

УДК 519.23: 004.6

DOI: [https://doi.org/10.31521/modecon.V39\(2023\)-11](https://doi.org/10.31521/modecon.V39(2023)-11)

Кушнір О. К., кандидат економічних наук, доцент, доцент кафедри економіки підприємства, Кам'янець-Подільський національний університет імені Івана Огієнка, м. Кам'янець-Подільський, Україна

ORCID: 0000-0003-2679-2782

e-mail: oksana.kushnir@kpnpu.edu.ua

Чаплінський В. Р., кандидат економічних наук, старший викладач кафедри економіки підприємства, Кам'янець-Подільський національний університет імені Івана Огієнка, м. Кам'янець-Подільський, Україна

ORCID: 0000-0002-3209-1475

e-mail: chaplinskyi.vadym@kpnpu.edu.ua

Статистичні методи аналізу великих даних

Анотація. У статті проаналізовано значення великих даних з часу їх створення та розуміння бізнесом їх важливості. Обґрунтовано використання великих даних, оскільки вони допомагають виявити нові можливості та підвищити ефективність діяльності, що допоможе збільшити прибуток, завоювавши більше клієнтів. Встановлено, що наявні підходи та методи аналізу інформації уже не виконують повноцінно свої функції та стають менш актуальними, тому виникає потреба пошуку нових можливостей. Для різноманітних великих даних потрібні нові статистичні ідеї та методи, що зосереджені на адаптації стандартних статистичних моделей до великих даних, розмір яких перевищує потужність одного комп'ютера через його великий обсяг і високу швидкість. Запропоновано використання методів статистичного аналізу, що передбачає збір, упорядкування, аналіз даних на основі встановлених принципів для виявлення закономірностей і тенденцій та виконує кілька функцій: прогнозування, моделювання, створення моделей, зниження ризику та виявлення тенденцій. Статистичний аналіз дозволить отримати висновки з даних. Розглянуто методи статистичного аналізу, які застосовуються для аналізу доказів своїх гіпотез, щоб допомогти бізнесу робити прогнози та приймати рішення щодо їхніх продуктів і послуг; для отримання точної інформації з числових даних. Статистичні методи використовуються для виявлення закономірностей і кореляцій, створених аналізом даних, і намагаються підтвердити їх за допомогою строгих наукових методологій. Доведено, що використання статистичних методів для аналізу великих даних потребує розробки принципово нової політики з управління інформацією, її захисту, стосунків із респондентами, процесу підготовки фахівців.

Ключові слова: великі дані; аналіз великих даних; статистичні методи; регресія; середнє арифметичне; вибірка; гіпотеза.

Kushnir O. K., Candidate of Economic Sciences, Associate Professor, Associate Professor of the Department of Economics of Enterprise, Kamianets-Podilskyi Ivan Ohienko National University, Kamianets-Podilskyi, Ukraine

Chaplinsky V. R., Candidate of Economic Sciences, Senior Lecturer of the Department of Economics of Enterprise, Kamianets-Podilskyi Ivan Ohienko National University, Kamianets-Podilskyi, Ukraine

Statistical Methods for Big Data Analysis

Abstract. Introduction. Big data has been in the spotlight since its inception as organizations understand its importance and use it in their business. The adoption of big data helps businesses discover new opportunities and improve efficiency, which helps increase profits by attracting more customers. In today's world, the concept of big data is considered the most important for the following reasons: Cost reduction – big data technologies are more cost-effective. And it is the best tool for storing large amounts of data at a lower cost; Fast decision making. With in-memory analytics and the ability to analyze new data sources, big data helps companies analyze data and information faster than ever before.

Purpose. Analyze the features of statistical methods of big data analysis with the growing importance of the features of big data presentation technologies.

