



KAPITEL 2 / CHAPTER 2²

BIG DATA AS A SOURCE OF INFORMATION SUPPORT FOR MODERN ENTERPRISES

DOI: 10.30890/2709-2313.2024-29-00-022

Вступ

Труднощі воєнного стану породили для багатьох підприємств проблеми недостатності ресурсів, як фінансових, так і людських, що визвало необхідність пошуку інноваційних рішень, спрямованих на недопущення збільшення впливу дестабілізуючих чинників та подальшого зростання розриву з країнами, що лідирують у сфері нових технологій.

Виникла потреба у оцінці та підсиленні чинників, що дозволяють здійснювати ефективну діяльність, одним із яких є інформаційне забезпечення, яке повинно представляти відлагоджений механізм збору, оброблення, збереження та передання інформації. Вартість достовірної та вчасно наданої інформації є неоціненною, її наявність – умова задоволення потреб підприємств, підвищення їх конкурентоспроможності за рахунок прийняття обґрунтованих управлінських рішень.

Якісна інформація повинна відповідати вимогам, що відображає особливості різних етапів життєвого циклу підприємства; відповідно має значення вибір пріоритетних джерел інформації для вирішення тих завдань, які властиві конкретному етапу.

Інформаційна підтримка діяльності підприємства передбачає:

- визначення необхідності у інформаційному забезпеченні в залежності від етапу життєвого циклу та потреб учасників, що приймають участь у реалізації цілей цього етапу;
- аналіз джерел інформації, які є доступними для підприємства, оцінка ступеня їх доступності та можливості одержання вичерпної інформації;
- вибір типу необхідної інформації (первинна чи вторинна) з урахуванням переваг та недоліків цих типів інформації та фінансових можливостей

²Authors: Gavrillo Tetiana



підприємства;

- оцінка способів збору інформації – зусиллями власних працівників чи з допомогою структур, що надають послуги по наданню необхідної інформації;
- визначення методів та інструментів обробки інформації, що дозволяють одержати бажані результати незалежно від рівня структурованості даних (структуровані чи неструктуровані).

Потужним ресурсом підвищення якості інформаційного забезпечення підприємств є великі дані (*Big Data*), які дають можливість на новому рівні оцінити явища та процеси в різних сферах суспільного життя.

Ряд науковців проявляють інтерес до розуміння сучасних тенденцій розвитку *Big Data* з огляду на особливості сучасної епохи цифровізації та потреб підприємств у трансформації власних бізнес-моделей. Клоповим І.О., Шапуровим О.О. досліджена сутність *Big Data* та здійснена класифікація інформації, що належить до *Big Data* [1]. Чаплінський В.Р., Кушнір О.К., Свідер О.П. аналізують проблеми неструктуризованих великих даних та сучасні технології графічної інтерпретації результатів аналізу *Big Data* [2].

Балабановим О.С. здійснене дослідження основних напрямків, задач та типів результатів глибокого аналізу комп'ютеризованих *Big Data* [3]. Вакшинська Н.Ю., Шандрівська О.Є. оцінюють стан та майбутні тенденції розвитку світового та вітчизняного ринків *Big Data* під час війни та післявоєнний період [4].

Незважаючи на наявність достатньої кількості публікацій, феномен *Big Data* не досліджений в повній мірі та потребує поглибленої оцінки з урахуванням сучасних процесів цифровізації та підвищених вимог до платформ управління великими даними.

2.1. Сутність та характеристики великих даних

Великі дані на сьогодні є невід'ємною умовою розвитку бізнесу. Вивчення



великих даних дає можливість підприємствам оптимізувати бізнес-процеси, знайти більш ефективні способи організувати виробництво, досягти кращого рівня фінансової результативності. Оброблення великих даних дозволяє одержати інформацію про слабкі місця в діяльності підприємств та визначити причини їх виникнення, а також зрозуміти умови розвитку та зростання.

Використання великих даних наближує підприємства до споживачів та сприяє підвищенню задоволення їх потреб за рахунок персоналізації продуктів та послуг на основі визначення демографічних параметрів, уподобань клієнтів, особливостей їх поведінки.

Появу терміну *Big Data* відносять до початку 1990-х років і пов'язують з особистістю Джона Р. Меші, який популяризував цей термін і завдяки цьому увійшов у історію, як «батько великих даних». Існує також думка, що першість щодо використання поняття *Big Data* належить Дугласу Лейні (1997 рік). Також в деяких джерелах приводиться ім'я Роджера Мугаласа та компанії *O'Reilly Media*, як авторів ідеї *Big Data* (2005 рік).

Як вважають фахівці, великі дані представляють собою масив, що включає структуровану і неструктуровану інформацію, а також комплекс методів, підходів та інструментів, що забезпечують оброблення та зберігання даних. Аналітична компанія *Gartner* визначає великі дані як інформаційні активи, що відрізняються великим обсягом, високою швидкістю та (або) великою різноманітністю, що призводить до формування потреби у нових формах обробки з метою збільшення ефективності прийнятих рішень, генерації ідей та удосконалення процесів.

Аналітичною фірмою *IDC* викладений підхід до великих даних, як до нового покоління технологій і архітектур, що призначені для одержання інформації, що має корисну значущість, використовуючи велику кількість різноманітних даних. За розумінням *Forrester Research*, до *Big Data* відносяться структуровані і неструктуровані дані такого масивного об'єму, що стає складним використання традиційних баз даних та програмних засобів для їх обробки.

Компанією *Gartner* у 2015 році *Big Data* перестали розглядатися в числі



популярних трендів, так як, на думку аналітиків компанії, вони містять у своєму складі велику кількість технологій, які використовуються підприємствами, в певній мірі мають відношення до інших актуальних сфер, тому уже перетворилися в повсякденний робочий інструмент. Однак, це не означає, що *Big Data* не розглядається як драйвер розвитку інформаційного середовища, що забезпечує умови сталого функціонування для підприємств.

