

УДК 004.94:004.738.5:658.78

DOI: 10.31673/2412-9070.2026.017401

А. А. БАЛВАК, аспірант;  
ORCID: 0000-0002-6441-8225В. В. ЗІНЧЕНКО, аспірант,  
ORCID: 0009-0000-7087-1678

Державний університет інформаційно-комунікаційних технологій, Київ

## ФОРМУВАННЯ ВИБІРКИ ДАНИХ ТА КРИТЕРІЇ ЯКОСТІ ДЛЯ МОДЕЛЮВАННЯ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ СКЛАДСЬКОЇ ЛОГІСТИКИ

*У сучасному світі стрімкий розвиток сектору електронної комерції створює безпрецедентне навантаження на логістичні системи, вимагаючи від складських комплексів високої адаптивності та швидкості обробки замовлень. Традиційні підходи до управління складом, що базуються на фіксованому закріпленні місць зберігання та класифікації товарів за показником обсягу продажів, виявляються неефективними в умовах високої мінливості попиту та жорстких вимог до термінів доставки. У статті досліджено проблематику підвищення продуктивності складських процесів шляхом переходу від суто кількісного нарощування фізичних потужностей до впровадження інтелектуальних алгоритмів обробки даних.*

*Метою статті є визначення передумов та інструментарію для алгоритмічної оптимізації розміщення товарних позицій (SKU) на основі виявлення прихованих закономірностей сумісного попиту. Спочатку було проведено аналіз сучасних технологічних трендів, який продемонстрував зміщення акцентів від систем «людина до товару» до роботизованих концепцій «товар до людини» (MRFS, AGV), ефективність яких суттєво залежить від якості алгоритмів маршрутизації та призначення місць зберігання.*

*Із зростанням обсягів даних ключовим фактором успіху стає формування якісної інформаційної бази для моделювання. У роботі детально розглянуто критерії релевантності даних для логістичних задач, серед яких виокремлено наявність точних часових міток, транзакційну цілісність та деталізацію фізичних параметрів вантажу. Особливу увагу приділено як перевагам, так і потенційним викликам роботи з відкритими джерелами даних. Проведено системний огляд восьми відібраних датасетів з репозиторіїв Kaggle та UCI Machine Learning Repository, проаналізовано їх придатність для вирішення специфічних задач: від кластеризації часових рядів та прогнозування пікових навантажень до оптимізації пакування.*

*Окремо висвітлено роль внутрішніх корпоративних даних підприємств, зокрема операційних логів систем управління складом, які дозволяють врахувати реальні обмеження інфраструктури та поведінкові фактори персоналу. Зроблено висновок, що цифрова трансформація сучасного логістичного центру являє собою синергію роботизованих технологій та методів машинного навчання. Результати дослідження підтверджують доцільність застосування кластеризації для динамічного перерозподілу SKU, що дозволяє мінімізувати дистанції переміщення та трансформувати логістику з моделі реагування на попит постфактум у систему завчасного планування на основі прогнозів.*

**Ключові слова:** складська логістика; оптимізація розміщення товарів; кластеризація даних; машинне навчання; системи «товар до людини»; системи «людина до товару»; електронна комерція; аналіз великих даних; прогнозування попиту.

### Постановка проблеми

Стрімкий розвиток електронної комерції (e-commerce) створює суттєве навантаження на функціонування логістичних центрів. Сучасні ринкові умови характеризуються значним зрос-

танням кількості дрібних замовлень, жорсткими вимогами до скорочення термінів доставки, зокрема впровадженню послуг доставки на той же або на наступний день та високою мінливістю попиту. За таких обставин традиційні методи управління складом виявляються неефективними. Насамперед це стосується підходів, що спираються на фіксоване закріплення місць зберігання та класифікацію товарів лише за одним показником, найчастіше за обсягом товарообігу. Вони не забезпечують необхідної адаптивності до динаміки запитів, що призводить до інертності процесів та нераціональних маршрутів комплектування.