Results. It has been established that the existing approaches and methods of information analysis no longer fully fulfill their functions and are becoming less relevant, therefore there is a need to find new opportunities. The diversity of big data requires new statistical ideas and methods that focus on adapting standard statistical models to big data, the size of which exceeds the capacity of a single computer due to its large volume and high speed. The use of statistical analysis methods is proposed, which involve the collection, organization and analysis of data based on established principles to identify patterns and trends and perform several functions: forecasting, modeling, creating models, reducing risk and identifying trends. Statistical analysis will allow you to draw conclusions from the data. Learns the statistical analysis techniques used to analyze the evidence for one's hypotheses to help businesses make predictions and make decisions about their products and services; get accurate information from numerical data. Statistical methods are used to identify patterns and correlations created by data analysis and attempt to confirm using rigorous scientific methodologies.

¹Стаття надійшла до редакції: 30.04.2023

Received: 30 April 2023

Conclusions. It has been proven that the use of statistical methods for the analysis of big data requires the development of a fundamentally new policy regarding information management, its protection, relations with respondents, and the process of training specialists.

Keywords: big data; big data analysis; statistical methods; regression; mean; sample; hypothesis.

JEL Classification: C10.

Постановка проблеми. Статистичний аналіз являється потужним інструментом, який необхідно використовувати для розуміння великих даних і прийняття рішень. Існують різні типи методів статистичного аналізу, які можна застосовувати до широкого діапазону даних, галузей і програм. Знання різних методів статистичного аналізу та способів їх використання може допомогти досліджувати дані, знаходити закономірності та виявляти тенденції ринку.

Аналіз останніх досліджень та публікацій. Сучасні методи аналізу даних були розвинені у працях Ю. П. Адлера, С. А. Айвазяна, В. Є. Бахрушина, Б. В. Гнеденко, А. М. Дуброва, І. А. Ібрагімова, А. Г. Івахненка, Дж. Кіфера, К. Х. Крамера, Б. Ю. Лемешка, Ю. В. Лінника, Г. В. Мартинова, В. В. Налімова, І. М. Парасюка, Ю. В. Прохорова, С. Р. Рао, А. Стьюарта, Дж. Тьюкі, А. Хьютсона та багатьох інших дослідників [1]. Оскільки останнім часом значного поширення набувають нові технології й методи аналізу великих даних, зокрема методи статистичного аналізу даних, а також методики й засоби статистичного контролю за якістю на виробництві та в управлінні організаціями, тому тема залишається актуальною.

Формулювання цілей дослідження. Метою дослідження є вивчення особливостей статистичних методів аналізу великих даних з урахуванням означення та особливостей застосування технологій великих даних.

Основні результати дослідження. Великі дані були в центрі уваги з часу їх створення, оскільки організації

розуміють їх важливість і використовують для свого бізнесу. Введення великих даних допомагає бізнесу виявити нові можливості та підвищити їх ефективність, що допоможе збільшити їх прибуток, завоювавши багато клієнтів. У сучасному світі поняття великих даних вважаються найважливішими з таких причин: зниження витрат – великі технології передачі даних є більш економічні. І це найліпший інструмент для зберігання величезних даних з меншими витратами; швидке прийняття рішень – за допомогою аналітики в пам'яті та можливості аналізувати нові джерела даних Big data допомагають бізнесу швидше аналізувати дані та інформацію, ніж раніше. На основі навчання за допомогою аналізу бізнес може прийняти розумне рішення; нові продукти та функції – завдяки правильній аналітиці великі концепції даних знають потреби та задоволення клієнтів. Тому вони завжди доставляють те, що хочуть клієнти. Деякі компанії також створюють нові продукти, використовуючи аналітику великих даних, щоб задовольнити своїх клієнтів [2, с. 371].

Внесок у розвиток великих даних зробила активна цифровізація документів, книг, фільмів. Ще одне гігантське джерело даних – пристрої та датчики IoT (Інтернету речей). IoT і машинне навчання – ключові драйвери зростання глобального ринку Big Data, який, згідно з прогнозом Statista/Wikibon, до 2027 року складе \$103 млрд (рис. 1).

