

<https://doi.org/10.31891/2307-5732-2026-365-91>  
УДК 004.89:658.8

СВИЦ ДМИТРО

Національний університет “Львівська політехніка”

<https://orcid.org/0009-0007-0206-3266>

e-mail: dmytro.svyshch.mnsam.2024@lpnu.ua

БАСЮК ТАРАС

Національний університет “Львівська політехніка”

<https://orcid.org/0000-0003-0813-0785>

e-mail: Taras.M.Basyuk@lpnu.ua

## ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ ПРОДАЖІВ НА ОСНОВІ АНСАМБЛЕВИХ МОДЕЛЕЙ МАШИННОГО НАВЧАННЯ ТА СЕМАНТИЧНОЇ ІНТЕГРАЦІЇ ДАНИХ CRM І ERP

У статті розглянуто вирішення проблеми підвищення точності прогнозування продажів в умовах ринкової невизначеності шляхом семантичної інтеграції даних CRM та ERP систем. Метою дослідження є розробка архітектури та реалізація інтелектуальної інформаційної системи із застосуванням гібридних ансамблевих моделей машинного навчання. В умовах високої волатильності ринку та ізольованості інформаційних систем підприємства традиційні методи прогнозування попиту, що базуються виключно на транзакційній історії, демонструють низьку ефективність, особливо для специфічних товарів «довгого хвоста». Побудова прогнозів без урахування поведінкових індикаторів клієнтів та складських обмежень призводить до розриву між комерційними планами та виробничими можливостями. Для моделювання бізнес-процесів розробленої системи застосовано методологію IDEF0, що дозволило деталізувати етапи збору та інженерії ознак. Програмну архітектуру побудовано за принципом Data Lakehouse із впровадженням модуля збереження ознак. В основу прогностичного ядра покладено гібридне зважене ансамблювання моделей машинного навчання: ARIMA для стабілізації часових трендів, XGBoost для обробки нелінійних сигналів з CRM та LSTM для виявлення довгострокових патернів. Навчання та валідація проводилися на масиві даних понад 4 тисяч активних SKU за період 18 місяців. Наукову новизну становить запропонований механізм динамічного розподілу ваг між базовими моделями ансамблю обернено пропорційно до їхньої локальної помилки на ковзному валідаційному вікні. Доведено, що використання ансамблевої моделі перевершує класичні підходи, знижуючи симетричну середню абсолютну помилку (sMAPE) до 12,9 %. Експериментально підтверджено, що застосування складних моделей машинного навчання доцільне за наявності історичних даних глибиною від 12 до 18 місяців. Також встановлено критичну залежність точності прогнозу від частоти оновлення вхідних ознак: затримка обробки даних з CRM більше ніж на 7 днів призводить до стрімкого зростання помилки. Практичне впровадження результатів дослідження мінімізує вплив людського фактору в процесах планування продажів та операцій (S&OP), а також дозволяє скоротити випадки відсутності товару на складі (out-of-stock) на 9 %.

**Ключові слова:** прогнозування продажів, машинне навчання, семантична інтеграція, CRM-системи, ERP-системи, ансамблеві моделі.

SVYSHCH DMYTRO, BASYUK TARAS

Lviv Polytechnic National University

## INTELLIGENT SALES FORECASTING SYSTEM BASED ON ENSEMBLE MACHINE LEARNING MODELS AND SEMANTIC INTEGRATION OF CRM AND ERP DATA

This paper addresses the challenge of enhancing sales forecasting accuracy amidst market uncertainty through the semantic integration of CRM and ERP system data. The primary objective of this study is to design the architecture for and implement an intelligent information system utilizing hybrid ensemble machine learning models. Given high market volatility and the frequent siloing of enterprise information systems, traditional demand forecasting methods—relying solely on transactional history—exhibit poor performance, particularly for niche “long-tail” products. Generating forecasts without factoring in customer behavioral indicators and inventory constraints creates a disconnect between commercial planning and production capacity. The IDEF0 methodology was employed to model the business processes of the proposed system, facilitating a detailed breakdown of the data collection and feature engineering stages. The software architecture is based on the Data Lakehouse paradigm, incorporating a dedicated Feature Store module. The predictive core relies on a hybrid weighted ensemble of machine learning models: ARIMA is used to stabilize temporal trends, XGBoost processes non-linear signals originating from the CRM, and LSTM detects long-term patterns. Model training and validation were executed using a dataset comprising over 4,000 active SKUs spanning an 18-month period. The scientific novelty lies in the proposed mechanism for the dynamic allocation of weights among the base ensemble models, which are distributed inversely proportional to their local error within a sliding validation window. Results demonstrate that the ensemble model outperforms classical approaches, successfully reducing the symmetric Mean Absolute Percentage Error (sMAPE) to 12.9%. Experimental findings confirm that deploying complex machine learning models is optimal when historical data spanning 12 to 18 months is available. Furthermore, the study identifies a critical correlation between forecasting accuracy and the update frequency of input features: delaying CRM data processing by more than seven days results in a sharp escalation of the error rate. The practical implementation of these research findings minimizes human factor interference in Sales and Operations Planning (S&OP) processes and reduces out-of-stock occurrences by 9%.