Основна ідея дослідження полягає в тому, що для підвищення продуктивності складських процесів суто кількісне збільшення фізичних потужностей (стелажів, навантажувачів, площ зберігання) часто виявляється недостатнім або економічно невиправданим. Критичним фактором успіху стає програмна складова – інтелектуальні алгоритми, здатні обробляти великі масиви даних для прийняття рішень у реальному часі. Актуальним науковим завданням є мінімізація відстаней, що долаються при зборі замовлень, шляхом виявлення прихованих патернів сумісного попиту. Для цього доцільно застосовувати методи машинного навчання, зокрема кластеризацію часових рядів для оптимізації розміщення товарних позицій (Stock Keeping Unit, SKU) залежно від ймовірності їх спільного придбання.

Разом з тим, важливою умовою для ефективного впровадження таких інтелектуальних систем є наявність доступних та якісних даних. Джерелами інформації можуть виступати як відкриті репозиторії, зокрема платформи Kaggle та UCI Machine Learning Repository, так і внутрішня історія записів про виконані операції із систем управління складом підприємств. У цьому контексті виникає нагальна потреба у формуванні чітких критеріїв релевантності даних (наявність часових міток, деталізація до рівня SKU, ідентифікатори транзакцій тощо), які дозволять оцінити їх придатність для моделювання та прогнозування особливостей попиту.

### *Аналіз останніх досліджень і публікацій*

Огляд наукових робіт [1, 2] у сфері сучасних складських рішень демонструє виразну тенденцію переходу від традиційної концепції «людина до товару» (picker-to-parts) до більш продуктивної моделі «товар до людини» (parts-to-picker). Водночас попри загальний технологічний прогрес, склади типу picker-to-parts залишаються актуальними завдяки своїй гнучкості та нижчим початковим інвестиціям, особливо для обробки нестандартних вантажів, хоча такі системи й потребують значних витрат часу на переміщення комплектувальників [1].

У спектрі сучасних технологічних рішень, що трансформують складську логістику, особливу увагу привертають комплекси переміщення стелажів, відомі як мобільні роботизовані системи виконання замовлень (Mobile Robot Fulfillment Systems, MRFS). Найбільш показовим прикладом цієї категорії є автономні платформи KIVA, які реалізують концепцію «товар до людини» шляхом автоматизованого підіймання та транспортування цілих стелажів безпосередньо до станцій комплектування (рис. 1) [1]. Такий підхід дозволяє нівелювати непродуктивні витрати часу персоналу на переміщення територією складу. Водночас дослідження підтверджують, що операційна ефективність цих систем суттєвою мірою залежить від якості застосованих алгоритмів призначення місць зберігання та планування маршрутів руху роботів.

Поряд із рішеннями переміщення стелажів, значного поширення набуває технологія AGV-assisted picking, що базується на використанні автоматизованих керованих візків [1, 3]. Ці мобільні платформи інтегруються у процес відбору, функціонуючи у режимі супроводу комплектувальника або виконуючи автономне транспортування зібраних замовлень до зони відвантаження, що дозволяє суттєво знизити фізичне навантаження на персонал та прискорити внутрішньоскладські логістичні потоки.

Окремий вектор автоматизації, висвітлений у фаховій літературі, стосується впровадження роботизованого поштучного відбору та мобільних роботів. Зокрема, роботизовані маніпулятори (robotic arms) демонструють високу ефективність при обробці товарів різної конфігурації, тоді як автономні мобільні роботи забезпечують оптимізацію переміщення одиниць зберігання у динамічному середовищі складу [3].

Для мінімізації впливу людського фактору та підвищення точності операцій на етапі комплектування застосовуються високотехнологічні робочі станції. Їх оснащення системами світлової індикації pick-to-light або put-to-light дозволяє візуалізувати необхідні комірки для відбору або розміщення товару, що критично важливо для зменшення кількості помилок [1].

Враховуючи специфіку електронної комерції, яка характеризується замовленнями, що містять декілька товарних позицій, та значним відсотком повернень, критичного значення набувають комплекси автоматизованого сортування (bag sorter systems). Технології підвісних транспортерів із кишнями дозволяють раціонально консолідувати відправлення та забезпечувати оперативність їх обробки [1].

