), що вимагають моніторингу кожного потокового запису [7, 18]. Отримані результати узгоджуються з даними [23], розширюючи їхні висновки на пакетне

перенавчання XGBoost у B2B-контексті і додаючи кількісну оцінку відмінностей із статистичною верифікацією на значно більшому масиві реальних даних.

**Висновки.** Проведено систематичне порівняння одинадцяти стратегій формування навчального вікна для XGBoost на масиві 86 786 B2B-транзакцій реальної промислової ERP-системи протягом 76-місячного преквенційного горизонту.

Стратегія часового вікна 6 місяців (Time 6M) є оптимальною за сукупністю критеріїв: найвищий ROC-AUC (0,9440), найнижчий LogLoss (0,2732), висока стабільність (CV = 0,027), найкращий середній ранг (4,90) та помірно обчислювальне навантаження (0,22 с на цикл – удвічі менше за кумулятивну стратегію). Ефективність цього горизонту зумовлена структурним вирівнюванням із піврічними ритмами корпоративного планування: вікно одночасно захоплює актуальну ринкову динаміку та виключає нерелевантні патерни попередніх циклів.

Кумулятивна стратегія (Full History), попри максимальний обсяг навчальної вибірки, не демонструє переваг за якістю прогнозу. Деградація ROC-AUC до 0,809 в кризовому епізоді січня 2023 р. підтверджує ефект «катастрофічного запам'ятовування»: надмірне збереження застарілих патернів знижує адаптивність моделі в нестационарних умовах.

Відмінності між стратегіями є статистично значущими (критерій Фрідмана:  $\chi^2 = 99,89$ ;  $(p < 10^{-16})$ ;  $W = 0,131$ ). З 55 попарних порівнянь 17 є значущими (тест Вілкоксона з поправкою Бонферроні,  $\alpha = 0,05$ ). Стратегії з найвужчими вікнами (Time 3M, Count 3K) достовірно поступаються стратегіям із помірними вікнами, тоді як лідерська група (Time 6M, Frac 1/2, Full History) статистично нерозрізнявана, що допускає гнучкий вибір між ними залежно від обчислювальних обмежень.

Розробникам i-ERP рекомендовано впровадити стратегію 6-місячного ковзного вікна як базову архітектурну конфігурацію конвеєрів автоматичного перенавчання XGBoost-моделей. Перспективними напрямками подальших досліджень є адаптивні стратегії, що динамічно регулюють глибину навчального вікна залежно від рівня детектованого концептуального дрейфу; зважені вікна з вищим ваговим коефіцієнтом для новіших транзакцій; ансамблі, що комбінують моделі різних часових горизонтів.

#### ЛІТЕРАТУРА:

1. Grabski S. V., Leech S. A., Schmidt P. J. A Review of ERP Research: A Future Agenda for Accounting Information Systems. *Journal of Information Systems*. 2011. Vol. 25, No. 1. P. 37–78. DOI: 10.2308/jis.2011.25.1.37.