Підприємства при розробленні та реалізації інноваційних проєктів відчують потребу у якісному інформаційному забезпеченні, і це пов'язано з необхідністю використання величезних обсягів даних, з яких підприємства прагнуть одержати цінну інформацію для успішного досягнення цілей.

Первинно характеристика *Big Data* була надана Кліффордом Лінчем в статті «Як зростають ваші дані?», опублікованій в спеціальному випуску журналу *Nature* за 2008 рік. До основних ознак *Big Data* Лінч К. відносить наступні: *volume* (обсяг), *velocity* (швидкість), *variety* (різноманітність).

Ця модель одержала назву *3V*. Ознака «обсяг» передбачає, що вимірювання великих даних здійснюється у терабайтах, петабайтах та ексабайтах. Наприклад, аналітиками компанії IBS величина світового об'єму даних була оцінена в 2003 році на рівні 5 ексабайтів, в 2008 році – 0,18 зетабайта, в 2015 році – більше 6,5 зетабайтів, для 2020 року прогностичні дані склали 40-44 зетабайтів, а для 2025 року ця величина повинна збільшитися у 10 разів.

«Швидкість» означає, що продукування та генерування відбувається з великою швидкістю та потребує відповідного темпу оброблення. Важливим є застосування машинного навчання та штучного інтелекту, які забезпечують автоматичне знаходження певних закономірностей в зібраних даних та застосовують їх для трансформації в знання. Під ознакою «різноманітність» розуміється, що надходження даних може відбуватися із різних джерел, у різноманітних форматах та формах.

У наступних моделях (*4V*, *5V*, *7V*) з'явилися нові ознаки, що дозволили поглибити розуміння сутності великих даних і привести його у відповідність до сучасного стану інформаційного простору та потреб бізнесу. Модель *5V* включає



додаткові ознаки – «*veracity*» (правдивість) та «*value*» (цінність). Під правдивістю розуміється рівень якості та достовірності даних; недостовірні дані можуть привести до хибного трактування явищ і процесів і, як наслідку, прийняття неефективних інноваційних рішень.

Корисність даних означає те, що із них можна одержати цінну інформацію, що відповідає нагальним потребам підприємства. При одержанні цінної інформації із загальних масивів інформації повинен застосовуватися інструментарій, якому властива унікальність, хоча при збиранні та аналізі великих даних підприємства можуть користуватися стандартними засобами.

До моделі *7V* включені, порівняно з моделлю *5V*, дві ознаки – «*variability*» (мінливість сприйняття) та «*visualization*» (візуалізація). Мінливість сприйняття даних зумовлює варіабельність застосовуваних алгоритмів для забезпечення достовірного відображення сутності даних в процесі їх розшифрування. Візуалізація забезпечує можливість краще сприймати дані на основі їх графічного представлення, яке не повинно носити спрощений характер, а враховувати достатню кількість змінних для здійснення ефективного аналізу.

У сучасну епоху цифровізації великі дані стали надзвичайно важливим активом для підприємств. Однак, попри їх незаперечні переваги, їм властиві і вразливі місця, про які треба завжди пам'ятати (таблиця 1).

До основних принципів *Big Data* відносяться:

- адаптованість інструментарію оброблення *Big Data* до збільшення об'єму інформаційних масивів;
- надійність використовуваного інструментарію за рахунок застосування сучасних технологій, що мінімізують можливість впливу збійних ситуацій на функціонування системи в цілому;
- локалізація даних, що передбачає акумулювання операцій оброблення та зберігання даних в межах однієї обчислювальної потужності.

Викладені принципи повинні бути врахованими на всіх етапах роботи з великими даними – збору, оброблення, очищення, аналізу. Збір даних може здійснюватися з використанням різноманітних джерел – соціальних мереж,



медіа-ресурсів, загальнодоступних веб-сайтів, архівів, баз та сховищ даних, хмарних сховищ, даних датчиків.

Таблиця 1 – Переваги та недоліки *Big Data*

Певаги <i>Big Data</i>	Недоліки <i>Big Data</i>
Створюють умови для оптимізації бізнес-процесів на підприємстві за рахунок аналізу продуктивності праці працівників та стану виробництва	Можливість виникнення загрози для безпеки персональних даних та порушення норм приватності
Сприяють удосконаленню процесів управління ресурсами	Неграмотний вибір та застосування інструментарію збору та оброблення великих даних може призвести до хибних заключень
Розширюють можливості для взаємодії зі споживачами, більш якісно задовольняючи їх потреби за рахунок адаптації продуктів і послуг, та забезпечують формування нової цільової аудиторії	Упровадження технології <i>Big Data</i> та створення відповідної інфраструктури передбачає суттєві витрати
Дозволяють пришвидшити та збільшити дієвість управлінських рішень за рахунок оперування фактами, а не на основі припущень	Імовірність порушення в результаті використання даних конфіденційності інформації стосовно конкретних осіб
Слугують умовою для удосконалення інноваційної діяльності на підприємстві, що стає визначальним чинником підвищення конкурентоспроможності	Загроза, у випадку надмірних обсягів даних, одержання необ'єктивних результатів в результаті оброблення та аналізу
Забезпечують можливість виявляти та прогнозувати майбутні ринкові тенденції та на основі цього розробляти стратегію розвитку підприємства	Необхідність адаптації системи великих даних до особливостей підприємства, що потребує гармонізації використовуваного інструментарію
Допомагають вчасно ідентифікувати ризики для підприємства, що дає можливість застосовувати заходи щодо їх попередження та мінімізації впливу у випадку настання ризикової ситуації	Робота з <i>Big Data</i> потребує наявності у IT-менеджерів нових навичок, які суттєво відрізняються від роботи зі звичайними масивами даних

Джерело: складено автором

На етапі оброблення даних вирішальним є вибір способу оброблення – потокового чи пакетного. При пакетному обробленні даних вони організуються у великі блоки; цей спосіб оброблення є доцільним у випадку необхідності достатнього часового інтервалу між процедурою збору та аналізу даних. При використанні потокового оброблення великих даних розглядаються одночасно комплекти невеликих пакетів, що дозволяє скоротити час на прийняття рішень,



хоча сам процес є достатньо складним та потребує більших витрат.

