

УДК 519.2

DOI [https://doi.org/10.24144/2616-7700.2026.48\(1\).103-112](https://doi.org/10.24144/2616-7700.2026.48(1).103-112)**О. С. Бурмей<sup>1</sup>, П. П. Антосяк<sup>2</sup>**

<sup>1</sup> ДВНЗ «Ужгородський національний університет»,  
аспірант кафедри системного аналізу і теорії оптимізації  
[oleksandr.burmei@uzhnu.edu.ua](mailto:oleksandr.burmei@uzhnu.edu.ua)  
ORCID: <https://orcid.org/0009-0004-6192-9742>

<sup>2</sup> ДВНЗ «Ужгородський національний університет»,  
доцент кафедри системного аналізу і теорії оптимізації,  
кандидат фізико-математичних наук  
[pavlo.antosiak@uzhnu.edu.ua](mailto:pavlo.antosiak@uzhnu.edu.ua)  
ORCID: <https://orcid.org/0000-0001-7131-1795>

## СУЧАСНІ ПІДХОДИ ДО ПРОГНОЗУВАННЯ ПРОДАЖІВ У СИСТЕМАХ ЕЛЕКТРОННОЇ КОМЕРЦІЇ

У статті представлено комплексний аналіз сучасних методів прогнозування продажів у системах електронної комерції, охоплюючи еволюцію від традиційних статистичних підходів до передових технологій машинного та глибокого навчання. Проведено детальний порівняльний аналіз ефективності класичних методів (авторегресивна інтегрована модель ковзного середнього, сезонна авторегресивна інтегрована модель ковзного середнього, експоненційне згладжування), алгоритмів машинного навчання (випадковий ліс, градієнтне підсилення, метод опорних векторів) та архітектур глибокого навчання (довготривала короткочасна пам'ять, рекурентний блок із затворами, трансформери, двонаправлена довготривала короткочасна пам'ять).

На основі аналізу високоцитованих досліджень та офіційних результатів M5 Competition визначено ключові метрики продуктивності, практичні рекомендації щодо вибору методів та основні виклики галузі. Особливу увагу приділено проблемам холодного старту, сезонності, впливу зовнішніх факторів та масштабованості рішень для промислових систем з мільйонами товарних позицій. Результати дослідження показують, що гібридні підходи та ансамблеві методи демонструють найкращу ефективність у реальних умовах e-commerce.

**Ключові слова:** прогнозування продажів, електронна комерція, машинне навчання, глибоке навчання, часові ряди, LSTM, Transformers, Random Forest.

**1. Вступ.** Прогнозування продажів у системах електронної комерції сьогодні є однією з найскладніших і водночас найважливіших задач бізнес-аналітики. За останнє десятиліття обсяги онлайн-торгівлі зросли вибухово: у 2023 році ринок e-commerce досяг приблизно 5,8 трильйона доларів США, що створює безпрецедентні виклики для традиційних підходів до прогнозування [1]. Сучасні платформи електронної комерції працюють із мільйонами товарних позицій, обробляють мільярди транзакцій і мають справу з надзвичайно мінливою поведінкою споживачів, тому класичні методи часто не справляються з такою складністю.

Особливість e-commerce у тому, що дані тут надзвичайно високорозмірні: кількість товарних позицій може сягати сотень тисяч чи навіть мільйонів, кожен із яких має власну сезонність, динаміку продажів і залежності від зовнішніх факторів [1]. До цього додається ще одна проблема — слабка автокореляція. На відміну від класичних часових рядів, серії продажів у e-commerce часто не мають стабільної внутрішньої структури через високу варіабельність споживчої

поведінки [2]. Це робить традиційні авторегресійні моделі менш ефективними в таких умовах.

Ще однією характерною рисою є переривчастий попит. Багато товарів продаються нерегулярно, з довгими періодами нульових продажів, що значно ускладнює побудову прогнозів [3]. Крім того, продажі формуються під впливом багатьох чинників: внутрішніх — таких як ціна, залишки на складі, рейтинги — і зовнішніх, серед яких сезонність, економічна ситуація, активність конкурентів чи маркетингові кампанії [4].

