

## БАГАТОФАКТОРНИЙ СТАТИСТИЧНИЙ АНАЛІЗ УСПІШНОСТІ ЗАМОВЛЕНЬ ЯК ІНСТРУМЕНТ ОПТИМІЗАЦІЇ АВТОМАТИЗОВАНИХ СИСТЕМ КЕРУВАННЯ РЕСУРСАМИ

**Мірошниченко С.О.** студент, ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА», м. Запоріжжя, ORCID: <https://orcid.org/0009-0007-4868-3006>, e-mail: [serhii.miroshnychenko@mipolytech.eduication](mailto:serhii.miroshnychenko@mipolytech.eduication);

**Вовна О.В.** д-р техн. наук, професор, ТОВ «ТЕХНІЧНИЙ УНІВЕРСИТЕТ «МЕТІНВЕСТ ПОЛІТЕХНІКА», м. Запоріжжя, ORCID: <https://orcid.org/0000-0003-4433-7097>, e-mail: [oleksandr.vovna@mipolytech.eduication](mailto:oleksandr.vovna@mipolytech.eduication)

У статті розглянуто методологію багатофакторного статистичного аналізу для прогнозування успішності замовлень у B2B-сегменті як інструмент оптимізації автоматизованих систем керування ресурсами підприємства. Дослідження зосереджено на виявленні ключових факторів, що впливають на успішність замовлень, та розробці методології, яка автоматично адаптується до характеру розподілу даних, обираючи відповідні статистичні тести та методи аналізу. Представлено комплексний підхід, який включає статистичний аналіз історичних даних ERP-систем, побудову прогностичних моделей та їх інтеграцію в автоматизовані системи керування ресурсами підприємства. Запропонована методологія дозволяє підвищити точність прогнозування успішності замовлень, оптимізувати управління запасами та виробничими процесами, а також покращити якість прийняття управлінських рішень. Результати дослідження підтверджують доцільність застосування методології для підвищення ефективності використання ресурсів підприємства та зниження операційних ризиків.

**Ключові слова:** багатофакторний аналіз, статистичне прогнозування, успішність замовлень, B2B-сегмент, автоматизовані системи керування ресурсами, ERP-системи, оптимізація бізнес-процесів.

### Постановка проблеми

Ефективне керування ресурсами підприємства залишається однією з ключових проблем сучасного бізнесу [1]. При автоматизації цього процесу одним з найважливіших питань є контроль поточної та прогнозування майбутньої кількості продукції на складах [2], що дозволяє своєчасно виготовляти продукцію та, за необхідності, закупати компоненти для забезпечення виробничого процесу [3]. Основним підходом до прогнозування майбутньої кількості продукції є моніторинг динаміки замовлень на відповідні товари та аналіз їх відпуску зі складських приміщень. Для оперативного реагування на створені, але ще не погоджені, замовлення постає задача прогнозування їхньої успішності, спираючись на історичні дані системи планування ресурсів підприємства (ERP) [4].

Багатофакторний статистичний аналіз є потужним інструментом, який дозволяє виявляти приховані взаємозв'язки між різними характеристиками замовлень та їхньою кінцевою успішністю. Це дає можливість розробити моделі прогнозування, що допомагають оптимізувати бізнес-процеси та підвищити ефективність використання ресурсів підприємства [5], [6]. У контексті B2B-сегмента прогнозування успішності замовлень набуває критичного значення внаслідок підвищеної вартості транзакцій та стратегічної важливості довгострокових партнерських відносин. Аналіз історичних даних замовлень дозволяє виявити фактори, що впливають на ймовірність успішного виконання замовлення, що, в свою чергу, дає можливість проактивно управляти ризиками та ресурсами підприємства [7], [8].

### Аналіз останніх досліджень та публікацій

Дослідження в області прогнозування успішності бізнес-операцій та оптимізації ресурсів підприємства викликають значний інтерес як в академічному середовищі, так і в практичній сфері [9], [10]. Зокрема, автори [1] досліджували вплив застосування ERP на раціональне використання ресурсів компанії та на основі даних ERP для різних типів операційних моделей компанії виділили фактори раціонального використання ресурсів. У роботі підкреслено, що системна інтеграція сприяє оптимізації використання ресурсів через автоматизацію рутинних бізнес-процесів. Дослідження [11] продемонструвало, що імплементація систем планування ресурсів підприємства позитивно корелює з розширенням інтеграції постачальників, оптимізацією внутрішньої інтеграції та впровадженням екологічно орієнтованих практик управління ланцюгами постачання. Емпіричні результати, отримані зазначеними дослідниками, підтверджують гіпотезу про те, що дані технологічні рішення сприяють підвищенню операційної ефективності підприємств. Автори [12] провели аналіз історії, особливостей та критичних факторів успіху впровадження ERP в епоху Industry 4.0, визначивши ключові ознаки, серед яких підтримка топменеджменту, проєктна команда, IT-інфраструктура, комунікація, кваліфіковані співробітники, навчання та моніторинг.

У роботі [8] досліджено підвищення гнучкості та стійкості ланцюгів постачання за допомогою машинного навчання з фокусуванням на методах оптимізації для логістики та управління запасами. Автор продемонстрував, що застосування моделей машинного навчання дозволило значно покращити точність

прогнозування попиту та оптимізувати рівні запасів. Порівняння статистичних та методів машинного навчання для часових рядів у логістиці проводили автори [5], які виявили, що методи машинного навчання, такі як нейронні мережі та Random Forests, часто перевершують традиційні моделі часових рядів у точності прогнозування. Однак вони також відзначають, що статистичні методи залишаються актуальними через їхню інтерпретованість та ефективність в певних умовах. Застосування методів машинного навчання для прогнозування успіху стартапів досліджували в роботі [9], де автори порівняли три алгоритми машинного навчання: логістичну регресію, дерево рішень та Random Forest. Попри обмежені дані, вони отримали прийнятні результати з точки зору точності, чутливості та F-міри, а найкращою моделлю виявилось дерево рішень. Автори [13] вивчали вплив даних про роботу з відеоматеріалами на точність моделей прогнозування успішності студентів. Автори використовували алгоритми Naive Bayes, логістичної регресії, Random Forest та нейронних мереж і виявили, що включення даних про перегляд відео підвищило точність моделей на 10%, а найвищу точність прогнозування досягли моделі, побудовані з використанням алгоритмів Random Forest (87,1%) та нейронних мереж (85,3%). Дослідження [10] оцінює ефективність різних алгоритмів машинного навчання у прогнозуванні успіху стартапів та досліджує покращення продуктивності, досягнуте шляхом застосування методу головних компонент (PCA) до моделей. Зокрема, Support Vector Classifier показав значне підвищення продуктивності при використанні з PCA, фіксуючи точність 0,90, що демонструє важливість зменшення розмірності даних для підвищення точності предиктивних моделей.