Очищення даних переслідує мету – виключити імовірність виникнення помилок, неточностей та уникнути можливості некоректного подання даних. Після очищення даних збільшується рівень якості отриманих результатів за рахунок видалення тієї інформації, яка дублюється, неправильно відформатована чи є нерелевантною.

Етап аналізу великих даних – один із вирішальних етапів роботи з *Big Data*, на якому важливим є обґрунтований вибір видів та методів аналізу даних. При виборі виду аналізу можна скористатися описовим чи дослідницьким аналізом, аналізом висновків, прогнозним чи причинно-наслідковим аналізом.

Залежно від цілей, які прагне досягти підприємство, визначається доцільність застосування того чи іншого виду аналізу [5]. Для описового аналізу характерним є оцінка ефективності діяльності підприємства за певний термін часу з можливістю порівняння даних на початку і вкінці періоду. Суть дослідницького аналізу – у візуалізації аналізу даних, що дає можливість зрозуміти пріоритетність функцій та зв'язок між ними. Реалізація дослідницького аналізу може здійснюватися двома способами: аналіз однієї змінної (одновимірний) або двох чи трьох змінних (багатовимірний).

Застосування аналізу висновків ґрунтується на екстраполяванні та узагальненні даних стосовно більших груп на основі дослідження менших по розміру, але репрезентативних сукупностей даних. При прогнозному аналізі достовірність одержаних результатів залежить від якості вхідних даних; цьому виду аналізу є властивою орієнтація на оцінку майбутнього, ґрунтуючись не на фактичних даних, а на прогнозних.

Причинно-наслідковий аналіз дає можливість встановити взаємозв'язок, який існує між причинами та наслідками, та надати уявлення про те, яким чином дії чи обставини зумовлюють одержання певного результату.

Сучасними науковцями та практиками виділяються різноманітні методи та техніки аналізу великих даних (таблиця 2).



Таблиця 2 – Сучасні методи та техніки аналізу великих даних

№ з/п	Метод	Сутність методу
1	<i>Machine Learning</i> (машинне навчання)	Відноситься до галузі штучного інтелекту, реалізується у вигляді складного ітеративного процесу, що передбачає підбір необхідного алгоритму, процедуру збору та підготовки даних, налаштування та упровадження моделі
2	<i>Real-time Data Processing</i> (опрацювання даних у реальному часі)	Зводить до мінімуму час, що витрачається підприємством на аналіз даних, що одержуються у реальному часі
3	<i>Deep Analysis</i> (глибокий аналіз)	Дозволяє виявити складні зв'язки та патерни, використовуючи алгоритми та статистичні методи
4	<i>Social Network Analysis</i> (аналіз соціальних мереж)	Виявляються способи та особливості взаємодії людей та аналізується тип їх поведінки
5	<i>Text Analysis</i> (текстовий аналіз)	Дає можливість отримувати корисну інформацію із неструктурованих даних
6	<i>Geospatial Analysis</i> (геопросторовий аналіз)	Застосовуються методи статистичного аналізу та інформаційні технології до даних, що носять геопросторовий характер
7	<i>Data Mining</i> (інтелектуальний аналіз даних)	Дають можливість вирішити ряд завдань по класифікації, кластеризації, послідовній асоціації, візуалізації та ін.
8	<i>Crowdsourcing</i> (краудсорсинг)	Дані одержують, використовуючи велику кількість джерел, із залученням інтелектуальних здібностей невизначеної кількості осіб
9	<i>Simulation</i> (імітаційне моделювання)	Будуються моделі, які імітують поведінку різних систем шляхом побудови різноманітних сценаріїв
10	<i>A/B testing</i> (A/B-тестування)	Порівнюються гіпотези та приймаються рішення з використанням даних, а не інтуїції
11	<i>Artificial neural networks</i> (штучні нейронні мережі)	Створення обчислювальних моделей, які подібні за структурою та функціями нейронним мережам людського мозку
12	<i>Data Fusion and Data Integration</i> (злиття та інтеграція даних)	Застосовуваний інструментарій дає можливість інтегрування даних, одержаних із різних джерел, з подальшим їх аналізом
13	<i>Time series analysis</i> (аналіз часових рядів)	Здійснюється аналіз даних, які повторюються у розрізі певного терміну часу, з використанням цифрових та статистичних засобів
14	<i>Unsupervised learning</i> (навчання без учителя)	Використовуються технології машинного навчання, які виявляють неявні взаємозв'язки в сукупності даних

Авторська розробка

Важливим є вибір технологій для роботи з великими даними, які забезпечують можливість ефективно їх обробляти та одержувати точні та корисні результати.



1. *Hadoop*. Технологія, яка користується найбільшою популярністю за рахунок:

- використання фреймворків MapReduce та Hadoop Distributed File System;
- придатність до оброблення різних типів даних – текстових, відео та ін.;
- здатність обробляти дані у режимі реального часу (із-за наявності Apache Storm).

2. *Apache Spark*. Особливості технології:

- проведення операцій з *Big Data* у високошвидкісному режимі;
- здатність до підтримки різноманітних джерел *Big Data*;
- можливість надавати високорівневі API з орієнтацією на різні мови програмування, що дає можливість швидкого освоєння цієї технології розробникам з різним досвідом праці у сфері програмування.

3. *Apache Kafka*. Основні відмінності технології:

- відноситься до категорії розподілених систем, що здійснюють потокове оброблення даних;
- властива гарантованість при доставленні повідомлень;
- доступність та прозорість процесів масштабування.