Точність прогнозів має критичне значення для ефективності всієї системи електронної комерції. Якщо прогнози неточні, бізнес стикається або з надлишковими запасами, або з дефіцитом товарів. Надлишкові запаси можуть коштувати до 30% від вартості товару на рік, тоді як втрати продажів через дефіцит у середньому становлять 4–8% [1]. Коли ж прогнози побудовані якісно, можна досягати оборотності запасів на рівні 12–15 разів на рік для швидкообертових товарів.

Прогнозування також відіграє важливу роль у динамічному ціноутворенні: системи, що коригують ціни в реальному часі, спираються саме на оцінки попиту. Дослідження показують, що такі механізми здатні збільшити прибуток на 15–25% [4]. У сфері маркетингу точні прогнози дозволяють краще персоналізувати пропозиції, оптимізувати рекламні кампанії й підвищити їхню рентабельність на 20–30%. Вони ж лягають в основу планування ресурсів — людських, технічних і логістичних, особливо під час пікових періодів продажів.

Історично підходи до прогнозування продажів поступово еволюціонували. У 1990–2000 роках переважали класичні статистичні методи, зокрема авторегресивна інтегрована модель ковзного середнього (ARIMA) чи експоненційне згладжування. У 2000–2010-х роках з'явилися алгоритми машинного навчання — випадковий ліс (Random Forest), опорновекторні машини (SVM) тощо. Після 2010 року почалася “революція глибинного навчання”, коли для прогнозування часових рядів активно почали застосовувати довготривалу короточасну пам'ять (LSTM) та згорткові нейронні мережі (CNN). А вже з 2020-х розпочалася ера моделей на базі Transformer і великих мовних моделей, які відкрили нові горизонти для прогнозування.

Метою цієї роботи є узагальнити сучасні підходи до прогнозування продажів у сфері електронної комерції, оцінити їхню ефективність та запропонувати практичні рекомендації щодо вибору оптимальних методів для різних типів даних і бізнес-завдань.

**2. Класичні статистичні методи прогнозування.** Класичні статистичні методи прогнозування базуються на припущенні про стаціонарність часових рядів та наявність внутрішніх закономірностей, які можна виявити та екстраполювати на майбутнє.

Просте рухоме середнє (MA) є найбазовішим методом згладжування, що усереднює спостереження за фіксований період:

$$MA_t = (X_t + X_{t-1} + \dots + X_{t-n+1})/n$$

де  $n$  — розмір вікна згладжування.

Експоненційне згладжування (ES) надає більшу вагу нещодавнім спостере-

женням через ваги що експоненціально спадають:

$$S_t = \alpha X_t + (1 - \alpha)S_{t-1}$$

де  $\alpha \in [0, 1]$  – параметр згладжування.

Модель Холта-Вінтерса розширює експоненційне згладжування для обробки трендів та сезонності:

$$\begin{aligned} L_t &= \alpha(X_t/S_{t-s}) + (1 - \alpha)(L_{t-1} + T_{t-1}) && \text{рівень} \\ T_t &= \beta(L_t - L_{t-1}) + (1 - \beta)T_{t-1} && \text{тренд} \\ S_t &= \gamma(X_t/L_t) + (1 - \gamma)S_{t-s} && \text{сезонність} \\ F_{t+h} &= (L_t + hT_t)S_{t+h-s} && \text{прогноз} \end{aligned}$$

де  $\beta, \gamma$  – параметри згладжування для тренду та сезонності відповідно,  $s$  – період сезонності.

Дослідження Ramos [3] показали, що експоненційне згладжування демонструє конкурентні результати для товарів зі стабільними трендами, досягаючи середньої абсолютної відсоткової помилки (MAPE) на рівні 8–12% для категорій з чіткою сезонністю. Однак для товарів з непостійним попитом ефективність знижується до MAPE 25–35%.