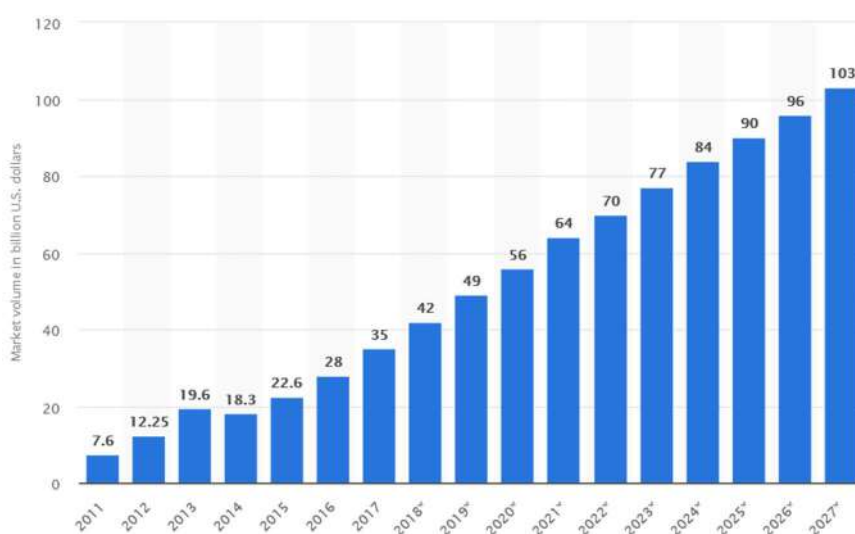


Рисунок 1 – Прогноз обсягу світового ринку великих даних 2011-2027 рр. (у млрд \$)

Джерело: побудовано авторами на основі [3]

Кількість внутрішньокорпоративних великомасштабних джерел даних істотно зростає: великі бази даних сьогодні виникають навіть на основі єдиного джерела потоків даних про відвідування Web-сайтів (click-stream), журналів програмних систем, архівів електронної пошти та форумів тощо. Для збирання розрізаних великомасштабних даних доцільним є підхід, який називають могутнім аналізом даних (MAD) [4, с. 85].

Наявні підходи та методи аналізу інформації уже не виконують повноцінно свої функції та стають менш актуальними, виникає потреба пошуку нових можливостей. Найкраще з такими викликами справляються методики Big Data Analytics. Big Data Analytics – це комплекс методик та підходів, що направлені на акумулювання, систематизацію та обробку великої кількості різної за своїми характеристиками інформації та формуванні на їх основі відповідних висновків та гіпотез. Відповідно функціональні зв'язки Big Data Analytics є досить розгалуженим та охоплюють такі елементи: технології візуалізації, статистичний аналіз, штучний інтелект, технології баз даних, технології розпізнавання образів,

об'єднавши які та класифікувавши можна отримати перелік методів аналітики великих даних [5].

Для різноманітних великих даних потрібні нові статистичні ідеї та методи, що зосереджені на адаптації стандартних статистичних моделей до великих даних, розмір яких перевищує потужність одного комп'ютера через його великий обсяг і високу швидкість. Статистичний аналіз передбачає збір, упорядкування, аналіз даних на основі встановлених принципів для виявлення закономірностей і тенденцій та виконує кілька функцій: прогнозування, моделювання, створення моделей, зниження ризику та виявлення тенденцій. Завдяки вдосконаленню технологій багато організацій мають величезні обсяги даних про кожен аспект своєї діяльності та ринків, що зміцнює їхню конкурентну перевагу, підвищує ефективність і оптимізує ресурси для максимальної віддачі інвестицій. Статистичний аналіз є науковим інструментом штучного інтелекту та машинного навчання [6-8].

Процес статистичного аналізу включає 5 етапів (рис. 2).