4. *MapReduce*. Технологія, створена компанією Google, призначена для:

- здійснення паралельного оброблення великих даних на розподіленій основі;
- орієнтована на застосування кластерів недорогого комп'ютерного обладнання;
- оброблятися можуть дані, що відносяться як до неструктурованих (місце збереження – файлова система), так і до структурованих (місце збереження – база даних).

5. *NoSQL*. Позначаються групи баз даних, які володіють властивостями, які не характерні для традиційних баз даних:

- відносяться до нереляційних типів баз даних, для яких не потребується фіксована схема;
- охоплює широкий набір технологій баз даних – структурованих,



напівструктурованих, неструктурованих та поліморфних;

– характеризуються масштабованістю, великою гнучкістю, потужною продуктивністю, високим рівнем функціональності;

– сумісні з більшістю сучасних додатків – мобільних, веб-додатків та ін.

6. *Deep Learning*. Технологія, що відноситься до машинного навчання, і для якої є характерним:

– застосування багат шарових нейронних мереж для аналізу даних;

– здатність аналізувати великі обсяги неструктурованих даних у вигляді зображень, аудіо та текстів;

– можливість здійснення складних аналізів та виявлення складних патернів.

2.2. Світова індустрія великих даних

За оцінкою фахівців, популярність *Big Data* буде невпинно зростати в найближчі роки, що буде сприяти активізації процесів розвитку в будь-яких сферах розвитку суспільного життя.

За прогнозами Statista/Wikibon, світовий ринок *Big Data* у 2024 році досягне рівня 84 млрд дол. США, у 2025 році – 90 млрд дол. США, у 2026 році – 96 млрд дол. США, у 2027 році – 103 млрд дол. США (рис.1).

Твердження, що використання *Big Data* необхідно лише для великих підприємств, не має сенсу, так як збільшується кількість середніх та малих підприємств, які усвідомлюють переваги великих даних та мають можливість упевнитися у безперешкодному доступі до одержання послуг щодо їх оброблення та аналізу.

Потреба у необхідній інфраструктурі для *Big Data* реалізується в теперішній час у більш доступній формі, урахувавши значне зростання обчислювальних потужностей за останній час, зниження вартості обладнання та можливість застосування сучасних хмарних сервісів.

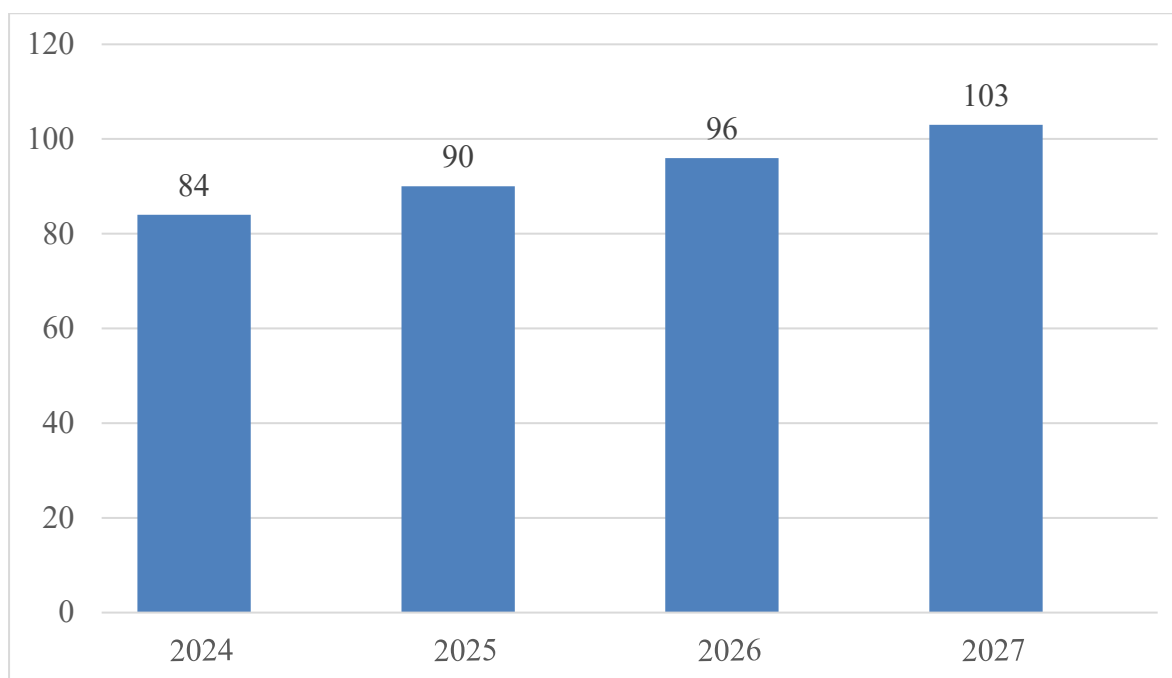


Рисунок 1 - Прогнозні обсяги світового ринку *Big Data*, млрд дол. США

Джерело: [6]

Однак, основною умовою результативності *Big Data* залишається не лише використання їх у діяльності підприємств, а і рівень обґрунтованості проведеного аналізу та достовірність зроблених висновків. Фахівцями великі дані порівнюються з нафтою, яка потребує очищення і переробки для того, щоб надалі її можна було використовувати.

Інтерес визивають сформульовані Бернардом Марром тенденції в сфері *Big Data*, які він визначав для 2022 року, але які будуть зберігати актуальність і в наступні роки [7].

1. *TinyML*. Ця технологія відноситься до інструментів машинного навчання, об'єднує програмне, апаратне забезпечення та алгоритми, які функціонують синхронізовано для забезпечення необхідного рівня продуктивності. Може бути реалізованою на пристроях, які мають низьке енергоспоживання і обмежену пам'ять, володіє потенціалом для застосування штучного інтелекту у величезній кількості пристроїв, що раніше не могли реалізувати таку можливість.

