

ПРОБЛЕМИ НАЦІОНАЛЬНОЇ ТА РЕГІОНАЛЬНОЇ ЕКОНОМІКИ

УДК 330.4: 338.22: 330.42

ПЕРСПЕКТИВИ ВИКОРИСТАННЯ АНАЛІЗУ ВЕЛИКИХ ДАНИХ В ЕКОНОМІЦІ УКРАЇНИ

Антонюк О.П., к.е.н.,
Бєлкіна І.А., к.е.н.

Державний вищий навчальний заклад «Національний гірничий університет»

Прогрес інформаційно-комунікаційних технологій, збільшення використання електронних та мережевих пристроїв, а також оцифрування операцій виробничих процесів, призводять до генерації величезної кількості даних щорічно. Аналіз великих даних дає змогу виявити приховані закономірності, які можуть бути застосовані у широкому спектрі галузей економіки України, і привести до створення нових продуктів та послуг, підвищення конкурентоспроможності та економічного зростання. Дослідження свідчать, що компанії, які використовують аналітику великих даних, можуть підвищити продуктивність на 5% - на 10%. Поряд з цим існує ряд певних організаційних та технологічних труднощів, пов'язаних з використанням даних великих обсягів.

Метою статті є визначення перспектив використання технологій аналізу великих даних для розвитку бізнесу в Україні. З одного боку, згідно світових тенденцій технології *Big data* поширюють вплив на стратегії електронної комерції, особливо в області маркетингу і продажів. Аналіз відповідних даних, буде ставати все більш і більш індивідуалізованим під потреби окремого клієнта, з іншого боку, Україна володіє високим потенціалом спеціалістів з аналізу даних, які можуть не лише для реалізовувати запропоновані алгоритми, а й пропонувати нові технології та рішення.

Ключові слова: великі дані, аналіз даних, неструктурована інформація, системи управління базами даних, індивідуалізована підтримка клієнтів

UDC 330.4: 338.22: 330.42

PERSPECTIVES OF USING ANALYSIS OF BIG DATA IN THE ECONOMY OF UKRAINE

Antoniuk O., PhD in Economics,
Bielkina I., PhD in Economics

National Mining University, Dnipro

The progresses of information and communication technologies, increasing the use of electronic and network devices, as well as the digitization of product process operations generate a huge amount of data annually. The analysis of big data allows us to identify hidden patterns that can be applied in a wide range of sectors of the Ukrainian economy

© Антонюк О.П., к.е.н., Бєлкіна І.А., к.е.н., 2017

and may lead to the creation of new products and services, increasing competitiveness and economic growth. The studies have shown that companies which used analytics of big data can increase their productivity by 5% - 10%. Herewith, there are certain organizational and technological problems associated with the use of big data.

The purpose of this article is to determine the perspectives of using big data analysis technologies for business development in Ukraine. On the one hand, according to the global trends, the big data technology's influence is expanding on e-commerce strategies, especially in the field of marketing and sales. The analysis of the relevant data will become much more individualized for the needs of a specific client. On the other hand, Ukraine has high potential of specialists in data analysis who can not only implement the proposed algorithms, but also offer new technologies and solutions.

Keywords: big data, data analysis, unstructured information, database management systems, individualized customer's support

Актуальність проблеми. За останнє десятиріччя, обсяги даних, що генеруються та реєструються електронними засобами та додатками зростає запаморочливими темпами. Разом зі стрімким накопиченням інформації швидкими темпами розвиваються і технології аналізу даних. Якщо ще кілька років тому було можливо, лише сегментувати клієнтів на групи зі схожими перевагами та потребами, то тепер можливо будувати моделі для кожного клієнта в режимі реального часу, аналізуючи, наприклад, його переміщення по мережі Інтернет при пошуку конкретного товару. Інтереси споживача можуть бути проаналізовані, і відповідно до побудованої моделі запропонована відповідна контекстна реклама або конкретні пропозиції товарів з бажаними властивостями. Модель також може налаштовуватися і перебудовуватися в режимі реального часу, що було недосяжно ще декілька років тому.

Аналіз останніх досліджень та публікацій. Останнім часом з'явилося багато додатків, що дозволяють користувачам легко отримати доступ і аналізувати дані в хмарі. Ці набори даних вільно розміщуються і доступні для використання в різних сховищах даних і аналітики за допомогою програмного забезпечення, з відкритим вихідним кодом, наприклад, Apache Spark, передовими технологіями Google, такими як Google BigQuery і Google Cloud DataFlow.