2. Jawad Z. N., Balázs V. Machine learning-driven optimization of enterprise resource planning (ERP) systems: a comprehensive review. *Beni-Suef University Journal of Basic and Applied Sciences*. 2024. Vol. 13. Art. 4. DOI: 10.1186/s43088-023-00460-y.
3. Romero D., Vernadat F. Enterprise information systems state of the art: Past, present and future trends. *Computers in Industry*. 2016. Vol. 79. P. 3–13. DOI: 10.1016/j.compind.2016.03.001.
4. Zdravković M., Panetto H. Artificial intelligence-enabled enterprise information systems. *Enterprise Information Systems*. 2022. Vol. 16, No. 5. Art. 1973570. DOI: 10.1080/17517575.2021.1973570.
5. Duan Y., Edwards J. S., Dwivedi Y. K. Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*. 2019. Vol. 48. P. 63–71. DOI: 10.1016/j.ijinfomgt.2019.01.021.
6. Villegas-Ch G., García-Ortiz D., Luján-Mora S. Machine Learning and Deep Learning Models for Demand Forecasting in Supply Chain Management: A Critical Review. *Applied System Innovation*. 2024. Vol. 7, No. 5. P. 93. DOI: 10.3390/asi7050093.
7. Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A survey on concept drift adaptation. *ACM Computing Surveys*. 2014. Vol. 46, No. 4. P. 1–37. DOI: 10.1145/2523813.
8. Lu J., Liu A., Dong F., Gu F., Gama J., Zhang G. Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*. 2019. Vol. 31, No. 12. P. 2346–2363. DOI: 10.1109/TKDE.2018.2876857.
9. Iwashita A. S., Papa J. P. Evolving strategies in machine learning: A systematic review of concept drift detection. *Information*. 2024. Vol. 15, No. 12. Art. 786. DOI: 10.3390/info15120786.
10. Abdul Razak M. S., Nirmala C. R., Sreenivasa B. R., Lahza H., Lahza H. F. M. A survey on detecting healthcare concept drift in AI/ML models from a finance perspective. *Frontiers in Artificial Intelligence*. 2023. Vol. 5. Art. 955314. DOI: 10.3389/frai.2022.955314.
11. Ditzler G., Roveri M., Alippi C., Polikar R. Learning in nonstationary environments: A survey. *IEEE Computational Intelligence Magazine*. 2015. Vol. 10, No. 4. P. 12–25. DOI: 10.1109/MCI.2015.2471196.
12. Kraus A., van der Aa H. Machine learning-based detection of concept drift in business processes. *Process Science*. 2025. Vol. 2. P. 5. DOI: 10.1007/s44311-025-00012-w.
13. Gama J., Medas P., Castillo G., Rodrigues P. Learning with drift detection. *Advances in Artificial Intelligence – SBIA 2004 : proceedings of the 17th Brazilian Symposium on Artificial Intelligence (Sao Luis, Maranhao, Brazil, Sept. 29 – Oct. 1, 2004) / ed. by A. L. C. Bazzan, S. Labidi*. Springer, 2004. P. 286–295. (Lecture Notes in Computer Science ; vol. 3171). DOI: 10.1007/978-3-540-28645-5\_29.
14. Baena-García M., Del Campo-Ávila J., Fidalgo R., Bifet A., Gavaldà R., Morales-Bueno R. Early drift detection method. *Proceedings of the 4th International Workshop on Knowledge Discovery from Data Streams (IWKDDs 2006)*. Berlin, 2006. P. 77–86.
15. Bifet A., Gavaldà R. Learning from time-changing data with adaptive windowing. *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM 2007)*. SIAM, 2007. P. 443–448. DOI: 10.1137/1.9781611972771.42.
16. Bayram F., Ahmed B. S., Kassler A. From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*. 2022. Vol. 245. Art. 108632. DOI: 10.1016/j.knsys.2022.108632.
17. Gomes H. M., Barddal J. P., Enembreck F., Bifet A. A survey on ensemble learning for data stream classification. *ACM Computing Surveys*. 2017. Vol. 50, No. 2. P. 1–36. DOI: 10.1145/3054925.
18. Suárez-Cetrulo A. L., Quintana D., Cervantes A. A survey on machine learning for recurring concept drifting data streams. *Expert Systems with Applications*. 2023. Vol. 213. Art. 118934. DOI: 10.1016/j.eswa.2022.118934.
19. Software engineering for machine learning: A case study / S. Amershi et al. *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP 2019)*. IEEE, 2019. P. 291–300. DOI: 10.1109/ICSE-SEIP.2019.00042.
20. Žliobaitė I., Gabrys B. Analysis of descriptors of concept drift and their impacts. *Informatics*. 2025. Vol. 12, No. 1. Art. 13. DOI: 10.3390/informatics12010013.
21. Agrahari S., Singh A. K. Concept drift detection in data stream mining: a literature review. *Journal of King Saud University – Computer and Information Sciences*. 2022. Vol. 34, No. 10. P. 9523–9540. DOI: 10.1016/j.jksuci.2021.11.006.
22. Žliobaitė I. Combining similarity in time and space for training set formation under concept drift. *Intelligent Data Analysis*. 2011. Vol. 15, No. 4. P. 589–611. DOI: 10.3233/IDA-2011-0484.
23. Kozal J., Guzy F., Woźniak M. Employing chunk size adaptation to overcome concept drift. *Journal of Universal Computer Science*. 2022. Vol. 28, No. 3. P. 249–268. DOI: 10.3897/jucs.80735.
24. Nikolaidis D., Doumpos M. Credit Scoring with Drift Adaptation Using Local Regions of Competence. *Operations Research Forum*. 2022. Vol. 3. P. 67. DOI: 10.1007/s43069-022-00177-1.
25. Webb G. I., Hyde R., Cao H., Nguyen H. L., Petitjean F. Characterizing concept drift. *Data Mining and Knowledge Discovery*. 2016. Vol. 30. P. 964–994. DOI: 10.1007/s10618-015-0448-4.
26. Iwashita A. S., Papa J. P. CODE: A moving-window-based framework for detecting concept drift. *Symmetry*. 2022. Vol. 14, No. 12. Art. 2508. DOI: 10.3390/sym14122508.