**Keywords:** sales forecasting, machine learning, semantic integration, CRM systems, ERP systems, ensemble models

Стаття надійшла до редакції / Received 17.03.2026  
Прийнята до друку / Accepted 11.04.2026  
Опубліковано / Published 28.05.2026



This is an Open Access article distributed under the terms of the [Creative Commons CC-BY 4.0](https://creativecommons.org/licenses/by/4.0/)

© Свиц Дмитро, Басюк Тарас

### Вступ. Загальна постановка проблеми

Відомі методи прогнозування попиту, такі як часові моделі (ARIMA, ETS) або інтуїтивні експертні оцінки, мають свої обмеження, оскільки вони здебільшого базуються на історичних транзакційних даних і не враховують поведінкові індикатори майбутнього попиту [1]. В умовах високої волатильності ринків, частого проведення промоакцій та ускладнення ланцюгів постачання, ізолюваність даних у системах CRM та ERP призводить до того, що прогнози будуються без належного контексту, що знижує їх точність та корисність. З огляду на це, виникає необхідність впровадження сучасних технологічних рішень, що здатні об'єднати розрізнені потоки даних. Одним із найперспективніших підходів є створення інтелектуальних систем із семантичною інтеграцією, що поєднують ранні сигнали продажів із виробничими обмеженнями для побудови ансамблевих моделей [2]. Застосування технологій машинного навчання (XGBoost, LSTM) та наскрізної інтеграції даних у прогнозуванні дозволяє значно підвищити ефективність існуючих процесів планування [3]. Використання збагачених даних з CRM (стадії угод, активність лідів) у поєднанні з операційними даними ERP-систем забезпечує гнучкість моделювання, мінімізуючи ризики відсутності товару на складі та оптимізуючи виробничі графіки. Це сприяє узгодженню планів продажів і закупівель, а також кращому управлінню запасами товарів «довгого хвоста» [4]. Проведене дослідження спрямоване на розробку та впровадження архітектурних рішень для автоматизованого збору та очищення даних, а також на перетворення прогнозованої аналітики в інструмент стратегічного управління підприємством.

### Зв'язок висвітленої проблеми із важливими науковими та практичними завданнями

Існуючі методи прогнозування продажів, що базуються виключно на історичних транзакційних даних, демонструють обмежену ефективність в умовах високої волатильності ринку та ізолюваності інформаційних систем підприємства. Використання підходів машинного навчання та семантичної інтеграції даних CRM та ERP є перспективним рішенням, яке забезпечує врахування як ранніх поведінкових сигналів клієнтів, так і реальних операційних обмежень ланцюгів постачання. Такий підхід дозволяє вирішити проблему розриву між комерційними планами та виробничими можливостями, мінімізуючи помилки прогнозування [5]. З огляду на це, важливим науково-практичним завданням є розробка інтелектуальної системи, що забезпечить наскрізну інтеграцію різномірних даних для вдосконалення процесів операційного планування [3].

Перспективність дослідження полягає в аналізі можливостей гібридних ансамблевих моделей (поєднання ARIMA, XGBoost, LSTM) для створення адаптивних інформаційних систем прогнозування. Зокрема, науково-практичну цінність становить запропонований механізм динамічного розподілу ваг між базовими моделями ансамблю залежно від їхньої локальної точності на специфічних сегментах товарів (SKU), що дозволяє ефективно обробляти та інтерпретувати великі масиви поведінкових та транзакційних даних [6].

Розв'язання цього завдання сприятиме розробці механізмів впровадження складних прогностичних алгоритмів у реальні бізнес-процеси. На практиці це дозволить зменшити випадки відсутності товару на складі, оптимізувати управління запасами товарів «довгого хвоста» та підвищити загальну економічну ефективність управління підприємством.