Спектр сучасних рішень доповнюють технології високої щільності зберігання. Зокрема, активного розвитку набули човникові комплекси (Shuttle-based Storage and Retrieval Systems), висока продуктивність яких підтверджена імітаційним моделюванням [4, 5]. Також значний інтерес представляють платформи блокового зберігання [2]. У системах цього типу роботи маневрують по поверхні щільної сітки контейнерів, завдяки чому максимізується використання складського об'єму.

Узагальнюючи огляд технологій, слід зазначити, що для автоматизованих комплексів класу «товар до людини» мінімізація відстаней переміщення залишається ключовим фактором продуктивності. Результати досліджень [4] підтверджують, що параметри маршрутизації та швидкості руху транспортерів безпосередньо впливають на енергоспоживання та пропускну здатність обладнання. Також зазначається, що загальна результативність роботизованих рішень (MRFS) визначається взаємоузгодженою оптимізацією призначення місць зберігання та планування шляху [6]. Відтак, навіть найсучасніше апаратне забезпечення втрачає свій потенціал при хаотичному або нераціональному розміщенню запасів, особливо коли сумісні товари рознесені по різних зонах складу. Зазначені фактори обумовлюють важливість вирішення наукового завдання алгоритмічної оптимізації розміщення товарів на основі інтелектуального аналізу даних про закономірності сумісного попиту.



Рис. 1. Приклад роботизованої системи KIVA, що переміщує стелажі

### *Основна частина*

У контексті цифрової трансформації логістичних процесів дані про транзакції трансформувалися з пасивного елемента обліку у критично важливий актив, що визначає операційну ефективність підприємства. Зазначені інформаційні масиви виступають необхідним підґрунтям для сучасних алгоритмів оптимізації. Без наявності якісних, деталізованих історичних даних неможливо побудувати адекватну модель попиту, здатну адаптуватися до стохастичної природи ринку та забезпечити ефективне функціонування інтелектуальних систем управління складом. Це зумовлено тим, що результативність впровадження методів машинного навчання, таких як кластеризація часових рядів або пошук асоціативних правил, прямо залежить від релевантності вхідного інформаційного потоку. Відтак, формування репрезентативної вибірки даних є першочерговим завданням, що передусє етапу моделювання.

Для проведення верифікованих наукових досліджень та тестування розроблених алгоритмів доцільно використовувати відкриті набори даних, розміщені на спеціалізованих платформах. Провідними ресурсами для отримання емпіричних даних є репозиторії Kaggle та UCI Machine Learning Repository, на яких надається доступ до реальних транзакційних логів глобальних ритейлерів та логістичних операторів.

Зазначені джерела дозволяють отримати масиви даних, що відповідають специфіці різних типів складських систем. Якщо для концепцій «людина до товару» критично важливими є параметри маршрутизації, фізичні виміри та структура замовлень, необхідні для моделювання групового відбору та оптимізації шляху комплектувальника, то для роботизованих комплексів «товар до людини» акцент зміщується на аналіз сумісності товарів. Останній вимагає наявності даних для кластеризації SKU за ймовірністю їх спільного попиту, що забезпечує раціональне розміщення сумісних позицій на мобільних стелажах для мінімізації кількості переміщень роботів. Саме тому у межах даного дослідження пошук наборів даних здійснювався цілеспрямовано для виявлення товарів, що замовляються разом. Така стратегія відбору ґрунтується на положенні, що ідентифікація прихованих патернів сумісного попиту виступає ключовим елементом реалізації гібридних стратегій зберігання. Відібрані датасети повинні підходити саме для цієї цілі, містячи історію спільних покупок, що дозволяє застосовувати алгоритми для підвищення щільності зберігання та скорочення часу комплектування.

### ***Критерії вибору та оцінки якості даних***

Вибір релевантних наборів даних базується на чіткій системі критеріїв, що гарантують придатність інформації для подальшого моделювання. Спираючись на методологічні підходи, викладені у наукових публікаціях [7, 8], було сформовано вимоги до якості даних. Згідно з цими дослідженнями великі дані вимагають перевірки не лише на формальну правильність, а й на змістовну відповідність для конкретних бізнес-задач. Серед ключових критеріїв релевантності насамперед виокремлюється наявність часових міток з обов'язковою точністю до хвилин або секунд, що дозволяє аналізувати внутрішньодобову сезонність та пікові навантаження. По-друге, критичною умовою для проведення аналізу споживчого кошика є деталізація даних до рівня унікальних ідентифікаторів товару та транзакції. Крім того, важливе значення мають кількісні та вартісні показники, зокрема поля з інформацією про кількість та ціну, які необхідні для класифікації об'єктів зберігання за важливістю шляхом ранжування за обсягом продажів та розрахунку швидкості обігу запасів. Нарешті, наявність поведінкових та географічних атрибутів, таких як ідентифікатор клієнта або геодані, дозволяє виявляти стабільні закономірності поведінки та оптимізувати розміщення товарів залежно від регіональних особливостей попиту.