Рисунок 2 – Етапи статистичного аналізу

Джерело: побудовано авторами

Статистичний аналіз це техніка для отримання висновків з даних. Методи статистичного аналізу застосовуються для аналізу доказів своїх гіпотез, щоб допомогти бізнесу робити прогнози та приймати рішення щодо їхніх продуктів і послуг; для отримання точної інформації з числових даних. Статистичні

методи використовуються для виявлення закономірностей і кореляцій, створених аналізом даних, і намагаються підтвердити їх за допомогою строгих наукових методологій. Статистичний аналіз використовується як життєво важливий контроль якості результатів аналізу даних. Він існує ще до того,

як існують теорії, які потрібно (не)підтвердити. Дані можуть бути попередньо оброблені та проіндексовані фахівцями з обробки даних, але сховища даних часто все ще занадто великі або складні, щоб дослідники

могли ними керувати. Таким чином, статисти можуть допомогти з кваліфікацією або узагальненням наявних даних, відповівши на запитання: Які викиди слід відкинути, а які врахувати? (рис. 3).

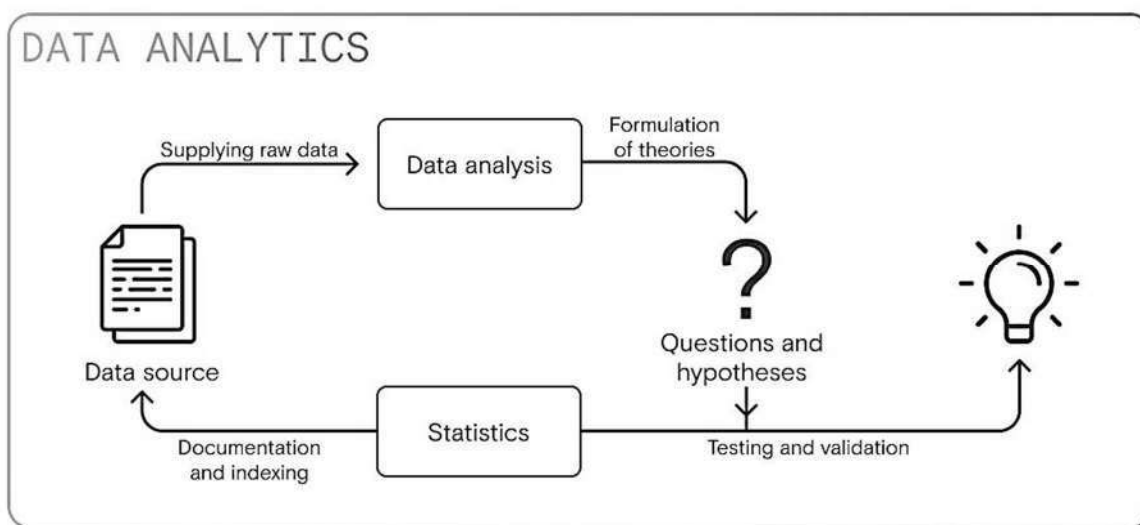


Рисунок 3 – Роль статистичного аналізу великих даних

Джерело: побудовано авторами на основі [9]

Існують різні статистичні методи, які можна використовувати для аналізу зібраних даних. До найбільш популярних і поширених методів статистичного аналізу відносяться [7]:

Метод середньої величини. Один із найпопулярніших методів статистичного аналізу в багатьох галузях, що використовується для набору даних з рівномірним розподілом або з розподілом у формі вигнутої кривої. Характерною особливістю цього розподілу є відсутність екстремальних значень. Розраховується за формулою:

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n}, \quad (1)$$

де x_i – окремі варіанти ознаки X ($i = 1, 2, 3, \dots, n$); n – кількість варіант.

Роль середніх величин полягає в узагальненні, тобто заміні безлічі індивідуальних значень ознаки середньої величини, що характеризує всю сукупність явищ.

Метод регресії. Статистичний метод для аналізу зв'язків між змінними. Ідея регресії полягає у вивченні зв'язку між причиною та наслідком. Існує багато типів моделей регресії, але однією з найбільш часто використовуваних є лінійна регресія. Регресійний аналіз використовується в тому випадку, якщо відношення між змінними можуть бути виражені кількісно у виді деякої комбінації цих змінних. Отримана комбінація використовується для передбачення значення, що може приймати цільова (залежна) змінна, яка обчислюється на заданому наборі значень вхідних (незалежних) змінних. Функція $f = (x_1, x_2, x_3, \dots, x_k)$, що описує залежність умовного середнього значення результативної ознаки у від