Проте, сам факт можливості зберігати дані і робити над ними обчислення ще не надає конкурентної переваги. Важливо зрозуміти, які саме обчислення потрібно виконати з метою отримання економічного прибутку. Спеціалісти по обробці великих даних затребувані на IT-ринку праці сьогодні, що відкриває широкі перспективи для працевлаштування

випускників ВНЗ технічних спеціальностей. В працях провідних закордонні фахівців, починаючи з МакКінслі [1], В. Майер-Шенбергер та К. Кукьєр, М. Гілберт і українських науковців з обробки даних В. Султан, Н. Шаховська, В. Москалюк [3] визначено організаційні та технічні засади обробки великих даних, але питання про одержання економічних переваг та захист приватних даних при застосуванні хмарних технологій потребують додаткового вивчення.

Мета роботи полягає в визначенні перспектив та особливостей застосування інструментарію для аналізу великих даних в економіці України.

Викладення основного матеріалу дослідження. *Google Trends* вказує на початок активного росту вживання словосполучення «*Big data*» у світі починаючи з 2012 року, в Україні з 2014 року (рис.1) :

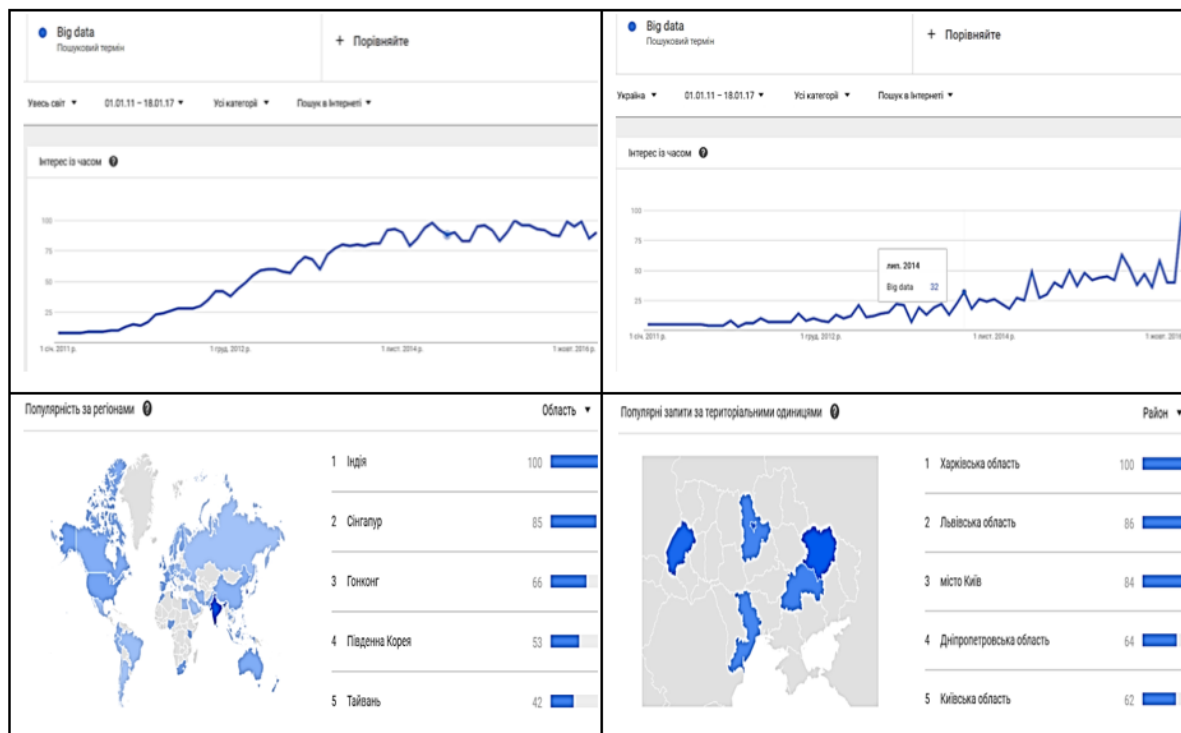


Рис.1. Статистика вживання словосполучення «*Big data*» в Україні та світі.
Джерело: авторська розробка

Як видно з рис.1 в Україні та світі термін «*Big data*» найчастіше застосовують в осередках розвитку ІТ індустрії, а саме в Харківській, Львівській, Київській та Дніпропетровській областях.