27. Mestiri S. Improved credit scoring model using XGBoost with Bayesian hyper-parameter optimization. *International Journal of Electrical and Computer Engineering*. 2021. Vol. 11, No. 6. P. 5477–5487. DOI: 10.11591/ijece.v11i6.pp5477-5487.
28. Ashofteh A., Bravo J. M. A conservative approach for online credit scoring. *Expert Systems with Applications*. 2021. Vol. 176. Art. 114835. DOI: 10.1016/j.eswa.2021.114835.
29. Gama J., Sebastião R., Rodrigues P. P. On evaluating stream learning algorithms. *Machine Learning*. 2013. Vol. 90, No. 3. P. 317–346. DOI: 10.1007/s10994-012-5320-9.
30. González Hidalgo J. I., Maciel B. I. F., Barros R. S. M. Experimenting with prequential variations for data stream learning evaluation. *Computational Intelligence*. 2019. Vol. 35. P. 670–692. DOI: 10.1111/coin.12208.
31. Dawid A. P. Statistical theory: The prequential approach (with discussion). *Journal of the Royal Statistical Society: Series A*. 1984. Vol. 147, No. 2. P. 278–292. DOI: 10.2307/2981683.
32. Chen T., Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, 2016. P. 785–794. DOI: 10.1145/2939672.2939785.
33. Demšar J. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*. 2006. Vol. 7, No. 1. P. 1–30. DOI: 10.5555/1248547.1248548.
34. Woolson R. F. Wilcoxon signed-rank test. *Wiley Encyclopedia of Clinical Trials*. 2007. DOI: 10.1002/9780471462422.eoct979.
35. Stratigakos A., Velásquez J., Johansson C. On the retraining frequency of global models in retail demand forecasting. *Machine Learning with Applications*. 2025. Vol. 19. Art. 100769. DOI: 10.1016/j.mlwa.2025.100769.
36. Overcoming catastrophic forgetting in neural networks / J. Kirkpatrick et al. *Proceedings of the National Academy of Sciences*. 2017. Vol. 114, No. 13. P. 3521–3526. DOI: 10.1073/pnas.1611835114.