заданих значень аргументів, називається функцією (рівнянням) регресії. Для точного опису рівняння регресії необхідно знати умовний закон розподілу результативного показника y . Таку інформацію отримати зазвичай не вдається, тому обмежуються пошуком відповідних апроксимацій для функції $f = (x_1, x_2, x_3, \dots, x_k)$, заснованих на вихідних статистичних даних. У рамках окремих модельних припущень про тип розподілу вектора показників $(y, x_1, x_2, x_3, \dots, x_k)$ може бути отриманий загальний вигляд рівняння регресії:

$$f(x) = M\left(\frac{y}{x}\right)x = (x_1, x_2, x_3, \dots, x_k). \quad (2)$$

Мета використання регресії: визначення ступеня детермінованості варіації критеріальної (залежної) змінної предикторами (незалежними змінними); пророкування значення залежної змінної за допомогою незалежної; визначення внеску окремих незалежних змінних у варіацію залежної.

Регресію не можна використовувати для визначення наявності зв'язку між змінними, оскільки наявність такого зв'язку і є передумова для її застосування.

Перевірка гіпотез. Застосовується для безперервних змінних шляхом прийняття рішення щодо ймовірності того, що параметр генеральної сукупності дорівнює заданому значенню або не дорівнює цьому заданому значенню. Висунута гіпотеза може бути вірною або невірною, тому виникає необхідність її перевірити. Перевірка гіпотези здійснюється за даними вибірки (n випадкових спостережень). Для цього використовують спеціально підібрану випадкову величину K (статистичний

критерій), точний або наближений розподіл якої відомий. Значення критерію, обчислене за даними вибірки, називають його спостережуваним (емпіричним) значенням і позначають $K_{сп}$. Існує багато статистичних критеріїв. Після вибору певного критерію, множину всіх його можливих значень розбивають на дві підмножини, що не перетинаються: одна з них містить значення критерію, при яких основна гіпотеза відхиляється, а друга – при яких вона приймається. Сукупність значень критерію, при яких основна гіпотеза відхиляється, називають критичною областю, а сукупність значень, при яких гіпотезу приймають, – областю прийняття гіпотези або областю допустимих значень. Основний принцип статистичної перевірки статистичних гіпотез полягає в наступному: якщо спостережуване значення критерію належить критичній області, то основну гіпотезу відхиляють; якщо спостережуване значення критерію належить області допустимих значень – основну гіпотезу приймають.

Довірчі інтервали. Застосовується для інтервальних змінних, надаючи інтервал із заданим рівнем довіри, наприклад 95%, 99% або 100%. Оскільки статистична оцінка невідомого параметра теоретичного розподілу є функцією від результатів спостережень, то поряд з точковою оцінкою θ^* необхідно також розглядати інтервальну оцінку даного параметру, яка характеризується двома числами – початком і кінцем інтервалу, до якого цей параметр θ належатиме з певною надійністю. При оцінюванні параметра одним числом θ^* отримується певне наближення до його істинного значення θ , проте отриманий результат не обов'язково буде збігатися з відповідним параметром генеральної сукупності. Виникає необхідність визначити точність, з якою здійснюється оцінювання, і його надійність. Тобто необхідно знати, до яких помилок може призвести заміна параметра θ його точковою оцінкою θ^* і з якою впевненістю можна стверджувати, що ці помилки не виходять за означені межі. Відповіддю на ці питання є побудова довірчого інтервалу, до якого із заздалегідь визначеною надійністю належатиме значення параметра θ , що оцінюється. Зазвичай, значення довірчої ймовірності обумовлюється заздалегідь і відповідно до цього значення обчислюються границі довірчого інтервалу. Загальноприйнятими є два рівні довірчої ймовірності: 0,95 та 0,99.