Аналіз наявних досліджень свідчить про високий потенціал методів машинного навчання для прогнозування в бізнес-контексті, однак також підкреслює необхідність розуміння базових статистичних властивостей даних. Попри значний прогрес у розвитку методологічного апарату прогнозування та оптимізації ресурсного забезпечення організацій, виявлено низку концептуальних і методологічних проблем, які потребують поглибленого наукового дослідження та системного методологічного вирішення. Також існує потреба в більш ґрунтовному дослідженні специфічних факторів, що впливають на успішність замовлень у B2B-сегменті [14]. Цей ринковий сектор характеризується унікальними особливостями, такими як складність процесу прийняття рішень, важливість довгострокових партнерських відносин та необхідність індивідуального підходу до кожного клієнта. Водночас актуальною є розробка методологічного підходу, який враховуватиме характеристики розподілу даних при виборі аналітичних методів, оскільки автоматичний відбір відповідних статистичних тестів залежно від властивостей емпіричних даних є критичним для підвищення предиктивної точності. Дослідження у сфері прогнозування B2B-продажів підтверджують складність цього

завдання та необхідність врахування специфічних факторів. Наприклад, дослідження, проведене на даних компанії зі списку Fortune 500, показало, що найточніша предиктивна модель досягла точності лише 80%, що свідчить про наявність значного потенціалу для вдосконалення [14]. Крім того, більшість досліджень зосереджується на загальних методах прогнозування попиту та управління запасами, без врахування специфіки B2B-сегмента та особливостей прогнозування успішності окремих замовлень [7]. Також багато дослідників для прогнозування бізнес-показників використовують готові методи машинного навчання без детального статистичного аналізу властивостей даних [9], [10]. Це може призводити до неоптимального вибору методів аналізу та, як наслідок, до неточних прогнозів.

Результати проведеного аналізу досліджень та публікацій демонструють наявність потреби в розробці методології, яка б автоматично адаптувалася до характеру розподілу даних, обираючи відповідні статистичні тести та методи аналізу. Такий підхід дозволив би підвищити точність прогнозування в умовах різномірних даних, що характерні для реальних бізнес-процесів [15]. Представлені дослідження також часто не приділяють достатньої уваги інтерпретації результатів статистичного аналізу та їх практичному застосуванню для оптимізації автоматизованих систем керування ресурсами. Розуміння значущості різних факторів та їх впливу на успішність замовлень має критичне значення для прийняття обґрунтованих управлінських рішень [4]. Також варто зазначити, що у більшості досліджень відсутній комплексний підхід до прогнозування успішності замовлень, який поєднуватиме статистичний аналіз з інтеграцією результатів у ERP-системи та інші платформи автоматизації підприємства [16]. Реалізація такого підходу забезпечить оптимальне використання аналітичних результатів для вдосконалення бізнес-процесів [2].

Таким чином, актуальною є розробка комплексної методології багатофакторного статистичного аналізу успішності замовлень, яка враховуватиме особливості B2B-сектора, автоматично адаптуватиметься до характеристик розподілу даних та забезпечуватиме інтеграцію результатів у корпоративні інформаційні системи.

### Мета статті

Мета цього дослідження полягає у розробці та науковому обґрунтуванні методології багатофакторного статистичного аналізу для прогнозування успішності замовлень у B2B-сегменті з метою оптимізації автоматизованих систем керування ресурсами підприємства.

Для досягнення цієї мети визначено наступні завдання дослідження:

1. Розробити методологію статистичного аналізу, яка враховує специфіку розподілу досліджуваних даних та забезпечує адекватний вибір статистичних методів і критеріїв оцінювання.

2. Провести багатфакторний статистичний аналіз історичних даних замовлень для виявлення факторів, що впливають на їхню успішність.

3. Розробити статистичну модель прогнозування успішності замовлень на основі виявлених значущих факторів.

4. Провести валідацію розробленої статистичної моделі та оцінити її прогностичну ефективність з визначенням найбільш впливових факторів успішності замовлень.

5. Розробити рекомендації щодо інтеграції результатів статистичного аналізу в автоматизовані системи керування ресурсами підприємства та визначити напрями подальших досліджень.

### Виклад основного матеріалу

У цій роботі авторами розроблено статистичну модель для прогнозування успішності замовлень на основі порівняльних статистичних тестів, що автоматично обирає методи аналізу та визначає статистичну значущість факторів, які впливають на результат. Методологія охоплює кілька етапів: завантаження та підготовку даних, обробку викидів, розподіл даних на тренувальну та тестову вибірки, оцінку нормальності розподілу даних, вибір відповідного статистичного тесту для аналізу значущості ознак, навчання моделі, прогнозування та оцінку її ефективності. Вихідний набір даних включає 86 794 записів, що містять інформацію про замовлення з 19 характеристиками (ознаками). Цільова змінна характеризує успішність замовлення – дорівнює 1 у випадку схваленого та успішного виконання або 0 в іншому випадку.