### Аналіз останніх досліджень та публікацій

Неточність прогнозування попиту завдає значних збитків як глобальним ланцюгам постачання, так і окремим ритейлерам. Неefективне управління складом, що проявляється у затоваренні або, навпаки, нестачі продукції (out-of-stock), щорічно коштує підприємствам мільярди доларів. Згідно з дослідженнями [7], помилка прогнозування лише на 1 % може призводити до втрати 2–4 % прибутку через порушення оборотності складських запасів. В умовах волатильності ринків, традиційні методи планування перестають бути ефективними.

Статистика вказує, що найбільші розриви у плануванні виникають на стику комерційних відділів та операційних департаментів, де дані часто ізолювані [4]. Водночас найбільше від розсинхронізації даних страждають компанії з широким асортиментом («довгий хвіст»), де попит на специфічні товари є важкопрогнозованим [8]. Додатково проаналізувавши методи, що застосовуються для прогнозування, можна зауважити, що більшість підприємств досі покладаються на експертну думку, ігноруючи потенціал семантичної інтеграції даних.

Дослідженню у сфері прогнозування часових рядів та інтеграції бізнес-систем присвячено праці багатьох науковців. Серед них варто виділити праці R. Hyndman та Y. Khandakar [9], які формалізували автоматичний підбір параметрів для класичних моделей ARIMA/ETS, що залишаються надійним бенчмарком. Значний внесок у розвиток методів градієнтного бустингу зробили T. Chen та C. Guestrin [10], чия розробка XGBoost дозволила ефективно обробляти табличні дані з нелінійними залежностями. У дослідженнях S. Hochreiter та J. Schmidhuber [11] закладено основи рекурентних мереж LSTM, здатних виявляти довгострокові часові патерни, що є критичним для сезонних товарів. Важливість гібридних підходів підтверджено у роботах S. Makridakis [12] на основі результатів змагань M4, де ансамблеві методи продемонстрували беззаперечну перевагу. Зазвичай для прогнозування продажів використовують традиційні статистичні методи: ковзне середнє, експоненційне згладжування або ARIMA [13]. Проте ці підходи мають суттєві недоліки: вони є одновимірними, тобто базуються виключно на історії минулих продажів. Окрім цього, побудова прогнозів без урахування складських обмежень ERP призводить до планів, які фізично неможливо виконати [14].

При цьому, використання методів машинного навчання дозволяє інтегрувати гетерогенні дані (текст, категорії, числа), покращуючи точність на рівні SKU. Це підтверджується, зокрема, дослідженнями S. Taylor та V. Letham [15], які розробили прогнозні моделі для масштабованого прогнозування з врахуванням свят та сезонності.

### Формулювання цілей статті

Метою дослідження є розробка архітектури та реалізація інтелектуальної інформаційної системи прогнозування продажів на основі семантичної інтеграції різнорідних даних з CRM та ERP систем із застосуванням гібридних ансамблевих моделей машинного навчання. Відповідно до сформульованої мети, у роботі визначено такі ключові завдання: проаналізувати проблематику неузгодженості планів продажів і виробництва в умовах ринкової невизначеності; порівняти ефективність класичних статистичних методів та сучасних алгоритмів штучного інтелекту; виконати системний аналіз об'єкта дослідження та сформулювати вимоги до інженерії ознак на основі поведінкових сигналів; спроектувати архітектуру системи з використанням підходів Data Lakehouse та Feature Store; розробити програмний модуль ансамблювання моделей (ARIMA, XGBoost, LSTM) та верифікувати роботу системи на реальних історичних даних для підтвердження її точності. Результати виконаної роботи вирішують актуальну проблему низької точності прогнозування попиту та виникнення ситуацій дефіциту товару.

### Основні результати досліджень

Застосування методів машинного навчання для прогнозування продажів потребує вибору відповідної алгоритмічної бази, що впливає на точність, інтерпретованість та обчислювальну складність системи. З огляду на проведений аналіз, вибір здійснювався між класичними стохастичними моделями (ARIMA/ETS), алгоритмами градієнтного бустингу (XGBoost) та рекурентними нейронними мережами (LSTM).