Великі дані (англ. *big data*) – в інформаційних технологіях, це серія підходів, інструментів і методів обробки структурованих і неструктурованих даних величезних обсягів і значного різноманіття, ефективних в умовах безперервного приросту, розподілу по численних

вузлах обчислювальної мережі, альтернативних традиційним системам управління базами даних і рішенням класу *Business Intelligence* [2].

Відмінність методів, що застосовують для аналізу *big data* від методів бізнес-аналізу (рис.2) обумовлена не лише обсягом даних, що обробляються, а насамперед їх неструктурованістю та швидкістю надходження.

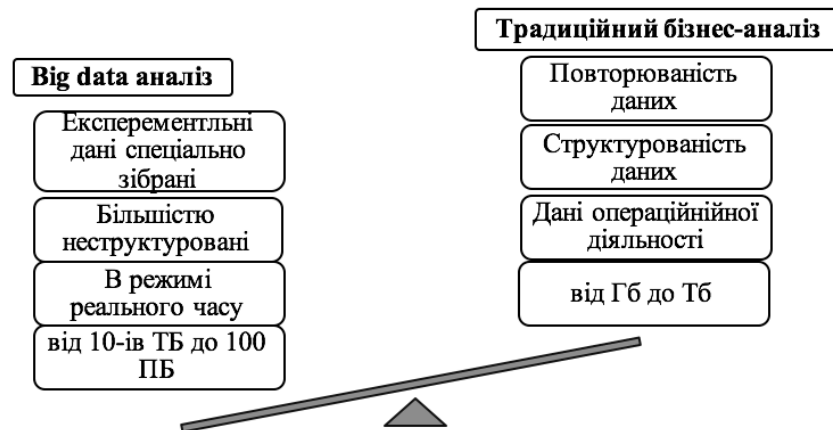


Рис. 2. Відмінність методів аналізу *big data* від бізнес-аналізу
Джерело: авторська розробка.

Тому системи для накопичення *Big data* мають задовольняти вимогам:

1. Горизонтальна масштабованість. Оскільки обсяг даних постійно зростає – будь-яка система, для обробки великих даних, повинна бути розширюваною.

2. Відмовостійкість. Принцип горизонтальної масштабованості передбачає, що машин в кластері може бути багато. Це означає, що частина цих машин буде гарантовано виходити з ладу. Методи роботи з великими даними повинні враховувати можливість таких збоїв і переживати їх без будь-яких значущих наслідків та втрат.

3. Локальність даних. У великих розподілених системах дані розподілені по великій кількості машин. Якщо дані фізично знаходяться на одному сервері, а обробляються на іншому – витрати на передачу даних можуть перевищити витрати на саму обробку. Тому одним з найважливіших принципів проектування *big data*-рішень є принцип локальності даних – по можливості обробляємо дані на тій же машині, на якій вони зберігаються. Технологічно це реалізується за допомогою механізмів *MapReduce*, *Spark*, *Tez* які в свою чергу реалізовано в екосистемі *Hadoop* [4].

Мета полягає в тому, щоб надати клієнтам індивідуальний і оптимізований торговий досвід в режимі реального часу. Інтернет-магазини можуть поліпшити свій бізнес з великими даними в наступних областях:

1. *Оптимізований портфель продуктів.* Аналіз великих обсягів структурованих даних клієнта дозволяє провести детальний аналіз цільової групи. На підставі результатів, портфель інтернет-магазину може бути адаптовано під потреби та вимоги конкретного користувача. Особливо великі інтернет-продавці можуть масштабувати свої пропозиції з *big data* краще для задоволення конкретних потреб клієнтів. З іншого боку *big data* також дозволяють прогнозувати потреби клієнтів і надають можливість майбутньої оптимізації продуктового портфеля. Таким чином, з використанням великих даних можна оптимізувати витрати на складі.

2. *Оптимізація ціни.* Завдяки *Big data*, стає можливим інтелектуальний аналіз даних і аналізу в реальному масштабі часу. Інтернет-магазин може налаштувати динамічну ціну продукту. Завдяки високій прозорості в Інтернеті, необхідно мати конкурентів завжди під наглядом і коригувати власну ціну для того, щоб залишатися конкурентоспроможними. *Big data* пропонують всебічний аналіз ринку для динамічної цінової політики.