Початковий етап передбачає збір історичних даних, після чого здійснюється ретельна перевірка та обробка пропущених значень шляхом заповнення медіаною відповідного стовпця, що є робастним підходом, стійким до викидів [17]. З метою забезпечення логічної консистентності масиву даних здійснюється корекція від'ємних значень шляхом їх заміни нульовими значеннями для тих параметрів, де негативні величини не мають змістовної інтерпретації в контексті фізичних або економічних процесів [18], [19], а також виявлення та обробка викидів методом міжквартильного розмаху (IQR) із порогом 1,5 [20], після чого значення обмежуються граничними межами (вінсоризація), що дозволяє зберегти всі спостереження, мінімізуючи вплив екстремальних величин [21]. Дані розподіляються на тренувальну та тестову вибірки у співвідношенні 70:30 із застосуванням стратифікованого розбиття для збереження пропорцій цільової змінної [22].

Ключовим елементом розробленої методології є систематичний аналіз характеру розподілу досліджуваних даних та науково обґрунтований вибір відповідного статистичного інструментарію. Оцінка нормальності розподілу проводиться для кожної числової ознаки окремо в групах успішних і неуспішних замовлень за допомогою трьох тестів: Shapiro-Wilk [23],

D'Agostino-Pearson [24] та Anderson-Darling [25]. Висновки про нормальність статистичного розподілу ознаки приймається за умови підтвердження цієї гіпотези принаймні двома з трьох застосованих тестів. На основі результатів оцінки нормальності обирається статистичний тест для аналізу значущості ознак. Для нормально розподілених даних застосовується Welch's t-test [26], а розмір ефекту обчислюється за Cohen's d [27]. У разі ненормального розподілу використовується непараметричний тест Mann-Whitney [28], де розмір ефекту оцінюється через площу під ROC-кривою (AUC) [29].

Навчання розробленої моделі включає три етапи: (1) відбір статистично значущих ознак ( $p \leq 0,05$ ), (2) розрахунок порогових значень та (3) визначення ваг ознак. Порогові значення обчислюються як середнє арифметичне між центральними тенденціями груп успішних і неуспішних замовлень, де використовуються середні значення за нормального розподілу або медіани за ненормального розподілу. Для кожної ознаки також визначається напрямок впливу та відповідна вага. Для параметричного тесту вага визначається як  $w = 0,5 + 0,5 \times \tanh\left(\frac{\text{Cohen's } d}{2}\right)$ , а для непараметричного як  $w = 2 \times |AUC - 0,5|$ , якщо  $AUC \geq 0,5$ , інакше 0. Ця процедура дозволяє уніфікувати розміри ефекту в межах інтервалу  $[0; 1]$ , що забезпечує порівнянність результатів незалежно від обраного статистичного методу. Прогнозування здійснюється шляхом обчислення «голосу» кожної значущої ознаки для спостереження: голос дорівнює вазі ознаки, якщо значення перевищує (або не перевищує) поріг залежно від напрямку впливу, інакше 0. Сумарний голос нормалізується до ймовірності успішності, а прогнозований клас визначається як 1, якщо ймовірність перевищує 0,5, інакше 0.

Прогнозована ймовірність успішності обчислюється за наступною формулою:

$$p = \frac{1}{|\mathcal{F}|} \sum_{f \in \mathcal{F}} \frac{w_f}{W} \cdot \frac{1 + \text{sgn}((x_f - \theta_f) \cdot d_f)}{2},$$

де  $p$  – прогнозована ймовірність успішності;  $\mathcal{F}$  – множина значущих ознак;  $|\mathcal{F}|$  – кількість ознак у множині  $\mathcal{F}$ ;  $w_f$  – вага ознаки  $f \in \mathcal{F}$ ;  $W$  – сумарна вага всіх ознак,  $W = \sum_{f \in \mathcal{F}} w_f$ ;  $x_f$  – значення ознаки  $f$  для спостереження;  $\theta_f$  – порогове значення для ознаки  $f$ ;  $d_f$  – напрямок впливу ознаки ( $d_f = +1$ , якщо більші значення асоціюються з успішністю, інакше  $d_f = -1$ );  $\text{sgn}()$  – функція, яка приймає значення 1 для додатних аргументів,  $-1$  для від'ємних та 0 для нульових.

Ця методологія забезпечує структурований підхід до аналізу даних і прогнозування, що дозволяє ідентифікувати ключові фактори успішності замовлень із високою статистичною достовірністю. Оцінка ефективності розробленої моделі включає розрахунок

комплексу метрик [30]: точність, збалансована точність, влучність, повнота, F1-score, ROC AUC та специфічність.

У результаті статистичного аналізу вхідного масиву даних було виявлено 17 статистично значущих ознак із 19 досліджуваних. Для кожної значущої ознаки було розраховано вагу (важливість) на основі розміру ефекту та визначено порогове значення для прийняття рішення. В табл. 1 наведено інформацію результатів аналізу для статистично значущих ознак. Варто зазначити, що для більшості досліджуваних факторів було встановлено позитивний напрямок впливу (+1), за

винятком змінних `hour_of_day` та `partner_avg_amount`, для яких виявлено негативний вплив (-1).

Результати аналізу демонструють, що комунікаційні фактори (кількість повідомлень) та показники попереднього досвіду співпраці з партнером (рівень успішності партнера) виявляють найбільш виражений детермінантний вплив на успішність замовлень. Це підтверджує важливість якісної комунікації та довготривалих партнерських відносин у B2B-сегменті.

Для кращого розуміння характеру впливу найбільш та найменш значущих ознак, було проведено порівняльний аналіз їх статистичних показників у групах успішних та неуспішних замовлень (табл. 2).