Узагальнені вимоги до структури та змісту даних, необхідних для моделювання логістичних процесів, наведено у таблиці.

### **Критерії оцінки якості наборів даних для моделювання логістичних операцій**

<b>Категорія критерію</b>	<b>Типовий ідентифікатор поля</b>	<b>Опис та обґрунтування важливості для складу</b>
Часова дискретність	Timestamp Precision	Наявність точної дати та часу. Критично для виявлення сезонності спільних покупок та динаміки попиту в часі
Транзакційна цілісність	Invoice / Order ID	Унікальний ідентифікатор замовлення. Необхідний для аналізу споживчого кошика з метою виявлення сумісності товарів у транзакції
Ідентифікація об'єкта	SKU / StockCode	Унікальний код товарної одиниці. Дозволяє відстежувати рух конкретного фізичного об'єкта, а не товарної групи, що є основою для точного адресного зберігання
Кількісні метрики	Quantity, Price	Обсяг замовлення за товарною позицією. Визначає частоту поповнення запасів та дозволяє проводити ранжування асортименту за пріоритетністю

Фізичні параметри	Dimensions, Weight	Габарити та вага. Параметр, необхідний для перевірки обмежень місткості комірок та моделювання фізичного розміщення на стелажах
Поведінковий вимір	User / Session ID	Дозволяє аналізувати стабільність попиту в розрізі конкретних клієнтів та групувати товари за моделями придбання
Геопросторові дані	Location / ZIP	Слугують основою для моделювання розподілу замовлень за регіонами, що впливає на стратегію відвантаження та консолідацію вантажів

### Характеристики наборів даних

На основі вищенаведених критеріїв було відібрано вісім найбільш релевантних наборів даних, які дозволяють комплексно дослідити проблему оптимізації розміщення товарів на складі. Кожен з датасетів має специфічні характеристики, що роблять їх придатними для тестування алгоритмів кластеризації попиту та оптимізації маршрутів відбору.

**E-Commerce Data** є класичним транзакційним набором даних, який містить понад 500 тисяч записів про операції британського онлайн-ритейлера, що спеціалізується на подарунках, здійснені протягом 2010–2011 років. Цей масив, доступний за посиланням <https://www.kaggle.com/datasets/carrie1/ecommerce-data>, повністю відповідає визначеним критеріям. Зокрема, наявність полів *InvoiceNo* та *StockCode* дозволяє точно ідентифікувати товари, що купуються разом, для вирішення задачі сумісного розміщення, тоді як *InvoiceDate* слугує основою для побудови часових рядів. Водночас важливою особливістю набору є інформація про скасування замовлень, що дає змогу моделювати процеси обробки повернень товарів від покупців.

**eCommerce Events History in Electronics** – це масив даних, доступний за покликанням <https://www.kaggle.com/datasets/mkechinov/ecommerce-events-history-in-electronics-store>, який містить повну хронологію поведінки користувачів на сайті магазину електроніки. На відміну від попереднього, цей набір дозволяє аналізувати не лише покупки, а й переглядання та додавання товарів у кошик. Завдяки високій точності часових міток (*event\_time*) та поля *user\_session*, стає можливим групування товарів за наміром користувача, а також прогнозування пікових навантажень ще до оформлення замовлення, що є важливим для завчасного поповнення зон відбору.

Наступним важливим джерелом є репрезентативний набір даних **Online Retail II UCI**, що охоплює транзакції за два роки. Цей масив, доступний за адресою <https://www.kaggle.com/datasets/mashlyn/online-retail-ii-uci>, містить ключові поля *InvoiceDate*, *StockCode* та *Quantity*. Завдяки значному часовому проміжку, з'являється можливість не лише виявляти сезонність, а й досліджувати довгострокові тренди. Це є необхідною умовою для перевірки стабільності кластерів товарів у часі, забезпечуючи надійність моделей прогнозування попиту у динамічному середовищі.