Відповідно прогнозу *ABI Research*, до 2030 року кількість пристроїв, що будуть застосовувати технологію *TinyML*, складе 2,5 млрд. Підтвердженням



об'єктивності цього прогнозу є результати досліджень консалтингової компанії *Silent Intelligence*, відповідно яким за наступні п'ять років вартісна оцінка ринку *TinyML* досягне рівня 70 млрд дол.

2. *Big Data і клієнтський досвід*. Поняття клієнтського досвіду виникло у 1960-х роках, коли з'явилося розуміння пріоритетності довгострокових відносин між клієнтами та підприємством порівняно з продуктом чи послугою.

За визначенням дослідницької компанії *Forrester Research*, під клієнтським досвідом слідує розуміти сприйняття клієнтами компанії як результат всіх взаємодій з нею. Компанією *Gartner* клієнтський досвід сприймається в якості сукупності відчуттів, що отримуються клієнтами за весь час контактування з брендом – починаючи з моменту ознайомлення, включаючи придбання товару та контакти з фірмою після здійснення покупки.

Застосування сервісів *Big Data* забезпечують створення портрета споживача, що дає можливість персоніфікувати процеси його обслуговування. Виникає можливість для оптимізації процесів, що відбуваються у електронній комерції, зменшення терміну очікування та обміну інформацією між структурними підрозділами підприємства у випадках звернення до служб підтримки клієнтів.

3. *AutoML*. За рахунок цього хмарного продукту, що об'єднує різні інструменти машинного навчання, розпізнаються зображення та відбувається пошук у хмарі прийнятної нейронної мережі, яка в найбільшій мірі відповідає цілям вирішуваного завдання. Автоматизоване машинне навчання відрізняється від традиційного відсутністю необхідності оволодіння всіма алгоритмами машинного навчання, їх особливостями застосування. Створені розробниками платформи та інструментарій можуть бути використаними для формування своїх програм машинного навчання.

4. *Штучний інтелект та синтетичні дані*. В деяких випадках у підприємства можуть виникати труднощі щодо збору та використання реальних даних, якщо це може привести до порушення їх конфіденційності, правил захисту чи взагалі існують перепони до їх доступу. Ця проблема може бути



вирішена шляхом створення синтетичних даних за допомогою технологій штучного інтелекту. Використання синтетичних даних здійснюється без порушення етичних та правових норм, мінімізує роботу з реальними даними, заощаджує час і ресурси.

5. *Розміщення аналітики даних у хмарних сховищах.* Хмарні сховища, що призначені для зберігання великих даних, функціонують на умовах високої відмовостійкості, масштабованості та безпеки розміщених даних. Вони можуть використовуватися для різноманітних програм, але можуть потребувати спеціальних знань для роботи з ними та бути достатньо дорогі.

Велика кількість компаній уже користується гібридними або мультимарними системами і перенесення аналітики в хмару буде наростати із збільшенням можливості одержання послуг від різних постачальників хмарних сервісів.

За твердженням представників компанії McKinsey, великі дані необхідно розглядати як важливий фактор виробництва нарівні з трудовими ресурсами чи виробничими активами. Використання *Big Data* може бути результативним в наступних сферах:

1. *Big Data в маркетингу.* Великі дані створюють умови для визначення поведінкових чи інших характеристик цільової аудиторії підприємства. *Big Data* дають можливість прогнозування клієнтської поведінки, виявлення чинників, які впливають на прийняття рішення клієнтами про покупку. Покращення комунікації зі споживачами, застосування персоніфікованого підходу до них дає можливість збільшити рівень лояльності та зацікавленості клієнтів, забезпечити зростання продажів, удосконалити рішення в сфері ціноутворення, збільшити клієнтську базу.

2. *Big Data у сфері логістики.* Логістична сфера пов'язана з постійним використанням великих даних, які створюють умови для удосконалення цілого ряду процесів – робота з перевізниками, управління логістичними витратами, робота складу, управління розкладом. *Big Data* дозволяє оптимізувати логістичні ланцюги, ефективно використовувати всі ресурси. За експертними оцінками,



аналіз великих даних щодо використання приміщень складів, визначення маршрутів та використання транспорту приводить до скорочення логістичних витрат на 15-25%.

3. *Big Data у виробництві.* Застосування аналізу великих даних у виробництві дозволяє оцінювати стан обладнання, збільшувати його продуктивність та скорочувати величину експлуатаційних витрат, зменшувати кількість простоїв. Важливим є розрахунок термінів окупності різних видів обладнання, визначення оптимальних технологічних режимів та відповідність існуючим нормативним вимогам [8]. Прикладом є авіаційна галузь: при здійсненні трансатлантичного польоту літаком Boeing 747 Jumbo виробляється 640 Тбайт даних за годину. Ураховуючи надзвичайну важливість інформації щодо функціонування всіх систем літака, вчасне оброблення та зберігання даних є суттєвим чинником забезпечення льотної безпеки та основою для виявлення причин порушення безпеки у випадках авіаційних інцидентів [9].

4. *Big Data в управлінні безпекою та ризиками.* Використання інтелектуального аналізу даних та машинного навчання дозволяє зменшувати ризики за рахунок попередження та виявлення шахрайських дій. Виникає можливість ідентифікації нових ризиків шляхом застосування шаблонів даних, що служать основою формування ефективного інструментарію ризик-менеджменту.

Інструментарій фрод-моніторингу, створений на основі *Big Data*, дозволяє здійснювати аналіз транзакцій в режимі реального часу відповідно до сформульованих критеріїв, а також відслідковувати дії користувачів з метою виявлення аномальних ситуацій чи нелогічної, з точки зору законності, їх поведінки.