Класичні моделі (ARIMA) вирізняються простотою та швидкістю навчання, але не здатні враховувати зовнішні фактори (ціни, промо акції). Градієнтний бустинг (XGBoost) ефективно працює з табличними даними та категоріальними ознаками з CRM, проте має обмеження у виявленні довгих часових залежностей. Нейронні мережі LSTM найкраще працюють з послідовностями, але потребують значних обсягів даних. Результати порівняльного аналізу методів наведені в Табл. 1.

Таблиця 1.

#### Порівняльний аналіз технологій прогнозування

Параметри порівняння	ARIMA (Stochastic)	XGBoost (Boosting)	LSTM (Deep Learning)	Ансамбль (Hybrid)
Робота з сезонністю	Висока (через диференціювання)	Середня (потребує feature engineering)	Висока (вивчає довгі цикли)	Дуже висока
Врахування промо/цін	Низька (тільки як екзогенні змінні)	Дуже висока (головна перевага)	Середня	Висока
Інтерпретованість	Висока (коефіцієнти формули)	Середня (Feature Importance)	Низька ("Black Box")	Середня
Вимоги до даних	Низькі (працює на малих даних)	Середні	Високі (потребує Big Data)	Середні
Швидкість навчання	Дуже швидко	Швидко	Повільно	Помірно
Точність на "довгому хвості"	Низька	Висока	Середня	Найвища

З огляду на визначені критерії, було обрано стратегію ансамблювання, яка дозволяє нівелювати слабкі сторони інших методів. Основою розробленої системи є застосування трьох ключових компонентів: авторегресії, дерев рішень та рекурентних мереж.

Базою є модель ARIMA, яка дозволяє стабілізувати часовий ряд, [9] та описується рівнянням (1):

$$y'_t = c + \phi_1 y'_{t-1} + \dots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} + \epsilon_t \quad (1)$$

де  $y'_t$  - диференційований ряд;  $\phi_i$  - параметри авторегресії, що описують залежність від попередніх значень;  $\theta_j$  - параметри ковзного середнього, які враховують помилки минулих періодів;  $\epsilon_t$  - білий шум.

Однак ARIMA не здатна врахувати складні нелінійні взаємодії між ознаками з CRM (наприклад, вплив ймовірності закриття угоди та обсяг замовлення). Для цього застосовується алгоритм XGBoost [10], який мінімізує регуляризовану цільову функцію (2):

$$\mathcal{L}(\phi) = \sum_i l(\hat{y}_i, y_i) + \sum_k \Omega(f_k) \quad (2)$$

де  $l$  - функція втрат;  $\Omega(f_k)$  - доданок регуляризації для запобігання перенаванчанням:

$$\Omega(f) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (3)$$

де  $T$  - кількість листків у дереві, а  $w$  - ваги у листках. Це дозволяє моделі відбирати лише найбільш значущі сигнали CRM-системи, відсіюючи шум.

Для моделювання довгих послідовностей та пам'яті про минулі події (наприклад, ефект від промо-акції, що відбулася місяць тому) використовується архітектура LSTM (Long Short-Term Memory). На відміну від звичайних нейромереж, LSTM має механізм «воріт», що регулюють потік інформації. Математично процес оновлення стану комірки описується системою рівнянь:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{4}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{5}$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \tag{6}$$

де  $f_t$  – «ворота забування», що вирішують, яку інформацію видалити з пам'яті;  $i_t$  – «вхідні ворота», що визначають нову інформацію для збереження;  $\sigma$  – сигмоїдна функція активації, що повертає значення від 0 до 1. Оновлений стан комірки  $C_t$ , який і є носієм «довгострокової пам'яті» про поведінку клієнтів, обчислюється залежністю:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \tag{7}$$

Така архітектура дозволяє системі «пам'ятати» релевантні CRM-сигнали протягом довгих проміжків часу і «забувати» неактуальні тренди, що забезпечує високу адаптивність прогнозу. Особливість запропонованого рішення полягає у використанні адаптивного механізму зваженого ансамблювання базових моделей. Фінальний прогноз  $\widehat{y}_{ens}$  формується не простим усередненням, а шляхом призначення ваг  $\omega_m$  кожній моделі обернено пропорційно до її помилки (наприклад, sMAPE) на ковзному валідаційному вікні:

$$\omega_m = \frac{sMAPE_m^{-1}}{\sum_{j=1}^3 sMAPE_j^{-1}} \tag{8}$$

де  $m \in \{ARIMA, XGBoost, LSTM\}$ . Тоді результуючий прогноз обчислюється як лінійна комбінація (9):

$$\widehat{y}_{ens} = \omega_1 \widehat{y}_{ARIMA} + \omega_2 \widehat{y}_{XGBoost} + \omega_3 \widehat{y}_{LSTM} \tag{9}$$

Такий архітектурний підхід дозволяє системі автоматично адаптуватися до поведінки конкретного SKU: якщо для товару зі стабільним попитом класична ARIMA демонструє високу точність, її вага зростає; натомість для товарів із сильним впливом промо-акцій (ознак з CRM) система динамічно віддає перевагу XGBoost або LSTM.