Для вирішення задач ієрархічного прогнозування доцільно використати датасет **M5 Forecasting - Accuracy**, який розміщено за посиланням <https://www.kaggle.com/c/m5-forecasting-accuracy/data>. Цей масив містить історію продажів понад 30 тисяч товарів за 5 років, доповнену календарем подій та цінами. Структура даних робить їх ідеальними для кластеризації часових рядів із застосуванням алгоритму динамічного трансформування часу, що дозволяє групувати товари за формою кривої попиту незалежно від часового зсуву. Такий підхід сприяє оптимізації логістики шляхом рівномірного розподілу товарів з піковим навантаженням по різних зонах складу для уникнення заторів.

Для дослідження специфіки продуктового ритейлу варто звернути увагу на масив даних **Instacart Market Basket Analysis**, який налічує понад 3 мільйони замовлень та доступний на

сторінці <https://www.kaggle.com/datasets/psparks/instacart-market-basket-analysis>. Його унікальний атрибут *add\_to\_cart\_order* відображає послідовність додавання товарів у кошик, що дозволяє відтворити природну логіку покупця. Така аналітика є визначальною для оптимізації алгоритмів побудови маршрутів комплектування, дозволяючи планувати траєкторію руху таким чином, щоб важкі та габаритні позиції відбиралися першочергово. Крім того, аналіз даних у полі *reordered* допомагає виділити ядро асортименту для організації виділених зон прискореного відбору.

Найбільш орієнтованим на логістичні завдання є набір даних **Brazilian E-Commerce Public Dataset by Olist**, який складається з 9 взаємопов'язаних таблиць і доступний за посиланням <https://www.kaggle.com/datasets/olistbr/brazilian-ecommerce>. Його суттєвою перевагою є наявність у таблиці *olist\_products\_dataset* фізичних параметрів товарів: ваги (*product\_weight\_g*), довжини (*product\_length\_cm*), висоти (*product\_height\_cm*) та ширини (*product\_width\_cm*). Ця інформація дозволяє вирішувати задачу пакування та підбирати оптимальну тару ще до початку комплектування. Окрім того, наявність координат продавців та покупців дає змогу використовувати дані масиву для вирішення завдань геопросторового аналізу та оптимізації розміщення нових фулфілмент-центрів.

Ще одним цінним джерелом для аналізу є датасет **Favorita Grocery Sales Forecasting**, що містить інформацію про продажі великої еквадорської мережі супермаркетів. Цей масив, доступний за посиланням <https://www.kaggle.com/c/favorita-grocery-sales-forecasting/data>, вирізняється наявністю макроекономічних показників, зокрема цін на нафту, які суттєво впливають на економіку регіону. Особливістю даних є поле *onpromotion*, в якому вказано чи пропонувався товар у рамках спеціальної пропозиції у конкретному магазині на певну дату. Такий рівень деталізації дозволяє моделювати динамічну адаптацію складського простору під час промоакцій. Аналіз сплесків продажів дає змогу розраховувати коефіцієнти розширення зон відбору, забезпечуючи автоматичне резервування комірок для акційних товарів та запобігаючи переповненню складських площ.

Для вирішення специфічних задач логістики індустрії моди доцільно використовувати набір даних **H&M Personalized Fashion Recommendations**, доступ до якого відкрито за адресою <https://www.kaggle.com/c/h-and-m-personalized-fashion-recommendations/data>. Даний масив містить транзакції та метадані товарів, а головне – зображення, що дозволяє застосовувати методи комп'ютерного зору. Використання векторних представлень зображень уможливило впровадження стратегії візуальної диференціації товарів, оскільки розміщення схожих зовні позицій на відстані одна від одної суттєво зменшує ризик помилок при комплектуванні. Водночас дані дозволяють групувати товари за принципом колекцій, адаптуючи складські процеси до сезонних змін асортименту швидкої моди.