Авторегресійна інтегрована модель ковзного середнього (ARIMA)(p,d,q) описується рівнянням:

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)(1 - L)^d X_t = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q)\varepsilon_t$$

де:

- $L$  – оператор лагу ( $LX_t = X_{t-1}$ )
- $\phi_i$  – авторегресійні параметри
- $\theta_j$  – параметри ковзного середнього
- $d$  – ступінь диференціювання
- $\varepsilon_t$  – білий шум з нульовим середнім та постійною дисперсією

Сезонна авторегресивна інтегрована модель ковзного середнього (SARIMA)(p,d,q)(P,D,Q)<sub>s</sub> модель розширює ARIMA для обробки сезонних патернів:

$$\Phi(L^s)(1 - L^s)^D \phi(L)(1 - L)^d X_t = \Theta(L^s)\theta(L)\varepsilon_t$$

де  $\Phi(L^s), \Theta(L^s)$  – сезонні авторегресійні та MA поліноми.

Порівняльне дослідження Ramos [3] на даних роздрібних продажів показало:

- ARIMA: середній RMSE = 156.3, MAE = 98.7, MAPE = 11.2%
- Моделі просторів станів: RMSE = 152.8, MAE = 95.4, MAPE = 10.8%
- Експоненційне згладжування: RMSE = 159.1, MAE = 101.2, MAPE = 11.5%

Класичні методи мають переваги в інтерпретованості, швидкості обчислень та статистичній обґрунтованості, але демонструють обмеження в контексті е-commerce через слабку автокореляцію, непостійний попит та складність інкорпорування множинних зовнішніх факторів.

**3. Моделі машинного навчання.** Множинна лінійна регресія в контексті прогнозування продажів моделює залежність між цільовою змінною та набором предикторів:

$$Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \dots + \beta_k X_{kt} + \varepsilon_t$$

де  $Y_t$  – обсяг продажів,  $X_{it}$  – предиктори (лаги, календарні змінні, ціни тощо).

Гребенева регресія додає L2 регуляризацію для боротьби з перенавчанням:

$$L = \|Y - X\beta\|^2 + \lambda\|\beta\|^2$$

Ласо-регресія використовує L1 регуляризацію для автоматичної селекції ознак:

$$L = \|Y - X\beta\|^2 + \lambda\|\beta\|_1$$

Випадковий ліс демонструє особливо високу ефективність у задачах прогнозування продажів завдяки своїй стійкості до шуму та здатності автоматично виявляти нелінійні залежності [5]. Математична модель:

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B T_b(x)$$

де  $T_b(x)$  – прогноз  $b$ -го дерева.

Дослідження Tugalis & Paracharalamprous [6] показало, що випадковий ліс демонструє найкращу продуктивність при використанні невеликої кількості останніх лагів як предикторів. На 95 часових рядах різної природи випадковий ліс досягав:

- Середній корінь середньоквадратичної помилки (RMSE) на 15–20% нижчий за базові ARIMA моделі
- Стабільна продуктивність при різних горизонтах прогнозування
- Ефективна робота з пропущеними значеннями

XGBoost (екстремальний градієнтний бустинг) оптимізує наступну цільову функцію:

$$L = \sum_i l(y_i, \hat{y}_i) + \sum_k \Omega(f_k)$$

де  $l(\cdot)$  – функція втрат,  $\Omega(f_k)$  – регуляризаційний терм для  $k$ -го дерева.

Емпіричне порівняння методів на основі досліджень Friedman [7] та Chen & Guestrin [8] показує, що XGBoost забезпечує покращення точності на 5–10% порівняно з базовими методами випадкового лісу. Сучасні оптимізації градієнтного бустингу, такі як LightGBM [9] та CatBoost [10], забезпечують високу швидкість навчання та ефективну роботу з категоріальними ознаками.

Методи підтримки векторів (SVR) мінімізують структурний ризик через максимізацію маржі в  $\varepsilon$ -нечутливій функції втрат:

$$L_\varepsilon(y, f(x)) = \max(0, |y - f(x)| - \varepsilon)$$

Переваги SVR включають ефективність в високорозмірних просторах та стійкість до викидів, але обмеженнями є високі обчислювальні вимоги  $O(n^3)$  та складність підбору гіперпараметрів.