Таблиця 1

Ваги та порогові значення значущих ознак замовлень

| Ознака                                                                                                     | Порогове значення | Вага (важливість) |
|------------------------------------------------------------------------------------------------------------|-------------------|-------------------|
| Кількість повідомлень в замовленні ( <code>order_messages</code> )                                         | 7,00              | 0,5421            |
| Відсоток успішних замовлень у клієнта ( <code>partner_success_rate</code> )                                | 62,10             | 0,3956            |
| Кількість змін в замовленні ( <code>order_changes</code> )                                                 | 2,50              | 0,2828            |
| Загальна кількість повідомлень клієнта ( <code>partner_total_messages</code> )                             | 224,50            | 0,1861            |
| Загальна кількість замовлень клієнта ( <code>partner_total_orders</code> )                                 | 24,50             | 0,1777            |
| Кількість позицій в замовленні ( <code>order_lines_count</code> )                                          | 2,50              | 0,1666            |
| Строк співпраці з клієнтом в днях ( <code>partner_order_age_days</code> )                                  | 736,00            | 0,1564            |
| Середня кількість змін в замовленнях у клієнта ( <code>partner_avg_changes</code> )                        | 3,89              | 0,1436            |
| Середня сума в успішних замовленнях клієнта ( <code>partner_success_avg_amount</code> )                    | 904,14            | 0,1307            |
| Сума замовлення ( <code>order_amount</code> )                                                              | 512,88            | 0,1179            |
| Середня кількість повідомлень в неуспішних замовленнях клієнта ( <code>partner_fail_avg_messages</code> )  | 5,62              | 0,0871            |
| Середня кількість змін в успішних замовленнях клієнта ( <code>partner_success_avg_changes</code> )         | 4,29              | 0,0856            |
| Середня кількість змін в неуспішних замовленнях клієнта ( <code>partner_fail_avg_changes</code> )          | 2,93              | 0,0777            |
| Година дня, коли зроблено замовлення ( <code>hour of day</code> )                                          | 11,00             | 0,0767            |
| Середня кількість повідомлень в успішних замовленнях клієнта ( <code>partner_success_avg_messages</code> ) | 9,44              | 0,0645            |
| Середня сума в замовленнях клієнта ( <code>partner_avg_amount</code> )                                     | 1127,29           | 0,0364            |
| Квартал року, коли зроблено замовлення ( <code>quarter</code> )                                            | 2,50              | 0,0182            |

Таблиця 2

Порівняння статистичних показників найбільш та найменш впливових ознак

| Ознака                                    | Середнє   |         | Медіана   |         |
|-------------------------------------------|-----------|---------|-----------|---------|
|                                           | неуспішні | успішні | неуспішні | успішні |
| <code>order_messages</code>               | 6,09      | 10,08   | 5,00      | 9,00    |
| <code>partner_success_rate</code>         | 45,90     | 67,46   | 50,00     | 74,20   |
| <code>order_changes</code>                | 2,71      | 3,74    | 2,00      | 3,00    |
| <code>partner_success_avg_messages</code> | 8,71      | 9,70    | 9,48      | 9,50    |
| <code>partner_avg_amount</code>           | 1835,28   | 1745,37 | 1133,22   | 1121,35 |
| <code>quarter</code>                      | 2,46      | 2,49    | 2,00      | 3,00    |

Статистичний аналіз даних, представлених у Таблиці 2, свідчить про наявність статистично значущих відмінностей у показниках центральної тенденції (середніх арифметичних та медіанних значеннях) найвпливовіших предикторів між групами успішних та неуспішних замовлень. Наприклад, середня кількість повідомлень в успішних замовленнях (10,08) перевищує цей показник для неуспішних замовлень (6,09) в 1,77 рази.

Для оцінки ефективності розробленої статистичної моделі було проведено тестування на незалежній тестовій вибірці (30% початкових даних, 26039 замовлень). Результати оцінки представлено в таблиці 3.

Таблиця 3

Результати оцінки якості статистичної моделі

| Метрика                                   | Значення | Інтерпретація                                        |
|-------------------------------------------|----------|------------------------------------------------------|
| Точність (Accuracy)                       | 0,6557   | Частка правильно класифікованих замовлень            |
| Збалансована точність (Balanced Accuracy) | 0,6582   | Середнє арифметичне чутливості та специфічності      |
| Влучність (Precision)                     | 0,7694   | Частка дійсно успішних серед передбачених як успішні |
| Повнота (Recall)                          | 0,6489   | Частка виявлених успішних замовлень                  |
| F1-міра                                   | 0,7040   | Гармонічне середнє точності та повноти               |
| ROC AUC                                   | 0,7224   | Площа під ROC-кривою, якість класифікації            |
| Специфічність (Specificity)               | 0,6675   | Частка правильно ідентифікованих неуспішних          |

Емпіричні результати засвідчують, що розроблена модель демонструє загальну точність класифікації 65,57%, що відповідає задовільному рівню ефективності для завдань прогнозування економічних індикаторів у бізнес-середовищі [6]. Збалансована точність (65,82%) свідчить про рівномірну ефективність моделі для обох класів (успішні та неуспішні замовлення) [31]. Влучність 76,94% означає, що серед замовлень, які модель класифікує як успішні, близько 77% дійсно є успішними [32]. Це важливий показник для бізнесу, оскільки дозволяє ефективно розподіляти ресурси, концентруючись на замовленнях з високою ймовірністю успіху.

Значення ROC AUC 0,7224 вказує на здатність моделі розрізняти успішні та неуспішні замовлення з точністю, що на 22,24% перевищує випадкове передбачення [33], [34]. Згідно з інтерпретацією AUC [35], значення в діапазоні 0,7-0,8 свідчить про задовільну якість моделі.

Розподіл прогнозованих ймовірностей за фактичними класами, що наведено на рис. 1, показує, що модель здатна розділяти успішні та неуспішні замовлення, хоча існує значна область перекриття, де ймовірності коливаються в діапазоні від 0,3 до 0,7. Це свідчить про складність задачі прогнозування успішності замовлень та необхідність подальшого вдосконалення моделі.

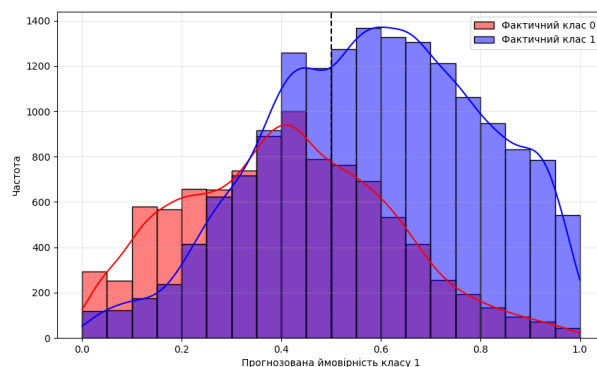


Рис. 1 – Розподіл прогнозованих ймовірностей за фактичними класами

Аналіз матриці помилок показує, що модель правильно ідентифікувала 6415 неуспішних та 10660 успішних замовлень, при цьому 3195 неуспішних були помилково класифіковані як успішні, а 5769 успішних – як неуспішні. Це вказує на потенціал для подальшого вдосконалення моделі.