Використання зазначених наборів даних забезпечує всебічне охоплення проблематики оптимізації складських процесів, дозволяючи тестувати алгоритмічні моделі як на рівні аналізу попиту, так і на рівні фізичних обмежень складського простору.

### **Використання внутрішніх даних підприємств**

Поряд із відкритими наборами даних не менш важливим джерелом для побудови адаптивних складських систем є корпоративні дані підприємств, що генеруються у системах управління складом та планування ресурсів. Специфіка збору внутрішніх даних полягає у необхідності фіксації не лише комерційних фактів продажу, а й операційних логістичних параметрів. До таких показників належать точні часові мітки початку та завершення операцій відбору, маршрути руху комплектувальників та дані про завантаженість техніки.

Огляд актуальних наукових праць виявляє різноманіття стратегій використання даних відповідно до завдань оптимізації. Показовим прикладом використання деталізованих операційних логів є робота [9], в якій для оптимізації процесів пакування замовлень було використано дані системи голосового відбору великого роздрібного складу у Фінляндії. Дослідники проаналізували записи майже 35 000 сформованих маршрутів збору, що містили близько 100 тис. клієнтських замовлень, накопичених протягом трьох місяців. Важливим аспектом цього

масиву даних була наявність детальної інформації про кожну операцію: адресу комірки та точний час виконання відбору, номер пакету, ідентифікатор замовлення й виконавця. Такий рівень деталізації дозволив авторам не лише відтворити реальні маршрути, але й перевірити ефективність запропонованого алгоритму імітаційного відпалу на реальних історичних даних, досягнувши скорочення дистанції переміщення на 15,7 %.

Інший підхід до формування набору даних для тренування інтелектуальних систем продемонстровано у дослідженні [10]. Для підготовки агента глибокого навчання з підкріпленням автори використали історичні дані про операції розміщення та вилучення товарів, зібрані протягом 14 місяців на складі з частковою автоматизацією. База даних містила 12 100 записів, кожен з яких включав інформацію про дату та час призначення місця зберігання, ідентифікатор типу товару, кількість одиниць, дату першого пакування та тип операції (первинне надходження або повернення). Цей масив було розділено на тренувальний (12 місяців) та тестовий (2 місяці) набори, що дозволило моделі вивчити сезонність та частоту звернень до SKU без участі експертів. Згідно з отриманими результатами розроблена модель глибокого навчання з підкріпленням дозволила знизити транспортні витрати на представленому у дослідженні складі на 6,3 % порівняно з методом, що використовується.

Важливість інтеграції фізичних параметрів товарів із даними про попит підкреслюється у роботі [11]. Для розробки моделі розміщення продукції автори використали історичні дані колумбійської компанії з виробництва продуктів харчування. Вхідні дані включали параметри конструкції складу (розміри стелажів, ширина проходів) та характеристики підйомно-транспортного обладнання. Також використовувалася інформація про попит на SKU за певні періоди, що дозволило врахувати ступінь однорідності замовлень. Такий підхід уможливив класифікацію товарів за методами Парето-аналізу та показником відношення об'єму до частоти замовлень, базуючись на реальній історії відвантажень. Застосування розробленої моделі дозволило також змодельовати сценарії розміщення, що мінімізують загальний час складських операцій.

Водночас існують випадки, коли пряме використання первинних даних є ускладненим через недосконалість наявних систем управління складом. У дослідженні [12] розглянуто приклад дистрибуційного центру хлібобулочних виробів, де програмне забезпечення не враховувало вагу продуктів при побудові маршрутів, змушуючи персонал вручну корегувати шлях для запобігання пошкодженню крихких товарів. Для вирішення цієї задачі автори згенерували моделі ситуацій, які базувалися на поведінкових закономірностях реальних даних, зокрема середній кількості замовлень на день, типах продуктів та місткості зон. Але алгоритми були адаптовані для математичного моделювання, що дозволило врахувати обмеження, які ігнорувала штатна система, та запропонувати оптимізовану модель спільного призначення місць зберігання та маршрутизації.