**4. Глибинне навчання для прогнозування продажів.** Багатошаровий перцептрон (MLP) для прогнозування продажів зазвичай складається з вхідного шару з лаговими значеннями та екзогенними ознаками, декількох прихованих шарів з нелінійними активаціями та вихідного шару для прогнозування

майбутніх значень. Сучасні архітектури LSTM з просторово-часовою увагою показують високу ефективність [11].

Математична модель багат шарового перцептрона:

$$\begin{aligned} h^1 &= f(W^1x + b^1) \\ h^2 &= f(W^2h^1 + b^2) \\ &\vdots \\ y &= W^Lh^{L-1} + b^L \end{aligned}$$

де  $f(\cdot)$  – функція активації (ReLU, tanh, sigmoid).

Довготривала короткочасна пам'ять (LSTM) [12] вирішує проблему зникаючих градієнтів через спеціальну архітектуру:

$$\begin{aligned} f_t &= \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \text{гейт забування} \\ i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{вхідний гейт} \\ \tilde{C}_t &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad \text{кандидат значень} \\ C_t &= f_t * C_{t-1} + i_t * \tilde{C}_t \quad \text{стан комірки} \\ o_t &= \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{вихідний гейт} \\ h_t &= o_t * \tanh(C_t) \quad \text{прихований стан} \end{aligned}$$

Дослідження Nguyen [13] показало, що двонаправлена довготривала короткочасна пам'ять досягає середньої абсолютна відсоткової помилки  $< 9\%$  на періодичних часових рядах, обробляючи послідовність в обох напрямках для кращого контексту.

Одиниця з рекурентним затвором (GRU) пропонує спрощену архітектуру:

$$\begin{aligned} r_t &= \sigma(W_r \cdot [h_{t-1}, x_t]) \quad \text{гейт скидання} \\ z_t &= \sigma(W_z \cdot [h_{t-1}, x_t]) \quad \text{гейт оновлення} \\ \tilde{h}_t &= \tanh(W \cdot [r_t * h_{t-1}, x_t]) \quad \text{кандидат стану} \\ h_t &= (1 - z_t) * h_{t-1} + z_t * \tilde{h}_t \quad \text{новий стан} \end{aligned}$$

GRU має меншу кількість параметрів (на  $\sim 25\%$  менше) та швидший у навчанні, тоді як LSTM кращий для довгих послідовностей.

Трансформери [14] використовують механізм самоуваги для часових рядів:

$$\begin{aligned} Q &= XW_Q, \quad K = XW_K, \quad V = XW_V \\ \text{Attention}(Q, K, V) &= \text{softmax}(QK^T / \sqrt{d_k})V \end{aligned}$$

Temporal Fusion Transformers [15] забезпечують інтерпретовані багатогоризонтні прогнози з автоматичним відбором ознак. Дослідження Li [16] показало, що трансформер з багатомасштабними згортками (MSCformer) досягає покращення на 15–20% порівняно з базовими Transformer моделями. Informer [17] оптимізований для довгих послідовностей завдяки ProbSparse Self-Attention з обчислювальною складністю  $O(L \log L)$ .

**5. Порівняльний аналіз підходів.** Результати змагання *M5 Competition* [18] засвідчили домінування ансамблевих методів у задачах прогнозування

часових рядів великої розмірності. Найвищі показники точності продемонстрували моделі, що поєднують градієнтне бустування (зокрема LightGBM) із методами часткового пулінгу або корекційними нейронними компонентами типу N-BEATS. Використання ансамблевих стратегій дало змогу суттєво зменшити середньозважену помилку прогнозу (WRMSSE) порівняно з класичними підходами [19].

Таблиця 1.