3. *Оптимізований інтернет-магазин.* Завдяки використанню великих обсягів даних і швидких технологій веб-серверів, можна забезпечити динамічні веб-сайти. Різні стартові сторінки або цільові сторінки можуть відображатися в залежності від регіону або цільової групи. Крім того, можуть бути відображені різні переваги щодо асортименту продукції для чоловіків і жінок. Завдяки аналізу великих обсягів даних не існує практично ніяких обмежень можливих переваг для оптимізації онлайн-магазину.

4. *Оптимізація інтернет реклами.* У минулому показ реклами не був ефективним. На даний час з використанням технології *big data* онлайн роздрібні торговці можуть орієнтувати цільові рекламу та пропозиції для своїх клієнтів, що дозволяє залучити нових клієнтів. У режимі реального часу реклама дешевша і ефективніша, тому з великими даними інтернет-магазини можуть скоротити витрати на рекламу, а сама вона стає змістовнішою та ефективнішою.

5. *Оптимізація обслуговування клієнтів.* Якщо клієнт не задоволений продуктом і скаржився по телефону, а не по електронній пошті на обслуговування, то з часом така інформації буде втрачена. Великою перевагою є, якщо співробітник служби підтримки або маркетингу зможе використовувати повну історію клієнта, збагачену будь-якою соціальною інформацією, з засобів масової інформації під час спілкування, цей сценарій можливо реалізувати з *big data*. Різноманітність цінної довідкової інформації про клієнта дозволить службі підтримки істотно поліпшити ставлення клієнтів. Якість підтримки клієнтів також є конкурентною перевагою [5].

В якості визначальних характеристик для великих даних визначають «п'ять V» – *volume* (обсяг), *velocity* (швидкість), *variety* (різноманітність) (рис.3), останнім часом додано ще дві важливі характеристики – *veracity* (достовірність) і *variability* (мінливість):



Рис.3 Основні характеристики *big data*.

Джерело: [3].

Природа та походження *big data* обумовлюють використання як традиційних методів обробки даних та і специфічних для великих даних:

- *Data Mining* – навчання асоціативним правилами, класифікація (методи категоризації нових даних на основі принципів, раніше застосованих до вже наявних даних), кластерний аналіз, регресійний аналіз;

- *Краудсорсінг* – категоризація та збагачення даних силами широкого, невизначеного кола осіб;

- *Змішання і інтеграція даних* – набір технік, що дозволяють інтегрувати різноманітні дані з різноманітних джерел для можливості глибокого аналізу;

- *Машинне навчання*, включаючи навчання з учителем і без учителя, а також використання моделей, побудованих на базі статистичного аналізу або машинного навчання для отримання комплексних прогнозів на основі базових моделей;

- *Штучні нейронні мережі, мережевий аналіз, оптимізація*, в тому числі генетичні алгоритми;

- *Просторовий аналіз* – використання топологічної, геометричної і географічної інформації в даних;

- *Статистичний аналіз: α/β* – тестування і аналіз часових рядів;

- *Візуалізація аналітичних даних* – подання інформації у вигляді малюнків, графіків, схем і діаграм з використанням інтерактивних можливостей та анімації як для результатів, так і для використання в якості вихідних даних для подальшого аналізу.

Поряд з широкими перспективами використання *big data*, існують певні організаційні, технічні та технологічні проблеми, що потребують усвідомлення та врахування при застосуванні *big data* [6,7].

1. Зберігання і управління. Обсяг даних в сотні терабайт або петабайт не дозволяє легко зберігати і управляти ними за допомогою традиційних реляційних баз даних.

Big data зазвичай зберігаються і організуються в розподілених файлових системах. У загальних рисах, інформація зберігається на декількох (іноді тисячах) жорстких дисках, на стандартних комп'ютерах. Так звана «карта» (*map*) відстежує, де (на якому комп'ютері і / або диску) зберігається конкретна частина інформації. Для забезпечення відмовостійкості та надійності, кожен частину інформації зазвичай зберігають кілька разів, наприклад – тричі.

2. Неструктурована інформація. Більшість даних *big data* є неструктурованими, тобто велика частина зібраної інформації в розподіленій файловій системі складається з неструктурованих даних, таких як текст, зображення, фотографії або відео.

Це має свої переваги і недоліки. Перевага полягає в тому, що можливість зберігання великих даних дозволяє зберігати "всі дані", не турбуючись про те, яка частина даних актуальна для подальшого аналізу

і прийняття рішення. Недоліком є те, що в таких випадках для отримання корисної інформації потрібна подальша обробка цих величезних масивів даних. Хоча деякі з цих операцій можуть бути простими (наприклад, прості підрахунки, і т.д.), інші вимагають більш складних алгоритмів, які повинні бути спеціально розроблені для ефективної роботи на розподіленій файловій системі.