Визначення розміру вибірки. Визначення кількості спостережень або повторень для включення в статистичну вибірку. При визначенні розміру вибірки враховуються два важливі моменти: бажаний рівень точності та необхідний рівень точності. Перевага вибірки полягає в тому, що вона забезпечує, у порівнянні з суцільним спостереженням, економію трудових, фінансових ресурсів та часу. Найважливішою умовою при формуванні вибірки є забезпечення її репрезентативності, тобто при будь-якому способі

відбору елементів із генеральної сукупності одержана вибірка повинна якомога точніше її представляти. Якщо є обсяг генеральної сукупності N , гранична похибка Δx або Δp , імовірність граничної похибки t , %, дисперсія сукупності σ^2 ; то оптимальний розмір вибірки, що буде складатись з n елементів можна знаходити за формулами:

для повторного добору:

$$n = \frac{t^2 \sigma^2}{\Delta^2}, \quad (3)$$

$\Delta = \Delta x$ або $\Delta = \Delta p$;

для безповторного добору:

$$n' = \frac{n}{1 + \frac{n}{N}}, \quad (4)$$

з урахуванням (3) та перетворень отримаємо:

$$n' = \frac{t^2 \sigma^2 N}{\Delta^2 N + t^2 \sigma^2}. \quad (5)$$

Використання статистичних методів для аналізу великих даних потребує розробки принципово нової політики з управління інформацією, її захисту, стосунків із респондентами, процесу підготовки фахівців. В частині методології проблемою використання великих даних є забезпечення їх репрезентативності. За відсутності формалізованої генеральної сукупності як такої, виникають складнощі у визначенні як цільової сукупності загалом, так і вибіркової сукупності. Як відомо, традиційні статистичні спостереження базуються на переписах та реєстрах, на основі яких можна визначити усі типи сукупностей, необхідні для реалізації статистичного спостереження. Водночас великі дані формуються спонтанно й поза межами реєстрів, їхня хаотичність, неструктурованість та динамічність ускладнює статистичні процеси. При цьому, наявні статистичні методи розраховані на послідовний, глибокий та тривалий аналіз даних невеликих за розміром вибірок і це, як відомо, значно гальмує статистичний виробничий процес. Ще однією методологічною проблемою використання великих даних у статистичних цілях є забезпечення відповідності технічного арсеналу статистичних досліджень у частині вимірювання якості даних, загальної обмеженості сфери застосування даних із зовнішніх джерел, складності інтегрування інформації різного зовнішнього походження у статистичні бази даних для отримання якісного кінцевого продукту. Існують також юридичні проблеми, пов'язані з унормуванням доступу та використанням великих даних, перш за все із захистом персональних даних громадян та індивідуальних даних підприємств; фінансові проблеми щодо забезпечення оптимізації співвідношення витрат і отриманих переваг; управлінські проблеми, які полягають у розробці відповідної політики та принципів управління й захисту даних [10, с. 27].

Шляхи подолання проблем використання великих даних для статистичного аналізу представлені у таблиці 1.

Таблиця 1 Шляхи подолання проблем використання великих даних для статистичного аналізу