#### REFERENCES:

1. Grabski, S. V., Leech, S. A., & Schmidt, P. J. (2011). A review of ERP research: A future agenda for accounting information systems. *Journal of Information Systems*, 25(1), 37–78. <https://doi.org/10.2308/jis.2011.25.1.37>
2. Jawad, Z. N., & Balázs, V. (2024). Machine learning-driven optimization of enterprise resource planning (ERP) systems: A comprehensive review. *Beni-Suef University Journal of Basic and Applied Sciences*, 13, Article 4. <https://doi.org/10.1186/s43088-023-00460-y>
3. Romero, D., & Vernadat, F. (2016). Enterprise information systems state of the art: Past, present and future trends. *Computers in Industry*, 79, 3–13. <https://doi.org/10.1016/j.compind.2016.03.001>
4. Zdravković, M., & Panetto, H. (2022). Artificial intelligence-enabled enterprise information systems. *Enterprise Information Systems*, 16(5), Article 1973570. <https://doi.org/10.1080/17517575.2021.1973570>
5. Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data – evolution, challenges and research agenda. *International Journal of Information Management*, 48, 63–71. <https://doi.org/10.1016/j.ijinfomgt.2019.01.021>
6. Villegas-Ch, G., García-Ortiz, D., & Luján-Mora, S. (2024). Machine learning and deep learning models for demand forecasting in supply chain management: A critical review. *Applied System Innovation*, 7(5), Article 93. <https://doi.org/10.3390/asi7050093>
7. Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37. <https://doi.org/10.1145/2523813>
8. Lu, J., Liu, A., Dong, F., Gu, F., Gama, J., & Zhang, G. (2019). Learning under concept drift: A review. *IEEE Transactions on Knowledge and Data Engineering*, 31(12), 2346–2363. <https://doi.org/10.1109/TKDE.2018.2876857>
9. Iwashita, A. S., & Papa, J. P. (2024). Evolving strategies in machine learning: A systematic review of concept drift detection. *Information*, 15(12), Article 786. <https://doi.org/10.3390/info15120786>
10. Abdul Razak, M. S., Nirmala, C. R., Sreenivasa, B. R., Lahza, H., & Lahza, H. F. M. (2023). A survey on detecting healthcare concept drift in AI/ML models from a finance perspective. *Frontiers in Artificial Intelligence*, 5, Article 955314. <https://doi.org/10.3389/frai.2022.955314>
11. Ditzler, G., Roveri, M., Alippi, C., & Polikar, R. (2015). Learning in nonstationary environments: A survey. *IEEE Computational Intelligence Magazine*, 10(4), 12–25. <https://doi.org/10.1109/MCI.2015.2471196>
12. Kraus, A., & van der Aa, H. (2025). Machine learning-based detection of concept drift in business processes. *Process Science*, 2, Article 5. <https://doi.org/10.1007/s44311-025-00012-w>
13. Gama, J., Medas, P., Castillo, G., & Rodrigues, P. (2004). Learning with drift detection. In A. L. C. Bazzan & S. Labidi (Eds.), *Advances in Artificial Intelligence – SBIA 2004* (Vol. 3171, pp. 286–295). Springer. [https://doi.org/10.1007/978-3-540-28645-5\\_29](https://doi.org/10.1007/978-3-540-28645-5_29)
14. Baena-García, M., Del Campo-Ávila, J., Fidalgo, R., Bifet, A., Gavaldà, R., & Morales-Bueno, R. (2006). Early drift detection method. *Proceedings of the 4th International Workshop on Knowledge Discovery from Data Streams (IWKDDs 2006)*, 77–86.
15. Bifet, A., & Gavaldà, R. (2007). Learning from time-changing data with adaptive windowing. *Proceedings of the 2007 SIAM International Conference on Data Mining (SDM 2007)*, 443–448. SIAM. <https://doi.org/10.1137/1.9781611972771.42>