### Висновки

Розроблено та емпірично верифіковано методологію багатофакторного статистичного аналізу для прогнозування успішності замовлень у B2B-сегменті як інструмент оптимізації автоматизованих систем керування ресурсами підприємства.

Методологія автоматичного вибору статистичних тестів забезпечила надійний аналіз гетерогенних бізнес-даних. Виявлено сімнадцять статистично значущих предикторів успішності замовлень, найвагоміші з яких: кількість повідомлень (0,5421), рівень успішності партнера (0,3956) та кількість змін (0,2828).

Розроблена модель прогнозування продемонструвала показники: точність 65,57%, влучність 76,94%, ROC AUC 0,7224, підтвердивши визначальний вплив комунікаційних факторів та попереднього досвіду співпраці на успішність замовлень.

Значущість дослідження полягає у формуванні концептуально-методологічного апарату статистичного прогнозування бізнес-процесів. Практичне застосування включає оптимізацію складських запасів, пріоритезацію замовлень та стратегічне планування партнерських відносин.

Подальші дослідження доцільно спрямувати на впровадження методів машинного навчання, зокрема алгоритмів Random Forest та нейронних мереж, що

дозволить підвищити точність прогнозування та ефективність аналітичних процедур. Особливу увагу варто приділити розробці гібридних моделей, які поєднують традиційні методи статистичного аналізу з сучасними підходами машинного навчання, компенсуючи обмеження кожного з методів окремо. Не менш важливим є дослідження динамічних аспектів функціонування системи та впровадження механізмів автоматичної адаптації моделі до змінних умов середовища, що забезпечить стабільність та надійність отриманих результатів у довгостроковій перспективі.

Дослідження демонструє значний потенціал статистичних методів для оптимізації систем керування ресурсами підприємства та окреслює перспективи інтеграції з сучасними методами аналізу даних.

#### Перелік використаних джерел

- [1] Markuts V., Kyzenko O. ERP system as a tool for ensuring the rational use of company resources. *Scientific notes*. 2023. No. 32. Pp. 68-78. DOI: [https://doi.org/10.33111/vz\\_kneu.32.23.03.06.045.05\\_1](https://doi.org/10.33111/vz_kneu.32.23.03.06.045.05_1).
- [2] Qureshi M.R.N.M. Evaluating Enterprise Resource Planning (ERP) Implementation for Sustainable Supply Chain Management. *Sustainability*. 2022. Vol. 14, no. 22. Article 14779. DOI: <https://doi.org/10.3390/su142214779>.
- [3] Enterprise Resource Planning Systems for Health, Safety, and Environment Management: Analyzing Critical Success Factors / M. AlMarri et al. *Sustainability*. 2025. Vol. 17, no. 7. Article 2947. DOI: <https://doi.org/10.3390/su17072947>.
- [4] AI-driven demand forecasting: Enhancing inventory management and customer satisfaction / Olamide Raimat Amosu et al. *World Journal of Advanced Research and Reviews*. 2024. Vol. 23, no. 2. Pp. 708-719. DOI: <https://doi.org/10.30574/wjarr.2024.23.2.2394>.
- [5] Comparing Statistical and Machine Learning Methods for Time Series Forecasting in Data-Driven Logistics—A Simulation Study / L. Schmid et al. *Entropy*. 2024. Vol. 27, no. 1. Article 25. DOI: <https://doi.org/10.3390/e27010025>.
- [6] Statistical Predictions of Trading Strategies in Electronic Markets / Á. Cartea et al. *Journal of Financial Econometrics*. 2024. Vol. 23, iss. 2. Pp. 1-64. DOI: <https://doi.org/10.1093/jjfinec/nbae025>.
- [7] Walter A., Ahsan K., Rahman S. Application of artificial intelligence in demand planning for supply chains: a systematic literature review. *The International Journal of Logistics Management*. 2025. Vol. 36, no. 3. Pp. 672-719. DOI: <https://doi.org/10.1108/ijlm-02-2024-0120>.
- [8] Ahmadi Z. The Role of AI and Machine Learning in Supply Chain Optimization. *International Journal of Modern Achievement in Science, Engineering and Technology*. 2025. Vol. 2, no. 2. Pp. 1-8. DOI: <https://doi.org/10.63053/ijset.77>.
- [9] Applying Machine Learning Approach to Start-up Success Prediction / O. Piskunova et al. *Scientific Horizons*. 2021. Vol. 24, no. 11. Pp. 72-84. DOI: [https://doi.org/10.48077/scihor.24\(11\).2021.72-84](https://doi.org/10.48077/scihor.24(11).2021.72-84).
- [10] Choi Y. Startup Success Prediction with PCA-Enhanced Machine Learning Models. *Journal of technology management & innovation*. 2024. Vol. 19, no. 4. Pp. 77-88. DOI: <https://doi.org/10.4067/s0718-27242024000400077>.
- [11] Tarigan Z.J.H., Siagian H., Jie F. Impact of Enhanced Enterprise Resource Planning (ERP) on Firm Performance through Green Supply Chain Management. *Sustainability*. 2021. Vol. 13, no. 8. Article 4358. DOI: <https://doi.org/10.3390/su13084358>.
- [12] History, Features, Challenges, and Critical Success Factors of Enterprise Resource Planning (ERP) in The Era of Industry 4.0 / M. Al-Amin et al. *European Scientific Journal, ESJ*. 2023. Vol. 19, no. 6. Pp. 31-59. DOI: <https://doi.org/10.19044/esj.2023.v19n6p31>.
- [13] Determining the influence of data on working with video materials on the accuracy of student success prediction models / V. Pylypenko et al. *Eastern-European Journal of Enterprise Technologies*. 2024. Vol. 5, no. 4(131). Pp. 52-62. DOI: <https://doi.org/10.15587/1729-4061.2024.313333>.
- [14] Predicting and Defining B2B Sales Success with Machine Learning / S. Mortensen et al. *2019 Systems and Information Engineering Design Symposium (SIEDS) : conference proceedings, Charlottesville, 26 April 2019. NY, 2019. Pp. 352-356. DOI: <https://doi.org/10.1109/SIEDS.2019.8735638>*.
- [15] McQuitty S. The Purposes of Multivariate Data Analysis Methods: an Applied Commentary. *Journal of African Business*. 2017. Vol. 19, no. 1. Pp. 124-142. DOI: <https://doi.org/10.1080/15228916.2017.1374816>.
- [16] Prediction for the Inventory Management Chaotic Complexity System Based on the Deep Neural Network Algorithm / T. Lei et al. *Complexity*. 2023. Vol. 2023. Pp. 1-11. DOI: <https://doi.org/10.1155/2023/9369888>.
- [17] Oluwaseye Joel, L., Doorsamy W., Sena Paul B. A Review of Missing Data Handling Techniques for Machine Learning. *International Journal of Innovative Technology and Interdisciplinary Sciences*. 2022. Vol. 5(3). Pp. 971-1005. DOI: <https://doi.org/10.1515/IJTIS.2022.5.3.971-1005>.
- [18] Arimie C., Biu E., Ijomah M. Outlier Detection and Effects on Modeling. *Open Access Library Journal*. 2020. Vol. 7. Pp. 1-10. DOI: <https://doi.org/10.4236/oalib.1106619>.
- [19] Shved A. V., Davydenko Y. O. Outlier detection technique for heterogeneous data using trimmed-mean robust estimators. *Radio Electronics, Computer Science, Control*. 2022. No. 3. Pp. 50-57. DOI: <https://doi.org/10.15588/1607-3274-2022-3-5>.