*Проектування системи та архітектура*

Для систематизації процесів інтелектуальної системи та визначення взаємозв'язків між різнорідними потоками даних CRM та ERP використано методологію функціонального моделювання IDEF0. Цей підхід дозволяє відокремити безпосередні дані від механізмів їх обробки та керуючих бізнес-правил. На першому етапі проектування розроблено контекстну діаграму (Рис. 1), що визначає межі системи «Прогнозування продажів». На вході отримується інформація від ERP-системи про історичні транзакції та з CRM-системи про поведінкові події (стадії угод, активність лідів). Далі здійснюється їх обробка використовуючи ресурси хмарної платформи та ML-моделей побудованих на основі стеку бібліотек згідно з встановленими бізнес-правилами обробки даних та актуальним планом маркетингу. На виході отримуємо фінальні результати у вигляді прогнозу продажів та плану закупівель.

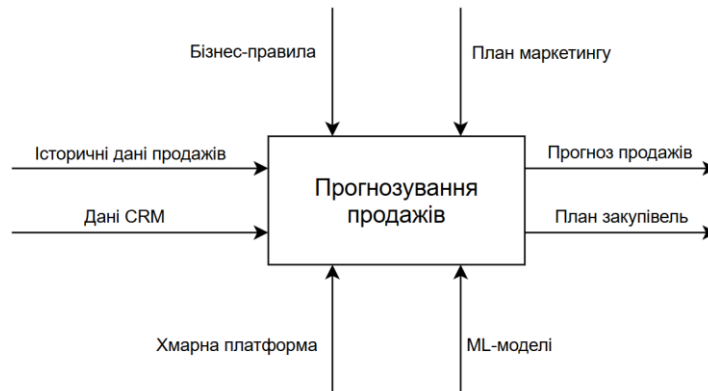


Рис.1. Контекстна діаграма проєктованої системи

Архітектура розробленої системи побудована за сучасним принципом Data Lakehouse, що дозволяє поєднати гнучкість зберігання неструктурованих даних (з CRM та логів) зі строгою типізацією та продуктивністю структурованих даних (з ERP), необхідних для ML-моделювання. Логічна архітектура процесу підготовки даних та прогнозування наведена на рис. 2, де проілюстровано повний цикл від збору необроблених («сирих») сигналів до формування фінального прогнозу.

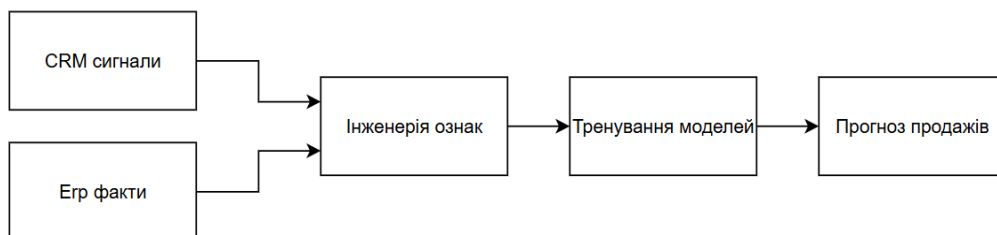


Рис.2. Загальна архітектура системи

Здійснено опис окремих процесів згідно зі схемою на рис. 2. Рівень вхідних даних реалізовано з використанням концепції Feature Store для стандартизованого, відтворюваного та безпечного зберігання ознак. Події надходять до системи через механізми CDC (Change Data Capture) для транзакційних систем та шини подій для потокових даних. Використовуючи сигнали з CRM було враховано безпеку даних за допомогою автоматичного маскуванню персональних даних клієнтів. До ML-пайплайнів потрапляють лише деперсоніфіковані ID та агреговані метрики. Виділено три типи сигналів: *профілі клієнтів* – деперсоніфіковані атрибути; *воронки угод* – дані про етапи переговорів; *журнали активності* – історія звернень, дзвінків, електронних листів. Оскільки структуровані дані про операційну діяльність можуть дублюватися або агрегуватися між CRM та ERP використовуються лише передача фактів з ERP-системи з яких ключовими є *рівні запасів* – поточні залишки на складах та *історія замовлень* – історичні дані про закупівлі та відвантаження.