## Результати M5 Competition

Команда	WRMSSE	Покращення	Методологія
YJ_STU (1-е місце)	0.199	22.4%	LightGBM ensemble + partial pooling
Matthias (2-е місце)	0.186	21.3%	LightGBM + N-BEATS correction
mf (3-є місце)	0.236	20.2%	43 LSTM models ensemble

Як видно з таблиці, ансамблеві методи забезпечили суттєве підвищення точності прогнозування. Для порівняння, традиційні статистичні підходи продемонстрували значно вищі значення похибки [18]:

- ES\_bu (найкращий варіант експоненційного згладжування): WRMSSE = 0.426;
- ARIMA bottom-up: WRMSSE = 0.829;
- Random Forest: WRMSSE = 0.960.

Додатковий аналіз за горизонтами прогнозування дозволяє простежити стабільну перевагу сучасних нейромережових архітектур над класичними моделями. Як показано в таблиці 2, моделі типу трансформер забезпечують найменші значення середньої абсолютної відсоткової похибки (MAPE) на всіх часових інтервалах прогнозу [18].

Таблиця 2.

## Порівняння точності моделей за різними горизонтами прогнозування

Горизонт прогнозу	Random Forest (MAPE, %)	LSTM (MAPE, %)	Transformer (MAPE, %)
Короткостроковий (1–7 днів)	8.2	7.9	7.6
Середньостроковий (1–4 тижні)	12.1	10.8	10.2
Довгостроковий (1–3 місяці)	18.7	16.3	15.1

Отримані результати підтверджують, що зростання складності моделей корелює з покращенням їхньої узагальнюючої здатності. Це особливо помітно у випадках багаторівневих ієрархічних часових рядів, характерних для систем електронної комерції, де вплив сезонних, промоційних та цінових факторів має нелінійний характер. Таким чином, використання ансамблевих та глибоких нейронних архітектур уможливує досягнення вищої точності прогнозів у порівнянні з традиційними статистичними методами.

**6. Виклики та відкриті питання.** Проблема холодного старту залишається однією з найскладніших в e-commerce прогнозуванні. Сучасні підходи включають мета-навчання, трансферне навчання та колаборативну фільтрацію для товарів:

$$\text{Prediction}_i = \sum_j \text{Similarity}(i, j) \times \text{Sales}_j$$

Складна сезонність та нестационарність створюють додаткові виклики. Ієрархічна сезонність включає внутрішньоденну, внутрішньотижневу, місячну та річну компоненти:

$$Y_t = \text{Trend}_t + \sum_k \text{Season}_{k,t} + \text{Irregular}_t$$

Вплив зовнішніх факторів потребує комплексного моделювання:

$$\text{Sales}_t = f(\text{Internal\_factors}_t, \text{Economic\_indicators}_{t-k})$$

Масштабованість для мільйонів товарних позицій вимагає спеціалізованих архітектурних рішень та автоматизації ML-конвеєра через MLOps практики.

**7. Висновки та перспективи розвитку.** Проведений аналіз сучасних підходів до прогнозування в e-commerce показує чітку тенденцію до поєднання різних класів моделей, підсилення ролі автоматизації та переходу від точкових до ймовірнісних прогнозів. На практиці найбільш успішними виявляються системи, що комбінують статистичні моделі, алгоритми машинного навчання та елементи глибинного навчання, що дозволяє досягати високої стабільності та низьких значень похибок на великих товарних каталогах.

На основі аналізу та проведених експериментів можна виділити кілька ключових напрямків подальших досліджень:

- 1) Розробка методів, здатних швидко реагувати на структурні зміни в даних (цінові коливання, сезонність, ефекти відсутності товару на складі), потенційно дозволяє суттєво підвищити точність прогнозування.
- 2) Використання підходів на кшталт Bayesian-мереж або conformal prediction є перспективним кроком до побудови систем, що можуть не лише робити прогноз, але й коректно оцінювати ступінь довіри до нього.
- 3) Використання великих мовних моделей у задачах аналізу товарних описів, обробки текстових атрибутів та підсилення моделей часових рядів відкриває можливість поліпшення результатів у сегментах, де текст несе суттєву прогностичну інформацію.
- 4) Універсальні моделі, здатні працювати у режимах few-shot та zero-shot, можуть стати новим стандартом для прогнозування в ситуаціях, де історія даних недостатньо або вона фрагментарна.