- [20] A Modern Introduction to Probability and Statistics / F.M. Dekking et al. London : Springer London, 2005. 488 p. DOI: <https://doi.org/10.1007/1-84628-168-7>.
- [21] MSCI global investable market value and growth index methodology. 2021. 37 p. URL: [https://www.msci.com/eqb/methodology/methods/MSCI\\_GIMIVGMethod\\_Feb2021.pdf](https://www.msci.com/eqb/methodology/methods/MSCI_GIMIVGMethod_Feb2021.pdf) (дата звернення: 05.01.2024).
- [22] Särndal C.-E., Swensson B., Wretman J. Model Assisted Survey Sampling (Springer Series in Statistics). Springer, 2003. 694 p.
- [23] Shapiro S. S., Wilk M. B. An analysis of variance test for normality (complete samples). *Biometrika*. 1965. Vol. 52, no. 3-4. Pp. 591-611. DOI: <https://doi.org/10.1093/biomet/52.3-4.591>.
- [24] D'Agostino R., Pearson E.S. Tests for departure from normality. Empirical results for the distributions of  $b_2$  and  $\sqrt{b_1}$ . *Biometrika*. 1973. Vol. 60, no. 3. Pp. 613-622. DOI: <https://doi.org/10.1093/biomet/60.3.613>.
- [25] Anderson T.W., Darling D.A. Asymptotic Theory of Certain «Goodness of Fit» Criteria Based on Stochastic Processes. *The Annals of Mathematical Statistics*. 1952. Vol. 23, no. 2. Pp. 193-212. DOI: <https://doi.org/10.1214/aoms/1177729437>.
- [26] Welch B.L. The generalization of 'student's' problem when several different population variances are involved. *Biometrika*. 1947. Vol. 34, no. 1-2. Pp. 28-35. DOI: <https://doi.org/10.1093/biomet/34.1-2.28>.
- [27] Cohen J. Statistical Power Analysis for the Behavioral Sciences. New York : Routledge, 2013. 567 p. DOI: <https://doi.org/10.4324/9780203771587>.
- [28] Mann H.B., Whitney D.R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*. 1947. Vol. 18, no. 1. Pp. 50-60. DOI: <https://doi.org/10.1214/aoms/1177730491>.
- [29] Swets J.A. Signal detection theory and ROC-analysis. New York : Academic Press, 1975. 277 p.
- [30] Ben-David S., Shalev-Shwartz S. Understanding Machine Learning: From Theory to Algorithms. Cambridge University Press, 2014. 449 p.
- [31] The Balanced Accuracy and Its Posterior Distribution / K.H. Brodersen et al. *2010 20th International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 23-26 August 2010. Pp. 3121-3124. DOI: <https://doi.org/10.1109/icpr.2010.764>.
- [32] Kelleher J.D., D'Arcy A., Namee B.M. Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. MIT Press, 2015. 624 p.
- [33] Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters*. 2006. Vol. 27, no. 8. Pp. 861-874. DOI: <https://doi.org/10.1016/j.patrec.2005.10.010>.
- [34] Nonparametric receiver operating characteristic curve analysis with an imperfect gold standard / J. Sun et al. *Biometrics*. 2024. Vol. 80, no. 3. Pp. 1-12. DOI: <https://doi.org/10.1093/biomtc/ujae063>.
- [35] Çorbacıoğlu Ş., Aksel G. Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value. *Turkish Journal of Emergency Medicine*. 2023. Vol. 23, no. 4. Pp. 195-198. DOI: [https://doi.org/10.4103/tjem.tjem\\_182\\_23](https://doi.org/10.4103/tjem.tjem_182_23).