Вхідні вектори даних надходять у центральний вузол – *розробку ознак та семантичну інтеграцію*. Саме тут відбувається реалізація конвеєрів машинного навчання (ML-пайплайнів) мовою Python. Для реалізації цього вузла використовуються бібліотеки Statsmodels (для базового статистичного аналізу), XGBoost та TensorFlow (для ML-моделювання). Основні етапи обробки: *семантична інтеграція* (виконується узгодження значень, зокрема етапи воронки з CRM та ERP зіставляються для створення єдиної метрики ймовірності покупки) та *інженерія ознак*. Здійснюється перетворення вхідних даних у стандартизований набір ознак.

Результатом роботи вузла обробки є набір фінальних ознак, які потрапляють до таблиці розрахованих ознак. Зазначений датасет є джерелом для сервісу прогнозів, який розгорнуто як високопродуктивний REST API. ERP-система інтегрована з цим API, що дозволяє їй в автоматичному режимі запитувати прогнози на рівні «SKU × клієнт» та автоматично створювати заявки на закупівлю на основі отриманих результатів.

#### Результати дослідження та експериментальна перевірка

Підсумкова частина роботи була присвячена оцінці результативності розробленої архітектури на вибірці даних за 18 місяців, що охоплювала понад 4 тис. активних SKU. Основними критеріями для оцінювання ефективності було обрано метрику sMAPE (симетрична середня абсолютна помилка у відсотках) та показник скорочення випадків відсутності товару. Результати порівняння методів наведені в табл. 2.

Таблиця 2

#### Порівняння методів прогнозування на рівні SKU×клієнт

Метод	Ознаки	sMAPE, %
ARIMA	Лаги, сезонність	18.7
Prophet	Календар, свята	17.9
XGBoost	+ CRM/ERP	14.2
LSTM	+ екзогенні регресори	13.8
Ансамбль	Комбінація	12.9

Результати підтвердили, що ансамблевий підхід є оптимальним (sMAPE становить 12,9 %), особливо завдяки залученню CRM-сигналів через модель XGBoost (sMAPE 14,2 %). Для глибшого аналізу стабільності архітектури та обґрунтування вибору параметрів системи було проведено два додаткові експерименти. Перший експеримент був спрямований на оцінку впливу глибини історичних даних (CRM та ERP) на точність прогнозів. Аналіз проводився з кроком у 3 місяці (від 3 до 18 місяців історії). Результати подано в табл. 3.

Таблиця 3

#### Динаміка зміни sMAPE залежно від обсягу історичних даних

Глибина історії	ARIMA (sMAPE, %)	Prophet (sMAPE, %)	XGBoost (sMAPE, %)	LSTM (sMAPE, %)	Ансамбль (sMAPE, %)
3 міс.	25.1	24.8	21.0	23.5	20.2
6 міс.	21.3	21.0	17.5	19.8	16.8
9 міс.	20.1	19.9	16.1	17.9	15.4
12 міс.	19.2	18.5	15.1	16.1	14.2
15 міс.	18.9	18.1	14.5	14.9	13.5
18 міс.	18.7	17.9	14.2	13.8	12.9

Для наочності ці результати візуалізовано на графіку (рис. 3).

Як видно з рис. 3, моделі класичного прогнозування (ARIMA, Prophet) швидко виходять на визначену точність після 9–12 місяців історії. Натомість моделі машинного навчання (XGBoost, LSTM та ансамбль) продовжують покращувати свої показники при збільшенні вибірки до 18 місяців, що підтверджує доцільність накопичення великих масивів даних у побудованому Lakehouse. Другий експеримент перевіряв чутливість моделей до частоти оновлення вхідних ознак (частоти перерахунку прогнозів). Оскільки архітектура передбачає розгортання REST API та інтеграцію через шини подій (CDC), це дозволяє системі працювати в режимі щоденного оновлення. Динаміку деградації точності внаслідок збільшення інтервалу оновлення наведено в табл. 4 та на рис. 4.