| Проблеми | Шляхи подолання |
|---|--|
| Перевантаження великими даними | Залучення фахівців певної кваліфікації, здатних впоратися з проблемами, які виникають з появою різноманітних джерел великих даних, що дозволить уникнути затримання розвитку. |
| Витрати по збору великих даних зростають швидше, ніж можливості суб'єктів щодо їх використання | Забезпечення відповідних темпів розвитку. Нема потреби братися за все відразу і прагнути збирати 100% інформації, що надходить з кожного нового джерела великих даних. Збір і вивчення зразків нових даних. За їх допомогою можна провести експериментальний аналіз, щоб визначити, що дійсно важливо в кожному джерелі та як кожен з них може бути використаний. |
| Ризик порушення конфіденційності, тобто забезпечення довіри громадськості та отримання її згоди на багаторазове використання даних і їх зв'язку із іншими джерелами | Запровадження правил обробки, зберігання та застосування великих даних, які повинні розвиватися разом з аналітичними можливостями. З розвитком сфери великих даних повинні розвиватися сфери самостійного і правового регулювання їх використання. Вжити заходи, щоб не допустити порушення конфіденційності. Організації повинні чітко пояснити, як вони будуть забезпечувати безпеку даних і як будуть їх використовувати, якщо хочуть отримати дозвіл користувачів на їх збір і аналіз. Розв'язання проблеми конфіденційності, пов'язаної з великими даними, інакше їх потенціал неможливо реалізувати повністю. Без належного обмеження великі дані можуть підняти таку хвилю протесту, що деякі їх джерела будуть повністю закриті. Критично важлива наявність саморегулювання. Учасники ринку повинні забезпечити саморегулювання і розробити правила, яких має дотримуватися кожен. Такі правила зазвичай більш ефективні та менш жорсткі, ніж ті, які вводяться державними органами, коли галузь не може контролювати себе самостійно. |
| Законодавчі | Розроблення системи нормативних документів правового регулювання використання великих даних, які на законодавчому рівні регламентуватимуть правила доступу до даних і їх використання. |
| Кадрові | Розроблення методів та інструментів для оцінки базових потреб у навичках використання великих даних. Інструктування з питань розробки модульної програми навчання спеціалістів. Координація навчальних програм, що стосуються більш поглибленої підготовки фахівців. Сприяння створенню глобальної мережі навчальних і дослідницьких установ для навчання і зміцнення потенціалу в галузі використання великих даних при підготовці статистичної інформації. |
| Управлінські | Формування національної політики й директив у галузі управління даними та їх захисту. Розроблення стратегічного та оперативного плану дій у галузі управління великими даними та захисту пропріетарних даних. |
| Методологічні | Підготовка методологічних матеріалів і проведення семінарів з питань забезпечення якості даних і застосування статистичних методів аналізу великих даних. Організація постійних консультативних центрів на базі навчально-наукових інститутів та закладів освіти. |
| Технологічні | Впровадження інноваційних рішень та розв'язання питань, пов'язаних з інформаційними технологіями. |

Джерело: складено авторами на основі [11, с. 360]

Висновки. Оскільки бізнес-середовище переживає період глобальної трансформації, експерти говорять не лише про потенційні можливості та результати, яких можна досягти завдяки використанню технологій великих даних, а й про реальні переваги, такі як збільшення прибутку, розширення бази клієнтів та підвищення операційної ефективності. Впровадження великих даних загалом змінює звичне IT-середовище компанії, дозволяючи автоматизувати деякі бізнес-процеси та підвищити ефективність роботи

співробітників компанії. Нові можливості дозволяють збирати статистику в режимі реального часу (онлайн), що дозволяє компаніям швидше реагувати на зміни ринку. Водночас технологія великих даних дає більш детальний і водночас надійний огляд загроз безпеці, а також використовує весь масив даних про діяльність компанії – від стану обладнання до аналізу мережевого трафіку – для швидкого та ефективного реагування на атаки та загрози в кіберпросторі.

Література:

1. Бахрушин В. Є. Методи аналізу даних : навчальний посібник для студентів. Запоріжжя : КПУ, 2011. 268 с.
2. Цифрова трансформація економіки : мікро- та макроаспекти : колективна монографія / за заг. ред. Н. А. Мазур, д.е.н., проф.; Кам'янець-Подільськ. нац. ун-т ім. І. Огієнка. Чернівці : Чернівець. нац. ун-т ім. Ю. Федьковича, 2022. 440 с.
3. Big data market size revenue forecast worldwide from 2011 to 2027. URL : <https://www.statista.com/statistics/254266/global-big-data-market-forecast/> (дата звернення : 10.03.2023 р.).