5. *Big Data в ритейлі.* Результати досліджень GE Capital Retail Bank свідчать про те, що суттєва більшість покупців (біля 81%) перед прийняттям рішення про придбання товару займається пошуком необхідної інформації, використовуючи онлайн-можливості [10]. Зростання стосується і ринку реклами – прогнозовані витрати світових ритейлерів, пов'язані з просуванням товарів в мережі, складуть



у 2024 році 133, 06 млрд дол. США [11].

Отже, будуть зростати об'єми інформації, що супроводжують розвиток цих явищ, і, відповідно, буде відчуватися потреба у відповідному інструментарії обробки великих даних.

6. *Big Data і сфера нерухомості*. Аналіз великих даних дозволяє вирішувати комплекс завдань у сфері нерухомості, одне із яких – пошук клієнтів (з урахуванням їх фінансового стану, вікових та поведінкових характеристик) для реалізації об'єктів за рахунок формування ефективної маркетингової стратегії.

Важливу роль *Big Data* можуть відігравати при оцінці вартості нерухомості у цих випадках аналізується рівень попиту на аналогічну нерухомість, існуюча пропозиція, показник продажів стосовно певного міста чи району, характеристики нерухомості і т.і. Одержана інформація може бути використана для формування ринкової ціни нерухомості та задоволення потреб продавців і покупців.

Для інвесторів *Big Data* забезпечує прийняття рішень щодо купівлі об'єктів нерухомості на основі співставлення цін на нерухомість та власне процедур купівлі у різних районах чи регіонах, оцінки доцільності придбання нерухомості в теперішній час чи згодом, урахуваючи прогнози щодо зміни цін, вартості кредитів та ін. Аналіз *Big Data* дає можливість уникнути шахрайських дій зі сторони орендарів, надаючи всі відомості про них орендодавцям; є також можливість знайти оптимальний варіант оренди залежно від одержаного прибутку та витрат, що пов'язані з орендними відносинами.

7. *Big Data як інструмент прийняття рішень*. Великі дані розглядаються як потужний інструмент прийняття стратегічних рішень. *Big Data* забезпечує для підприємств здійснення прогнозів щодо тенденцій у зовнішньому середовищі на основі застосування історичних даних та алгоритмів машинного навчання.

Наявність *Big Data* дає змогу менеджменту компанії приймати рішення, не покладаючись на інтуїцію чи досвід, а виходячи з реальної інформації та згенерованих на її основі знань. Рішення, прийняті на основі великих даних, відрізняються високим рівнем якості та обґрунтованості, урахувають імовірність



виникнення ризикових ситуацій, в тому числі і непередбачуваних. *Big Data* дає можливість оптимізувати бізнес-процеси, знаходити додаткові джерела доходів, підвищувати рівень інноваційності та конкурентоспроможності підприємства.

Для розуміння стану управління даними на підприємстві та його удосконалення доцільним є використання шестиступінної моделі зрілості управління даними (Data Governance Maturity Model), створеної компанією Gartner (таблиця 3).

Таблиця 3 – Етапи реалізації моделі зрілості управління даними

Найменування етапу	Сутність етапу
Етап 0. Необізнаний	Відсутність політики управління великими даними. Прийняття стратегічних рішень відбувається без використання <i>Big Data</i> – аналітики
Етап 1. Обізнаний	Важливість управління великими даними усвідомлюється топ-менеджментом фірми та працівниками ІТ-сфери. Однак, не визначене та документарно не закріплене рішення щодо політики у сфері управління великими даними
Етап 2. Реактивний	Всім працівникам властиве розуміння важливості процесів управління <i>Big Data</i> . Починається впровадження реактивних заходів щодо формування системи правил роботи з інформацією, але дисципліна щодо виконання цих правил є низькою
Етап 3. Проактивний	Система управління великими даними приймається і виступає у якості складової частини кожного проекту
Етап 4. Керований	Наявна конкретна програма щодо управління великими даними. Процесам та політикам дається визначення щодо кожної стадії життєвого циклу підприємства
Етап 5. Ефективний	Найвищий рівень зрілості управління даними, ефективний збір та аналіз даних, що забезпечують конкурентні переваги підприємству

Джерело: складено автором на основі [12]

Орієнтуючись на упровадження технологій *Big Data*, підприємства повинні застосовувати сучасний інноваційний інструментарій, що дозволяє оптимізувати процеси оброблення та зберігання великих даних.

Якщо топ-менеджментом підприємства прийнято рішення про застосування *Big Data*, можливий вибір однієї із двох альтернатив: іти шляхом застосування уже готових рішень, які надають вендори, чи створювати власну



платформу *Big Data*. У першому випадку підприємство може стикнутися з недостатньою адаптованістю продукту до особливостей його функціонування, що може потребувати значних грошових та часових витрат.

При розробці індивідуальної платформи буде урахована специфіка діяльності підприємства, механізм реалізації бізнес-процесів та потреби споживачів. Перевагами індивідуальних розробок є можливості передбачення процесів подальшого масштабування, на відміну від готових придбаних продуктів, які у разі виникнення такої ситуації будуть потребувати додаткових суттєвих витрат.

Індивідуальні аналітичні рішення в більшості випадків забезпечують інтеграцію з різними типами цифрових інструментів, платформ та систем, що також вигідно відрізняє їх від уже існуючих пакетних продуктів. Також може бути передбачена можливість забезпечити безпеку, за рішенням менеджменту підприємства, особливо важливої інформації, що є проблематичним при використанні уже існуючих платформ.

У той же час, створення аналітичних продуктів потребує наявності на підприємстві професійного персоналу з високим рівнем кваліфікації у сфері машинного навчання та великих даних, який зміг би створити по-справжньому ефективний продукт.

2.3. Великі дані в Україні

Українські підприємства в умовах війни працюють, долаючи цілий ряд викликів: клієнтський попит зазнав змін, виникли проблеми із кваліфікованими працівниками із-за мобілізації та міграції, зростають ціни на паливо, зменшилася платоспроможність клієнтів.