16. Bayram, F., Ahmed, B. S., & Kassler, A. (2022). From concept drift to model degradation: An overview on performance-aware drift detectors. *Knowledge-Based Systems*, 245, Article 108632. <https://doi.org/10.1016/j.knosys.2022.108632>
17. Gomes, H. M., Barddal, J. P., Enembreck, F., & Bifet, A. (2017). A survey on ensemble learning for data stream classification. *ACM Computing Surveys*, 50(2), 1–36. <https://doi.org/10.1145/3054925>
18. Suárez-Cetrulo, A. L., Quintana, D., & Cervantes, A. (2023). A survey on machine learning for recurring concept drifting data streams. *Expert Systems with Applications*, 213, Article 118934. <https://doi.org/10.1016/j.eswa.2022.118934>
19. Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP 2019)*, 291–300. IEEE. <https://doi.org/10.1109/ICSE-SEIP.2019.00042>
20. Žliobaitė, I., & Gabrys, B. (2025). Analysis of descriptors of concept drift and their impacts. *Informatics*, 12(1), Article 13. <https://doi.org/10.3390/informatics12010013>
21. Agrahari, S., & Singh, A. K. (2022). Concept drift detection in data stream mining: A literature review. *Journal of King Saud University – Computer and Information Sciences*, 34(10), 9523–9540. <https://doi.org/10.1016/j.jksuci.2021.11.006>
22. Žliobaitė, I. (2011). Combining similarity in time and space for training set formation under concept drift. *Intelligent Data Analysis*, 15(4), 589–611. <https://doi.org/10.3233/IDA-2011-0484>
23. Kozal, J., Guzy, F., & Woźniak, M. (2022). Employing chunk size adaptation to overcome concept drift. *Journal of Universal Computer Science*, 28(3), 249–268. <https://doi.org/10.3897/jucs.80735>
24. Nikolaidis, D., & Doumpos, M. (2022). Credit scoring with drift adaptation using local regions of competence. *Operations Research Forum*, 3, Article 67. <https://doi.org/10.1007/s43069-022-00177-1>
25. Webb, G. I., Hyde, R., Cao, H., Nguyen, H. L., & Petitjean, F. (2016). Characterizing concept drift. *Data Mining and Knowledge Discovery*, 30, 964–994. <https://doi.org/10.1007/s10618-015-0448-4>
26. Iwashita, A. S., & Papa, J. P. (2022). CODE: A moving-window-based framework for detecting concept drift. *Symmetry*, 14(12), Article 2508. <https://doi.org/10.3390/sym14122508>
27. Mestiri, S. (2021). Improved credit scoring model using XGBoost with Bayesian hyper-parameter optimization. *International Journal of Electrical and Computer Engineering*, 11(6), 5477–5487. <https://doi.org/10.11591/ijece.v11i6.pp5477-5487>
28. Ashofteh, A., & Bravo, J. M. (2021). A conservative approach for online credit scoring. *Expert Systems with Applications*, 176, Article 114835. <https://doi.org/10.1016/j.eswa.2021.114835>
29. Gama, J., Sebastião, R., & Rodrigues, P. P. (2013). On evaluating stream learning algorithms. *Machine Learning*, 90(3), 317–346. <https://doi.org/10.1007/s10994-012-5320-9>
30. González Hidalgo, J. I., Maciel, B. I. F., & Barros, R. S. M. (2019). Experimenting with prequential variations for data stream learning evaluation. *Computational Intelligence*, 35, 670–692. <https://doi.org/10.1111/coin.12208>
31. Dawid, A. P. (1984). Statistical theory: The prequential approach (with discussion). *Journal of the Royal Statistical Society: Series A*, 147(2), 278–292. <https://doi.org/10.2307/2981683>
32. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, 785–794. ACM. <https://doi.org/10.1145/2939672.2939785>
33. Demšar, J. (2006). Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research*, 7(1), 1–30. <https://doi.org/10.5555/1248547.1248548>
34. Woolson, R. F. (2007). Wilcoxon signed-rank test. *Wiley Encyclopedia of Clinical Trials*. <https://doi.org/10.1002/9780471462422.eoct979>
35. Stratigakos, A., Velásquez, J., & Johansson, C. (2025). On the retraining frequency of global models in retail demand forecasting. *Machine Learning with Applications*, 19, Article 100769. <https://doi.org/10.1016/j.mlwa.2025.100769>
36. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A. A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., Hassabis, D., Clopath, C., Kumaran, D., & Hadsell, R. (2017). Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13), 3521–3526. <https://doi.org/10.1073/pnas.1611835114>



Стаття поширюється на умовах ліцензії відкритого доступу CC BY 4.0

Дата першого надходження статті до видання: 12.04.2026  
 Дата прийняття статті до друку після рецензування: 07.05.2026  
 Дата публікації (оприлюднення) статті: 30.05.2026