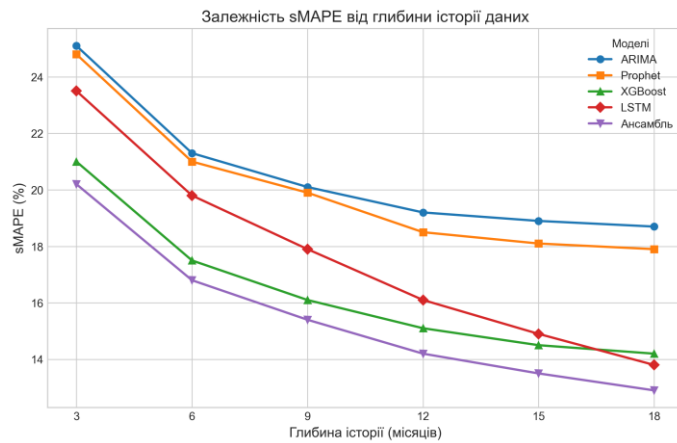


Рис. 3 Залежність точності прогнозування (sMAPE) від глибини історичних даних

Таблиця 4

**Вплив частоти оновлення прогнозів на метрику sMAPE**

Інтервал оновлення	ARIMA (sMAPE, %)	Prophet (sMAPE, %)	XGBoost (sMAPE, %)	LSTM (sMAPE, %)	Ансамбль (sMAPE, %)
1 день	18.7	17.9	14.2	13.8	12.9
3 дні	19.3	18.5	14.9	14.4	13.5
7 днів	20.8	19.9	16.1	15.9	14.8
14 днів	23.1	22.1	18.5	18.1	16.9
30 днів	27.5	26.3	22.8	21.9	20.5

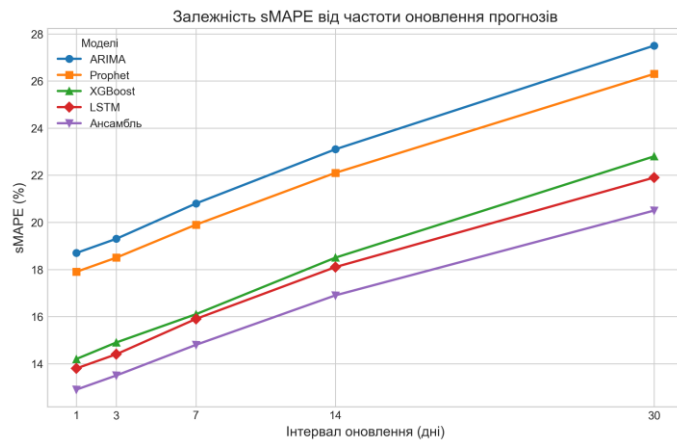


Рис. 4. Динаміка зростання помилки прогнозування залежно від інтервалу оновлення даних

Графік (рис. 4) ілюструє, що використання динамічних ознак (наприклад, актуальних етапів воронки з CRM або поточних залишків з ERP) вимагає високої частоти оновлень. Затримка в оновленні прогнозів навіть на 7 днів призводить до зростання помилки sMAPE ансамблевої моделі з 12,9 % до 14,8 %, а при місячному інтервалі система повністю втрачає свою ефективність (sMAPE > 20 %). Це в значній мірі виправдовує архітектурне рішення щодо використання потокової обробки та REST API замість пакетного перерахунку за розкладом один раз на місяць. Підсумовуючи результати всіх експериментів, можна стверджувати, що економічна доцільність інтеграції CRM та ERP даних прямо залежить від обраної архітектури обробки. Запропонована система на базі Feature Store з щоденним оновленням через REST API та ансамблевою моделлю прогнозування дозволила досягти sMAPE на рівні 12,9 %. Практичним наслідком застосування системи стане скорочення випадків відсутності товару на складі на 9 % у сегменті «довгого хвоста», що суттєво оптимізує процес автоматичного створення заявок на закупівлю.

**Висновки**

У статті вирішено актуальне науково-практичне завдання щодо підвищення точності планування попиту шляхом семантичної інтеграції даних CRM та ERP систем. Основні результати:

1. Проведено аналіз існуючих підходів до прогнозування та обґрунтовано вибір ансамблю моделей машинного навчання (ARIMA, XGBoost, LSTM), як ефективного способу поєднання транзакційної історії та ранніх поведінкових сигналів ринку.

2. Виконано процес створення інформаційної системи з супровідним ґрунтовним аналізом на базі методології функціонального моделювання IDEF0. Це забезпечило прозорий опис етапів акумулювання інформації та логіки обробки ознак.