По результатам дослідження *Київстар Бізнес* протягом листопада 2023 року, що проводилися серед 750 українських компаній (мікрокомпанії, малі, середні та великі компанії), діяльність 62% компаній відбувалася повноцінно;



24% компаній функціонували не на повну потужність; відновлювали роботу 4%; 2% компаній були вимушені призупинити діяльність або змінити її сферу; 5% закрили свій бізнес [13].

Отримана аналітика щодо визначення потреби у *Big Data* серед різних сфер бізнесу дозволила виявити, що найбільшу потребу демонструвала галузь ритейл (онлайн та офлайн) – 28,77%, друге місце відведене закладам харчування – 17,81%, сфера нерухомості та сфера послуг – 13,7% та 12,33% відповідно, виробництво та краса і спорт – 9,59% та 8,22%, медицина і ветеринарія – 6,85%, освіта і саморозвиток – 2, 74% [16] (рисунок 2).

Сегмент ритейлу, який лідирує по кількості запитів на великі дані, збереже тенденцію до зростання надалі. В Україні найбільшими ритейл-гравцями є Rozetka, OLX, Hotline, які активно використовують *Big Data* – продукти, хоча вони в певній мірі різняться з тими методами, що є розповсюдженими в зарубіжних країнах. Поясненням цьому є недостатній рівень розвитку вітчизняного ринку ритейлу порівняно з американським чи європейським ринками.

Ще один чинник, який можна розглядати як перепону до активного розвитку ринку ретейлу з використанням *Big Data*, є вартість упровадження, яка передбачає не лише наявність відповідних обчислювальних потужностей, а і фахівців у сфері *Big Data*. Вирішенню цієї проблеми сприяє використання хмарних послуг, які можуть надавати не тільки відомі у світовому масштабі компанії (*Amazon Web Service, Microsoft Azure, Google Cloud, Alibaba Cloud, IBM Cloud*), а і вітчизняні хмарні провайдери, які можуть забезпечити прийнятне поєднання цінових та якісних характеристик послуги.

Хмарні технології дають можливість не тільки оперувати з великими даними, для них є характерний швидкий доступ до них. Хмари є надійним інструментом для захисту інформації при виникненні форс-мажорних обставин, кризових ситуацій, хиб у мережах.

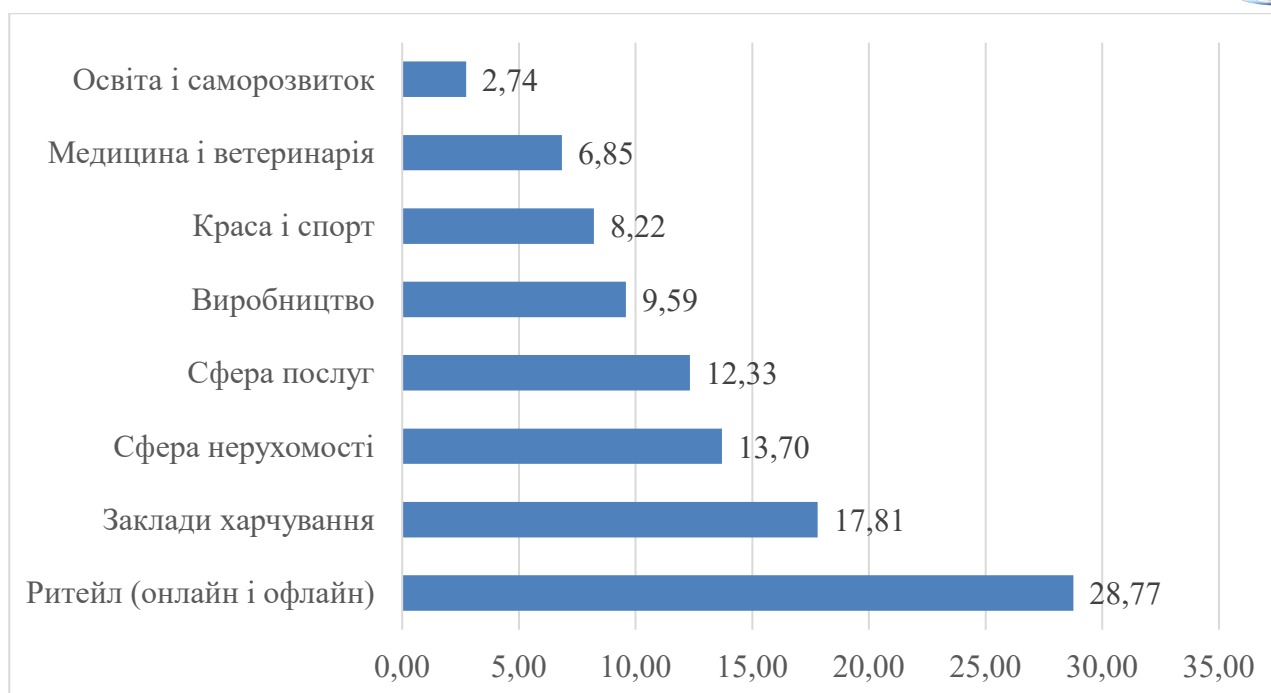


Рисунок 2 – Потреба у великих даних в різних галузях українського бізнесу, 2023 рік, %

Джерело: Побудовано автором на основі [14]

Для 2024 року будуть характерними наступні тренди у хмарних технологіях:

- використання штучного інтелекту як послуги, що створить оптимальні умови для оброблення та аналізу *Big Data*, що підвищить рівень обґрунтованості управлінських рішень;

- збільшення попиту на хмарні обчислення у реальному часі для більш оперативного отримання актуальних даних;

- збільшення популярності використання гібридних хмар та мультихмар;

- зростання актуальності інструментів захисту великих даних у хмарах, ураховуючи прогнози щодо активізації кіберзагроз та появи нових способів зловживань у кіберпросторі;

- наближення платформ постачальників хмарних послуг до користувачів з метою забезпечення надшвидкої обробки даних.