### **Висновки**

Узагальнюючи результати проведеного дослідження, можна констатувати, що цифрова трансформація сучасного логістичного центру виходить за межі простого впровадження автоматизації. Вона являє собою складну синергію роботизованих технологій та інтелектуальних алгоритмів обробки великих масивів даних. Аналіз сучасних підходів довів, що традиційні методи фіксованого зберігання втрачають ефективність в умовах електронної комерції, де критичними є швидкість та адаптивність.

Результати підтверджують доцільність застосування методів машинного навчання, зокрема кластеризації, для оптимізації розміщення товарних позицій та підвищення загальної продуктивності складських процесів. Аналіз окремих практичних кейсів засвідчує реальну ефективність таких підходів. Зокрема, як наведено у роботі [9], апробація алгоритму оптимізації на вибірці з 35 000 маршрутів продемонструвала скорочення дистанції переміщення на 15,7 %, а впровадження моделей глибокого навчання з підкріпленням для динамічного призначення місць зберігання у дослідженні [10] забезпечило зниження транспортних витрат на 6,3 % порівняно з існуючим методом.

Визначено, що використання відкритих репозиторіїв та корпоративних транзакційних даних для навчання моделей дозволяє переходити від реагування на зміни попиту постфактум до завчасного планування запасів. Це дає змогу виявляти приховані закономірності сумісного попиту ще до моменту формування замовлення. Водночас існують виклики, пов'язані з якістю вхідної інформації, оскільки для коректного моделювання критично важливою є наявність точних часових міток, фізичних параметрів вантажу та історії спільних покупок.

З огляду на вищезазначене, подальші наукові дослідження доцільно спрямувати на розробку та експериментальне тестування алгоритмів кластеризації часових рядів на базі відібраних наборів даних, що забезпечить динамічний перерозподіл SKU. Логічним продовженням роботи стане побудова комплексної симуляційної моделі складу для верифікації ефективності запропонованого кластерного підходу та його порівняння з традиційними методами, такими як випадкове або ABC-розміщення. Разом з тим, для підвищення точності прогнозування навантаження на роботизовані системи, у моделях маршрутизації необхідно враховувати фізичні параметри вантажів, зокрема їх вагу та габарити.

### Список літератури

1. Boysen N. Warehousing in the e-commerce era: A survey [Електронний ресурс] / Nils Boysen, René de Koster, Felix Weidinger // *European Journal of Operational Research*. – 2019. – Т. 277, № 2. – С. 396–411. – Режим доступу: <https://doi.org/10.1016/j.ejor.2018.08.023>
2. Low A. K. J. Review on Automated Storage and Retrieval System for Warehouse [Електронний ресурс] / Alex Kai Jie Low, Kok Swee Sim, Kai Liang Lew // *Journal of Informatics and Web Engineering*. – 2024. – Т. 3, № 3. – С. 77–97. – Режим доступу: <https://doi.org/10.33093/jiwe.2024.3.3.5>
3. AI-driven warehouse automation: A comprehensive review of systems [Електронний ресурс] / Enoch Oluwademilade Sodiya [та ін.] // *GSC Advanced Research and Reviews*. – 2024. – Т. 18, № 2. – С. 272–282. – Режим доступу: <https://doi.org/10.30574/gscarr.2024.18.2.0063>
4. Energy consumption model for shuttle-based Storage and Retrieval Systems / Ziwei Liu [та ін.] // *Journal of Cleaner Production*. – 2020. – Режим доступу: <https://doi.org/10.1016/j.jclepro.2020.124480>
5. SIMULATION ANALYSIS OF SHUTTLE BASED STORAGE AND RETRIEVAL SYSTEMS / Tone Lerher [та ін.] // *International Journal of Simulation Modelling*. – 2015. – Т. 14, № 1. – С. 48–59. – Режим доступу: [https://doi.org/10.2507/IJSIMM14\(1\)5.281](https://doi.org/10.2507/IJSIMM14(1)5.281)
6. Collaborative Optimization of Storage Location Assignment and Path Planning in Robotic Mobile Fulfillment Systems [Електронний ресурс] / Jianming Cai [та ін.] // *Sustainability*. – 2021. – Т. 13, № 10. – С. 5644. – Режим доступу: <https://doi.org/10.3390/su13105644>
7. Cai L. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era [Електронний ресурс] / Li Cai, Yangyong Zhu // *Data Science Journal*. – 2015. – Т. 14. – С. 2. – Режим доступу: <https://doi.org/10.5334/dsj-2015-002>
8. Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications [Електронний ресурс] / Benjamin T. Hazen [та ін.] // *International Journal of Production Economics*. – 2014. – Т. 154. – С. 72–80. – Режим доступу: <https://doi.org/10.1016/j.ijpe.2014.04.018>
9. A fast simulated annealing method for batching precedence-constrained customer orders in a warehouse [Електронний ресурс] / Marek Matusiak [та ін.] // *European Journal of Operational Research*. – 2014. – Т. 236, № 3. – С. 968–977. – Режим доступу: <https://doi.org/10.1016/j.ejor.2013.06.001>
10. Dynamic Storage Location Assignment in Warehouses Using Deep Reinforcement Learning [Електронний ресурс] / Constantin Waubert de Puiseau [та ін.] // *Technologies*. – 2022. – Т. 10, № 6. – С. 129. – Режим доступу: <https://doi.org/10.3390/technologies10060129>
11. Warehouse management optimization using a sorting-based slotting approach [Електронний ресурс] / Juan C. Duque-Jaramillo [та ін.] // *Journal of Industrial Engineering and Management*. – 2024. – Т. 17, № 1. – С. 133. – Режим доступу: <https://doi.org/10.3926/jiem.5661>