## MULTIFACTOR STATISTICAL ANALYSIS OF ORDER SUCCESS AS A TOOL FOR OPTIMIZATION OF AUTOMATED RESOURCE PLANNING AND CONTROL SYSTEMS

|                            |                                                                                                                                                                                                                                                                                                            |
|----------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>Miroshnichenko S.O.</b> | student, «TECHNICAL UNIVERSITY «METINVEST POLYTECHNIC» LLC, Zaporizhzhia, ORCID: <a href="https://orcid.org/0009-0007-4868-3006">https://orcid.org/0009-0007-4868-3006</a> , e-mail: <a href="mailto:serhii.miroshnichenko@mipolytech.education">serhii.miroshnichenko@mipolytech.education</a> ;          |
| <b>Vovna O.V.</b>          | D.Sc. (Engineering), professor, «TECHNICAL UNIVERSITY «METINVEST POLYTECHNIC» LLC, Zaporizhzhia, ORCID: <a href="https://orcid.org/0000-0003-4433-7097">https://orcid.org/0000-0003-4433-7097</a> , e-mail: <a href="mailto:oleksandr.vovna@mipolytech.education">oleksandr.vovna@mipolytech.education</a> |

Effective resource management remains a critical challenge in automated enterprise systems. This study addresses predicting order success in B2B segments through multifactor statistical analysis methodology for optimizing automated resource management systems based on historical ERP data. The methodology encompasses automatic selection of statistical tests based on data distribution characteristics. Data preprocessing includes outlier detection using interquartile range methods, missing value imputation, and stratified splitting. Normality assessment uses Shapiro-Wilk, D'Agostino-Pearson, and Anderson-Darling tests, with parametric t-tests for normal distributions and Mann-Whitney U tests for non-normal data. The study analyzed historical B2B order data examining 19 predictor variables across order characteristics, partner metrics, communication patterns, and temporal factors. Statistical significance was evaluated at  $\alpha \leq 0,05$ , with effect sizes calculated using Cohen's  $d$  and Area Under Curve. The predictive model employs weighted voting where significant features contribute to probability calculation based on statistical weight. Results identified seventeen statistically significant predictors. Most influential factors include order message count, partner success rate, and order

changes. The model achieved accuracy of 65,57%, precision of 76,94%, recall of 64,89%, F1-score of 70,40%, and ROC AUC of 0,7224, indicating satisfactory performance for business applications. Communication factors emerge as strongest predictors, with successful orders averaging 10,08 messages versus 6,09 for unsuccessful orders. Partner historical success rates show substantial predictive power, with successful orders associated with 67,46% partner success rates versus 45,90% for unsuccessful orders. The theoretical contribution involves developing adaptive statistical methodology maintaining interpretability while adjusting to data characteristics. Practical applications include inventory optimization, order prioritization, and partnership development. Future directions include machine learning integration and semantic analysis of communication data.

**Keywords:** statistical analysis, order success prediction, automated resource management systems, normality tests, non-parametric methods, ERP systems, resource optimization.

## References

- [1] V. Markuts, and O. Kyzenko, «ERP system as a tool for ensuring the rational use of company resources», *Scientific notes*, vol. 32, pp. 68-78, 2023. doi: **10.33111/vz\_kneu.32.23.03.06.045.051**.
- [2] M.R.N.M. Qureshi, «Evaluating Enterprise Resource Planning (ERP) Implementation for Sustainable Supply Chain Management», *Sustainability*, vol. 14, no. 22, article 14779, 2022. doi: **10.3390/su142214779**.
- [3] M. AlMarri, M. Al-Ali, M. Alzarooni, A. AlTeneiji, K. Al-Ali, and Z. Bahroun, «Enterprise Resource Planning Systems for Health, Safety, and Environment Management: Analyzing Critical Success Factors», *Sustainability*, vol. 17, no. 7, article 2947, 2025. doi: **10.3390/su17072947**.
- [4] O.R. Amosu, P. Kumar, Y.M. Ogunsuji, S. Oni, and O. Faworaja. «AI-driven demand forecasting: Enhancing inventory management and customer satisfaction», *World Journal of Advanced Research and Reviews*, vol. 23, no. 2, pp. 708–719, 2024. doi: **10.30574/wjarr.2024.23.2.2394**.
- [5] L. Schmid, M. Roidl, A. Kirchheim, and M. Pauly, «Comparing Statistical and Machine Learning Methods for Time Series Forecasting in Data-Driven Logistics – A Simulation Study», *Entropy*, vol. 27, no. 1, article 25, 2024. doi: **10.3390/e27010025**.
- [6] Á. Cartea, S.N. Cohen, R. Graumans, S. Labyad, L. Sánchez-Betancourt, L. van Veldhuijzen, «Statistical Predictions of Trading Strategies in Electronic Markets», *Journal of Financial Econometrics*, vol. 23, iss. 2, pp. 1-64, 2024. doi: **10.1093/jjfinec/nbae025**.
- [7] A. Walter, K. Ahsan, and S. Rahman, «Application of artificial intelligence in demand planning for supply chains: a systematic literature review», *The International Journal of Logistics Management*, vol. 36, no. 3, pp. 672-719, 2025. doi: **10.1108/ijlm-02-2024-0120**.
- [8] Z. Ahmadi, «The Role of AI and Machine Learning in Supply Chain Optimization», *International Journal of Modern Achievement in Science, Engineering and Technology*, vol. 2, no. 2, pp. 1-8, 2025. doi: **10.63053/ijset.77**.
- [9] O. Piskunova, L. Ligonenko, R. Klocho, T. Frolova, and T. Bilyk, «Applying Machine Learning Approach to Start-up Success Prediction», *Scientific Horizons*, vol. 24, no. 11, pp. 72-84, 2021. doi: **10.48077/scihor.24(11).2021.72-84**.
- [10] Y. Choi, «Startup Success Prediction with PCA-Enhanced Machine Learning Models», *Journal of technology management & innovation*, vol. 19, no. 4, pp. 77-88, 2024. doi: **10.4067/s0718-27242024000400077**.
- [11] Z.J.H. Tarigan, H. Siagian, and F. Jie, «Impact of Enhanced Enterprise Resource Planning (ERP) on Firm Performance through Green Supply Chain Management», *Sustainability*, vol. 13, no. 8, article 4358, 2021. doi: **10.3390/su13084358**.
- [12] M. Al-Amin, M.T. Hossain, M.J. Islam, and S. Kumar Biwas, «History, Features, Challenges, and Critical Success Factors of Enterprise Resource Planning (ERP) in The Era of Industry 4.0», *European Scientific Journal, ESJ*, vol. 19, no. 6, pp. 31-59, 2023. doi: **10.19044/esj.2023.v19n6p31**.
- [13] V. Pylypenko, V. Statsenko, T. Bila, and D. Statsenko, «Determining the influence of data on working with video materials on the accuracy of student success prediction models», *Eastern-European Journal of Enterprise Technologies*, vol. 5, no. 4(131), pp. 52-62. doi: **10.15587/1729-4061.2024.313333**.
- [14] S. Mortensen, M. Christison, B. Li, A. Zhu, and R. Venkatesan, «Predicting and Defining B2B Sales Success with Machine Learning», in *Proceedings of the 2019 Systems and Information Engineering Design Symposium (SIEDS)*, Charlottesville, USA, 2019, pp. 352-356. doi: **10.1109/SIEDS.2019.8735638**.
- [15] S. McQuitty, «The Purposes of Multivariate Data Analysis Methods: an Applied Commentary», *Journal of African Business*, vol. 19, no. 1, pp. 124-142, 2017. doi: **10.1080/15228916.2017.1374816**.
- [16] T. Lei, R.Y.M. Li, N. Jotikastira, H. Fu, and C. Wang, «Prediction for the Inventory Management Chaotic Complexity System Based on the Deep Neural Network Algorithm», *Complexity*, vol. 2023, pp. 1-11, 2023. doi: **10.1155/2023/9369888**.
- [17] L. Oluwaseye Joel, W. Doorsamy, and B. Sena Paul, «A Review of Missing Data Handling Techniques for Machine Learning», *International Journal of Innovative Technology and Interdisciplinary Sciences*, vol. 5, no. 3, pp. 971-1005, 2022. doi: **10.1515/IJTIS.2022.5.3.971-1005**.
- [18] C. Arimie, E. Biu, and M. Ijomah, «Outlier Detection and Effects on Modeling», *Open Access Library*