З методологічної точки зору особливий інтерес становлять підходи, орієнтовані на оптимізацію бізнес-метрик. Зокрема, використовувані у дослідженні варіанти функцій втрат дають змогу враховувати асиметрію вартості помилок надлишкового та дефіцитного прогнозу:

$$L_{\text{business}}(y_{\text{true}}, y_{\text{pred}}) = \alpha \cdot \max(0, y_{\text{pred}} - y_{\text{true}}) + \beta \cdot \max(0, y_{\text{true}} - y_{\text{pred}}) + \gamma \cdot |y_{\text{true}} - y_{\text{pred}}|$$

Практичний вибір методології у системах прогнозування має залежати передусім від масштабу даних: статистичні моделі залишаються доцільними для малих вибірок, класичні ML-підходи добре працюють у діапазоні до сотень тисяч записів, а глибинні архітектури є найбільш ефективними при роботі з великими та багатовимірними масивами даних.

Загалом можна очікувати, що подальша еволюція систем прогнозування продажів буде рухатися у напрямку більш «інтелектуальних» гібридних рішень,

які поєднують алгоритмічну потужність сучасних AI-технологій з глибоким доменним розумінням бізнес-процесів. Впровадження таких систем має потенціал суттєво підвищити точність прогнозування та одночасно скоротити операційні витрати компаній.

---

### Конфлікт інтересів

---

Автори заявляють, що не мають конфлікту інтересів щодо даного дослідження, включаючи фінансовий, особистий, авторський або будь-який інший, який міг би вплинути на дослідження, а також на результати, представлені в даній статті.

---

### Фінансування

---

Дослідження здійснено в рамках кафедральної науково-дослідної роботи «Моделі і методи системного аналізу в міждисциплінарних дослідженнях» (державний обліковий номер 0125U003246).

---

### Доступність даних

---

Усі дані доступні в цифровій або графічній формі в основному тексті рукопису.

---

### Використання штучного інтелекту

---

Автори підтверджують, що при створенні даної роботи вони не використовували технології штучного інтелекту.

---

### Внесок авторів

---

Бурмей О. С.: концептуалізація, огляд та систематизація літератури, методологія, написання — оригінальний проєкт. Антосяк П. П.: супервізія, методологічні консультації, написання — рецензування та редагування.

---

Авторські права ©



(2026). Бурмей О. С., Антосяк П. П. Ця робота ліцензується відповідно до Creative Commons Attribution 4.0 International License.

---

### Список використаної літератури

1. Rose, K. (2023). Retail Demand Forecasting for 1 Million Products. In *Advances in Intelligent Systems and Computing* (pp. 567–578). Springer. Retrieved from [https://doi.org/10.1007/978-3-031-27440-4\\_45](https://doi.org/10.1007/978-3-031-27440-4_45)
2. Kirichenko, L., Radivilova, T., & Zinkevich, I. (2017). Comparative Analysis of Conversion Series Forecasting in E-commerce Tasks. In *Advances in Intelligent Systems and Computing* (pp. 230–242). Springer, Cham. Retrieved from