12. Optimization of the Storage Location Assignment and the Picker-Routing Problem by Using Mathematical Programming [Електронний ресурс] / Johanna Bolaños Zuñiga [та ін.] // Applied Sciences. – 2020. – Т. 10, № 2. – С. 534. – Режим доступу: <https://doi.org/10.3390/app10020534>

A. Balvak, V. Zinchenko

## DATA SAMPLE FORMATION AND QUALITY CRITERIA FOR MODELLING INTELLIGENT WAREHOUSE LOGISTICS SYSTEMS

*In the modern world, the rapid growth of the e-commerce sector places unprecedented strain on logistics systems, requiring warehouse complexes to be highly adaptable and able to process orders quickly. Traditional approaches to warehouse management, based on fixed storage locations and the classification of goods by sales volume, are ineffective in environments with high demand variability and strict delivery time requirements. The article explores the issue of increasing the productivity of warehouse processes by moving from a purely quantitative increase in physical capacity to the implementation of intelligent data processing algorithms.*

*The purpose of this work is to determine the prerequisites and tools for algorithmic optimization of the placement of commodity items based on the identification of hidden patterns of joint demand. First, an analysis of current technological trends was conducted, which demonstrated a shift in emphasis from picker-to-parts systems to robotic parts-to-picker systems (MRFS, AGV), the effectiveness of which significantly depends on the quality of routing algorithms and the assignment of storage locations.*

*With the growth of data volumes, the key factor for success is the formation of a high-quality information base for modelling. The paper examines in detail the criteria for data relevance for logistics tasks, including the presence of accurate time stamps, transactional integrity, and detailed physical parameters of cargo. Particular attention is paid to both the advantages and potential challenges of working with open data sources. A systematic review of eight selected datasets from the Kaggle and UCI Machine Learning Repositories was conducted, and their suitability for solving specific tasks was analysed: from time series clustering and peak load forecasting to packaging optimization.*

*Separately, the role of internal corporate data of enterprises is highlighted, in particular operational logs of warehouse management systems, which allow taking into account real infrastructure limitations and behavioural factors of personnel. It is concluded that the digital transformation of a modern logistics centre is a synergy of robotic technologies and machine learning methods. The results of the study confirm the feasibility of using clustering for dynamic SKU redistribution, which allows minimizing movement distances and transforming logistics from a model of responding to demand post-facto into a system of planning based on forecasts.*

**Keywords:** warehouse logistics; optimization of goods placement; data clustering; machine learning; parts-to-picker systems; picker-to-parts systems; e-commerce; big data analysis; demand forecasting.

---

Надійшла до редакції: 25.11.2025

Прийнята до друку: 26.12.2025

Опубліковано: 27.02.2026

© 2026 Балвак А. А., Зінченко В. В. Цей матеріал ліцензовано за умовами СС ВУ 4.0.<https://creativecommons.org/licenses/by/4.0>