- Journal*, vol. 7, pp. 1-30, 2020. doi: **10.4236/oalib.1106619**.
- [19] A.V. Shved, Y.O. Davydenko, «Outlier detection technique for heterogeneous data using trimmed-mean robust estimators», *Radio Electronics, Computer Science, Control*, no. 3, pp. 50-57, 2022. doi: **10.15588/1607-3274-2022-3-5**.
- [20] F.M. Dekking, C. Kraaikamp, H.P. Lopuhaä, and L.E. Meester, *A Modern Introduction to Probability and Statistics*. Springer London, 2005. doi: **10.1007/1-84628-168-7**.
- [21] MSCI global investable market value and growth index methodology. [Online]. Available: [https://www.msci.com/eqb/methodology/meth\\_docs/MSCI\\_GIMIVGMethod\\_Feb2021.pdf](https://www.msci.com/eqb/methodology/meth_docs/MSCI_GIMIVGMethod_Feb2021.pdf). Accessed on: January 5, 2014.
- [22] C.-E. Särndal, B. Swensson, and J. Wretman, *Model Assisted Survey Sampling (Springer Series in Statistics)*. Springer, 2003.
- [23] S.S. Shapiro, and M.B. Wilk, «An analysis of variance test for normality (complete samples)», *Biometrika*, vol. 52, no. 3-4, pp. 591-611, 1965. doi: **10.1093/biomet/52.3-4.591**.
- [24] R. D'Agostino, and E.S. Pearson, «Tests for departure from normality. Empirical results for the distributions of  $b_2$  and  $\sqrt{b_1}$ », *Biometrika*, vol. 60, no. 3, pp. 613-622, 1973. doi: **10.1093/biomet/60.3.613**.
- [25] T.W. Anderson, and D.A. Darling, «Asymptotic Theory of Certain «Goodness of Fit» Criteria Based on Stochastic Processes», *The Annals of Mathematical Statistics*, vol. 23, no. 2, pp. 193-212, 1952. doi: **10.1214/aoms/1177729437**.
- [26] B.L. Welch, «The generalization of 'student's' problem when several different population variances are involved», *Biometrika*, vol. 34, no. 1-2, pp. 28-35, 1947. doi: **10.1093/biomet/34.1-2.28**.
- [27] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. New York, USA: Routledge, 2013. doi: **10.4324/9780203771587**.
- [28] H.B. Mann, and D.R. Whitney, «On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other», *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50-60, 1947. doi: **10.1214/aoms/1177730491**.
- [29] J.A. Swets, *Signal detection theory and ROC-analysis*. New York, USA: Academic Press, 1975.
- [30] S. Ben-David, and S. Shalev-Shwartz, *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014.
- [31] K.H. Brodersen, C.S. Ong, K.E. Stephan, and J.M. Buhmann, «The Balanced Accuracy and Its Posterior Distribution», in *Proceedings of the 20th International Conference on Pattern Recognition (ICPR)*, Istanbul, Turkey, 2010, pp. 3121-3124. doi: **10.1109/icpr.2010.764**.
- [32] J.D. Kelleher, A. D'Arcy, and B.M. Namee, *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. MIT Press, 2015.
- [33] T. Fawcett, «An introduction to ROC analysis», *Pattern Recognition Letters*, vol. 27, no. 8, pp. 861-874, 2006. doi: **10.1016/j.patrec.2005.10.010**.
- [34] J. Sun, C. Tang, W. Xie, and X.-H. Zhou, «Nonparametric receiver operating characteristic curve analysis with an imperfect gold standard», *Biometrics*, vol. 80, no. 3, pp. 1-12, 2024. doi: **10.1093/biometc/ujae063**.
- [35] Ş. Çorbacıoğlu, and G. Aksel, «Receiver operating characteristic curve analysis in diagnostic accuracy studies: A guide to interpreting the area under the curve value», *Turkish Journal of Emergency Medicine*, vol. 23, no. 4, pp. 195-198, 2023. doi: **10.4103/tjem.tjem\_182\_23**.

Стаття надійшла 18.05.2025

Стаття прийнята 02.06.2025

Стаття опублікована 30.06.2025

**Цитуйте цю статтю як:** Мірошніченко С.О., Вовна О.В. Багатофакторний статистичний аналіз успішності замовлень як інструмент оптимізації автоматизованих систем керування ресурсами. *Вісник Приазовського державного технічного університету*. Серія: Технічні науки. 2025. Вип. 50. С. 182-190. DOI: <https://doi.org/10.31498/2225-6733.50.2025.336372>.