4. Верес О. М., Оливко Р. М. Класифікація методів аналізу Великих даних. *Вісник Національного університету «Львівська політехніка». Серія : Інформаційні системи та мережі.* 2017. № 872. С. 84-92.
5. Чаплінський В. Р., Кушнір О. К., Свідер О. П. Аналіз великих даних та їх візуалізація для потреб бізнесу. *Ефективна економіка.* 2021. № 6. URL : <http://www.economy.nayka.com.ua/?op=1&z=8979> (дата звернення : 28.02.2023).
6. Wang C., Chen M. H., Schifano E., Wu J., Yan J. (2016) Statistical methods and computing for big data. *Stat Interface.* 9(4): 399-414. <http://doi.org/10.4310/SII.2016.v9.n4.a1> (дата звернення : 23.03.2023).
7. Kampakis, S. (2020). Demystifying Data Science and All the Other Buzzwords. In: *The Decision Maker's Handbook to Data Science.* Apress, Berkeley, CA. https://doi.org/10.1007/978-1-4842-5494-3_1 (дата звернення : 23.03.2023).
8. Beyer M., Laney D. (2012). The Importance of Big Data: A Definition. Gartner Inc. Electronic data. Stamford : *Gartner.* URL : <https://www.gartner.com/en/documents/2057415> (дата звернення : 25.03.2023).
9. The Difference Between Data Analytics and Statistics. <https://www.rudderstack.com/learn/data-analytics/the-difference-between-data-analytics-and-statistics/> (дата звернення : 25.03.2023).
10. Осауленко О. Г., Горобець О. О. Проблеми та перспективи імплементації великих даних в офіційну статистику України в умовах воєнного стану. *Сучасна статистика : проблеми та перспективи розвитку : матеріали XX Міжнар. наук.-практ. конф. К. : «Інформаційно-аналітичне агентство», 2022. С. 26-31.*
11. Корепанов О. С. Використання «великих даних» для статистичного забезпечення управління розвитком «розумних» сталих міст. *БізнесІнформ.* 2018. № 6. С. 356-361.

References:

1. Bakhrushyn, V. E. (2011). *Metody analizu danykh : navchal'nyj posibnyk dlia studentiv.* Zaporizhzhia : KPU [in Ukrainian].
2. *Cyfrova transformacija ekonomiky : mikro- ta makroaspekty : kolektyvna monografija.* (2022). Chernivtsi : Chernivec. nac. un-t im. Ju. Fedjkovycha [in Ukrainian].
3. Big data market size revenue forecast worldwide from 2011 to 2027. Retrieved from <https://www.statista.com/statistics/254266/global-big-data-market-forecast/>.
4. Veres, O. M., Olyvko, R. M. (2017). Classification of big data analysis methods. *Visnyk Natsional'noho universytetu «L'viv's'ka politekhnika»*, 872, 84-92 [in Ukrainian].
5. Chaplins'kyj, V. R., Kushnir, O. K., Svider, O. P. (2021). Analysis of big data and its visualization for business needs. *Efektivna ekonomika*, 6. Retrieved from <http://www.economy.nayka.com.ua/?op=1&z=8979> [in Ukrainian].
6. Wang, C., Chen, M. H., Schifano, E., Wu, J., Yan, J. (2016). Statistical methods and computing for big data. *Stat Interface*, 9(4): 399-414, doi: 10.4310/SII.2016.v9.n4.a1.
7. Kampakis, S. (2020). *The Decision Maker's Handbook to Data Science: A Guide for Non-Technical Executives, Managers, and Founders.* London, UK. Retrieved from <https://doi.org/10.1007/978-1-4842-5494-3>.
8. Beyer, M., Laney, D. (2012). The Importance of Big Data: A Definition. Gartner Inc. Electronic data. Stamford : Gartner, Retrieved from URL : <https://www.gartner.com/en/documents/2057415>.
9. The Difference Between Data Analytics and Statistics. Retrieved from <https://www.rudderstack.com/learn/data-analytics/the-difference-between-data-analytics-and-statistics/>.
10. Osaulenko, O. H., Horobets', O. O. (2022). Problems and prospects of implementation of big data in the official statistics of Ukraine in the conditions of martial law. *Suchasna statystyka : problemy ta perspektyvy rozvytku : materialy XX Mizhnar. nauk.-prakt. konf. Kyiv : «Informacijno-analityczne aghentstvo»*, 26-31 [in Ukrainian].
11. Korepanov, O. S. (2018). Using «big data» for statistical management of the development of «smart» sustainable cities. *BiznesInform*, 6, 356-361 [in Ukrainian].