В Україні функціонує ряд компаній, які можуть надавати послуги по вирішенню задач з використанням великих даних. Одна із них – компанія *Київстар, Big Data* – рішення від якої розробляються для клієнтів різної форми власності та незалежно від розміру – для малих, середніх та великих (таблиця 4).

Таблиця 4 – *Big Data* – рішення від *Київстар*

Послуга	Сутність послуги
Побудова портрета клієнта	Сформований портрет існуючої та потенційної клієнтської аудиторії є базою для оптимізації прийняття рішень замовниками послуг щодо задоволення потреб клієнтів
Застосування інструментарію геоаналітики	Проведений аналіз територій дозволяє визначити найбільш вдале розміщення об'єктів бізнесу (у випадках створення чи розвитку), зрозуміти доцільність застосування реклами та ін.
Формування теплових карт (при необхідності – кастомних)	Теплові карти дозволяють визначити зони зосередження цільової аудиторії; при розробленні кастомних теплових карт додатково надається інформація щодо робочих, домашніх зон, транзитного трафіку та ін.
Тригерні розсилки	Здійснюється надсилання повідомлень представникам цільової аудиторії у випадках безпосереднього знаходження біля місць зосередження точок бізнесу
Таргетовані розсилки	Застосування є доцільним при виведенні на ринок нового продукту, наданні нової послуги чи створенні нових підрозділів бізнесу без інвестування у великі рекламні компанії. Найбільший ефект досягається для приваблення клієнтів у високомаржинальних сферах, де покупки відбуваються спонтанно

Джерело: складено автором на основі [15]

Компанія *Vodafone* також використовує технологію *Big Data*, надаючи послуги замовникам:

- щодо аналізу клієнтів (місце проживання, коло інтересів та ін.);
- здійснення таргетованої реклами (з визначенням необхідного місця і часу для одержання максимального зворотного ефекту);
- застосування моделі *Look-a-like* (для визначення потенційних клієнтів, які з великою імовірністю можуть перетворитися в постійних);
- реалізація точної геоаналітики (визначення найбільш привабливих місць для розташування бізнесу).

Загальною проблемою функціонування компаній, що акумулюють великі дані, є забезпечення їх безпеки. 23 грудня 2023 року була здійснена масована



хакерська атака на віртуальну інфраструктуру *Kiivstar*, в результаті якої вона була частково зруйнована і виникла загроза інформаційним ресурсам компанії.

Начальником управління кібербезпеки СБУ було заявлено, що проникнення в систему *Kiivstar* було здійснено російськими хакерами ще у травні 2023 року. Представниками компанії було повідомлено, що за допомогою СБУ за лічені дні були відновлені всі системи та попереджені всі нові кібератаки, також не був зафіксований витік персональних та абонентських даних.

Очевидним є те, що кібербезпека буде залишатися однією із ключових умов розвитку *Big Data* в найближчому майбутньому. За оцінкою фахівців у сфері кібербезпеки, до основних проблем, яких можна уникнути, застосовуючи партнерство з MSP (постачальником управління послугами), відносяться [16]:

- недостатні навички по питанням кібербезпеки, які можуть бути компенсовані шляхом застосування комплексу інструментів, що дозволяють контролювати стан безпеки відносно всіх загроз;

- невідповідність існуючим нормативам, що може бути ліквідована у випадку виявлення потенційних проблем, пов'язаних як з мережею і ресурсами, так людськими чинниками ризику;

- кіберризик змінного характеру, для попередження яких MSP можуть застосовувати передову практику постачальників та партнерів;

- зміни бізнес-стратегій (впровадження двосторонньої стратегії – запобігання загроз та виявлення і пом'якшення їх дії); MSP повинні орієнтувати своїх клієнтів на те, щоб виявляти та реагувати на загрози, маючи плани дій для продовження роботи (у випадках виникнення загроз) чи швидкого його перезапуску;

- зростання масштабів послуг, що збільшує сферу можливих загроз; для вирішення цієї проблеми MSP повинні пропонувати індивідуальні рішення, урахувавши запити клієнтів та особливості конкретної галузі, застосовуючи розширену платформу послуг;

- використання трансформованих рішень з метою забезпечити кібербезпеку; MSP необхідно орієнтуватися на використання інноваційних



рішень, насамперед генеративного інтелекту, для поглибленої оцінки загроз та формування дій у межах спеціально розробленої платформи керованих послуг.

Висновки

Кількість інформації у світі постійно зростає, великі дані стали інструментом стратегічного менеджменту та засобом підвищення фінансової результативності функціонування підприємств. Якісна інформація, яку можна одержати в результаті оброблення та аналізу *Big Data*, є беззаперечною умовою успішної діяльності будь-якого підприємства.

Великі дані створюють можливість вирішення цілого ряду завдань, таких як поліпшення прогнозу попиту клієнтів, удосконалення процесу розроблення інноваційних продуктів та технологій з максимальною наближеністю їх до потреб споживача, вибір методів та форм оптимізації прийнятих рішень для персоналізації пропонованих клієнтам товарів та послуг.

Big Data, в силу наявності специфічних характеристик, основними із яких є об'єм, швидкість та різноманітність, потребує застосування нетрадиційного інструментарію аналізу та наявності кадрового забезпечення відповідної кваліфікаційної спрямованості. Важливим є обґрунтований вибір технологій для опрацювання великих даних, що буде служити умовою одержання достовірних результатів.

Перспективи розвитку індустрії *Big Data* як у світі, так і Україні, будуть пов'язані з використанням технологій машинного навчання, штучного інтелекту та синтетичних даних, хмарних продуктів. Особливої уваги будуть потребувати питання захисту великих даних, урахуваючи зростання загроз їх конфіденційності, цілісності та доступності зі сторони кіберпростору.