- [https://doi.org/10.1007/978-3-319-70581-1\\_16](https://doi.org/10.1007/978-3-319-70581-1_16)
3. Ramos, P., Santos, N., & Rebelo, R. D. (2015). Performance of state space and ARIMA models for consumer retail sales forecasting. *Robotics and Computer-integrated Manufacturing*, 34, 151–163. Retrieved from <https://doi.org/10.1016/J.RCIM.2014.12.015>
  4. Punia, S., Nikolopoulos, K., Singh, S. P., Madaan, J., & Litsiou, K. (2020). Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. *International Journal of Production Research*, 58(16), 4964–4979. Retrieved from <https://doi.org/10.1080/00207543.2020.1735666>
  5. Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32. Retrieved from <https://doi.org/10.1023/A:1010933404324>
  6. Tyrallis, H., & Papacharalampous, G. (2017). Variable Selection in Time Series Forecasting Using Random Forests. *Algorithms*, 10(4), 114. Retrieved from <https://doi.org/10.3390/a10040114>
  7. Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189–1232. Retrieved from <https://doi.org/10.1214/AOS/1013203451>
  8. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). Retrieved from <https://doi.org/10.1145/2939672.2939785>
  9. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154. Retrieved from [https://papers.nips.cc/paper\\_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf](https://papers.nips.cc/paper_files/paper/2017/file/6449f44a102fde848669bdd9eb6b76fa-Paper.pdf)
  10. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, 6638–6648. Retrieved from [https://papers.nips.cc/paper\\_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf](https://papers.nips.cc/paper_files/paper/2018/file/14491b756b3a51daac41c24863285549-Paper.pdf)
  11. Zhang, Y., Song, Y., & Wei, G. (2024). Spatial and temporal attention-based and residual-driven long short-term memory networks with implicit features. *Engineering Applications of Artificial Intelligence*, 133, 108549. Retrieved from <https://doi.org/10.1016/j.engappai.2024.108549>
  12. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780. Retrieved from <https://doi.org/10.1162/neco.1997.9.8.1735>
  13. Nguyen, D. Q., Phan, M. N., & Zelinka, I. (2021). Periodic Time Series Forecasting with Bidirectional Long Short-Term Memory. In *International Conference on Machine Learning* (pp. 234–245). Retrieved from <https://doi.org/10.1145/3453800.3453812>
  14. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. Retrieved from <https://doi.org/10.48550/arXiv.1706.03762>
  15. Lim, B., Arik, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, 37(4), 1748–1764. Retrieved from <https://doi.org/10.1016/j.ijforecast.2021.03.012>
  16. Li, A., Li, Y., Xu, Y., Li, X., & Zhang, C. (2024). Multi-scale convolution enhanced transformer for multivariate long-term time series forecasting. *Neural Networks*, 175, 106745. Retrieved from <https://doi.org/10.1016/j.neunet.2024.106745>
  17. Zhou, H., Zhang, S., Peng, J., Zhang, S., Li, J., Xiong, H., & Zhang, W. (2021). Informer: Beyond efficient transformer for long sequence time-series forecasting. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 35, No. 12, pp. 11106–11115). Retrieved from <https://doi.org/10.1609/aaai.v35i12.17325>
  18. Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364. Retrieved from <https://doi.org/10.1016/j.ijforecast.2021.11.013>
  19. In, Y., & Jung, J. Y. (2022). Simple averaging of direct and recursive forecasts via partial pooling using machine learning. *International Journal of Forecasting*, 38(4), 1386–1399. Retrieved from <https://doi.org/10.1016/j.ijforecast.2021.11.007>

20. Demand Forecasting Using Machine Learning and Deep Learning Approaches in the Retail Industry: A Comparative Study. (2023). In *Lecture Notes in Management and Industrial Engineering* (pp. 257–270). Springer. Retrieved from [https://doi.org/10.1007/978-3-031-25847-3\\_24](https://doi.org/10.1007/978-3-031-25847-3_24)

**Burmei O. S., Antosiak P. P.** Modern approaches to sales forecasting in e-commerce systems.

This article presents a comprehensive analysis of modern sales forecasting methods in e-commerce systems, covering the evolution from traditional statistical approaches to advanced machine learning and deep learning technologies. A detailed comparative analysis of the effectiveness of classical methods (ARIMA, SARIMA, exponential smoothing), machine learning algorithms (random forest, gradient boosting, support vector machines), and deep learning architectures (LSTM, GRU, Transformers, bidirectional LSTM) is conducted.

Based on the analysis of highly cited research and official M5 Competition results, key performance metrics, practical recommendations for method selection, and main industry challenges are identified. Special attention is given to cold start problems, seasonality, external factor influences, and scalability solutions for industrial systems with millions of product positions. Research results show that hybrid approaches and ensemble methods demonstrate the best efficiency in real e-commerce conditions.

**Keywords:** sales forecasting, e-commerce, machine learning, deep learning, time series, LSTM, Transformers, Random Forest.

Отримано: 16.10.2025

Прийнято: 27.11.2025

Опубліковано: 29.01.2026