3. Реалізовано прототип програмної системи, який надає можливість автоматизувати планування попиту з врахуванням поведінкових факторів клієнтів.

4. Показано, що розроблена архітектура відрізняється від існуючих аналогів здатністю враховувати ймовірнісні оцінки угод з CRM-системи. Це дозволило знизити помилку прогнозування (sMAPE) до 12,9 % та оптимізувати управління запасами товарів «довгого хвоста», практично унеможлививши виникнення дефіциту товару через невраховані фактори.

Подальші дослідження будуть зосереджені на оптимізації процесів інтеграції прогнозної аналітики безпосередньо в цикли операційного планування, а також на пошук рішень для повної автоматизації закупівель на основі отриманих прогнозів. Успішна реалізація таких підходів сприятиме підвищенню ефективності та стійкості ланцюгів постачання в умовах ринкової невизначеності.

### Література

1. Borucka A. Seasonal Methods of Demand Forecasting in the Supply Chain as Support for the Company's Sustainable Growth. *Sustainability*. 2023. Vol. 15, No. 9. P. 7399. DOI: <https://doi.org/10.3390/su15097399>.
2. Breiman L. Random forests. *Machine Learning*. 2001. Vol. 45, No. 1. P. 5–32. DOI: <https://doi.org/10.1023/A:1010933404324>.
3. Klaus H., Rosemann M., Gable G. G. What is ERP? *Information Systems Frontiers*. 2000. Vol. 2, No. 2. P. 141–162. DOI: <https://doi.org/10.1023/A:1026543906354>.
4. Payne A., Frow P. A strategic framework for customer relationship management. *Journal of Marketing*. 2005. Vol. 69, No. 4. P. 167–176. DOI: <https://doi.org/10.1509/jmkg.2005.69.4.167>.
5. Pasupuleti V., Thuraka B., Kodete C. S., Malisetty S. Enhancing supply chain agility and sustainability through machine learning: Optimization techniques for logistics and inventory management. *Logistics*. 2024. Vol. 8, No. 3. P. 73. DOI: <https://doi.org/10.3390/logistics8030073>.
6. Dudek G., Peřka P., Smyl S. A hybrid residual dilated LSTM and exponential smoothing model for midterm electric load forecasting. *IEEE Transactions on Neural Networks and Learning Systems*. 2022. DOI: <https://doi.org/10.1109/TNNLS.2020.3046629>.
7. Abolghasemi M., Gerlach R., Tarr G., Beh E. Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion. *arXiv preprint, arXiv:1909.13084*. 2019. DOI: <https://doi.org/10.48550/arXiv.1909.13084>.
8. Arvan M., Fahimnia B., Reisi M., Siemsen E. Integrating human judgement into quantitative forecasting methods: A review. *Omega*. 2018. Vol. 86. P. 237–252. DOI: <https://doi.org/10.1016/j.omega.2018.07.012>.
9. Hyndman R. J., Khandakar Y. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software*. 2008. Vol. 27, No. 3. P. 1–22. DOI: <https://doi.org/10.18637/jss.v027.i03>.
10. Chen T., Guestrin C. XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 2016. P. 785–794. DOI: <https://doi.org/10.1145/2939672.2939785>.
11. Hochreiter S., Schmidhuber J. Long short-term memory. *Neural Computation*. 1997. Vol. 9, No. 8. P. 1735–1780. DOI: <https://doi.org/10.1162/neco.1997.9.8.1735>.
12. Makridakis S., Spiliotis E., Assimakopoulos V. The M4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*. 2020. Vol. 36, No. 1. P. 54–74. DOI: <https://doi.org/10.1016/j.ijforecast.2019.04.014>.
13. Ensafi Y., Amin S. H., Zhang G., Shah B. Time-series forecasting of seasonal items sales using machine learning – A comparative analysis. *International Journal of Information Management Data Insights*. 2022. Vol. 2, No. 1. P. 100058. DOI: <https://doi.org/10.1016/j.ijime.2022.100058>.
14. Davenport T. H. Putting the enterprise into the enterprise system. *Harvard Business Review*. 1998. Vol. 76, No. 4. P. 121–131.
15. Taylor S. J., Letham B. Forecasting at scale. *The American Statistician*. 2018. Vol. 72, No. 1. P. 37–45. DOI: <https://doi.org/10.1080/00031305.2017.1380080>.