Отже, в той час як обсяг даних може рости в геометричній прогресії, можливості отримувати інформацію і діяти на основі цієї інформації, обмежені і будуть асимптотично досягати межі. Це дійсно велика проблема, пов'язана з аналізом неструктурованих даних.

3. Аналіз *big data*. Найбільш значним є використання всього обсягу даних для сегментації і кластеризації клієнтів, що дозволить ефективно будувати велику кількість моделей для невеликих кластерів.

Таким чином, один із способів отримання переваг від *big data* полягає в тому, щоб використовувати доступну інформацію для побудови великої кількості моделей для великого числа сегментів і, потім, за відповідною моделлю будувати прогнози.

У граничному випадку, кожна окрема «людина» у великому сховищі даних клієнтів може мати свою власну модель для прогнозування майбутніх покупок.

Це означає, що аналітична платформа, що підтримує сховища даних, повинна бути в змозі управляти сотнями або навіть тисячами моделей, і мати можливість перенастроювати їх, коли це необхідно.

4. Актуальними є питання про захист приватних даних при застосуванні хмарних технологій. Інформація про переміщення, дзвінки чи інші дії конкретної особи є конфіденційною. Ніхто не може її отримати від мобільного оператора, окрім відповідних органів, і то за рішенням суду. Але існує можливість використовувати загальну інформацію на користь людям. Наприклад, аналізуючи графік переміщення осіб протягом дня, керівництво міста могло б поліпшити стан дорожнього руху, оптимізувати його. Так, Vodafone Україна перед тим, як відкрити новий магазин, ретельно вивчає карту руху своїх абонентів, щоб зрозуміти, чи достатньо їх у цьому районі, чи буде затребуваний такий магазин саме в цьому місці. Але існує європейська практика, коли оператор ділиться знеособленою, загальною інформацією про своїх клієнтів, де немає адрес, номерів телефонів чи прізвищ.

Використовують лише аналітику: де живуть люди з такими-то інтересами, їхній вік та кількість. Така інформація допомагає розвиватися бізнесу, приймати більш зважені рішення про відкриття нових точок обслуговування клієнтів чи продажу товарів [8].

Висновки. Технології *big data* будуть впливати на стратегії електронної комерції. Особливо в області маркетингу і продажів, аналіз відповідних даних, буде ставати все більш і більш індивідуалізованим під потреби клієнта. Результати аналізу великих даних застосовують для динамічної адаптації інтернет-магазинів, оптимізації портфеля в режимі реального часу і індивідуалізованої підтримки клієнтів, у відповідності з їх споживчою поведінкою та уподобаннями. Також *big data* допоможуть боротися з шахрайством і знизити пов'язані з ним ризики.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ:

1. McKinsey. Big data: The next frontier for innovation, competition, and productivity /James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers. // Електр. ресурс. – Режим доступу: <https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation>
2. Большие данные // Електр. ресурс. – Режим доступу: ru.wikipedia.org/wiki/%D0%91%D0%BE%D0%BB%D1%8C%D1%88%D0%B8%D0%B5_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D0%B5
3. Пістунов І.М. Електронна економіка. Том 1. Криптовалюта. Big Data [Електронний ресурс]: Навч. посібник/ І.М. Пістунов, О.П. Антонюк – / М-во освіти і науки України; Нац. Гірн. ун-т. – Д.: НГУ, 2017. – 281 с. // Електр. ресурс. – Режим доступу: pistunovi.inf.ua/EE_KC_BD.pdf
4. What Is Apache Hadoop? // Електр. ресурс. – Режим доступу: <http://hadoop.apache.org/>
5. Big Data: Future Of E-Commerce Big Data: Future Of E-Commerce. // Електр. ресурс. – Режим доступу: <https://www.linkedin.com/pulse/20141112200850-1172947-big-data-future-of-e-commerce>
6. Big Data: A Revolution That Will Transform How We Live, Work, and Think / V. Mayer-Schönberger, Eamon Dolan/Mariner Books; Reprint edition. – 2014. – 272 p.
7. Силен Д. Основы Data Science и Big Data. Python и наука о данных /Д. Силен, А. Мейсман, М. Али/ Питер. – 2017. – 336 с.
8. Як Vodafone та інші великі компанії використовують наші дані// Електр. ресурс. – Режим доступу: <https://techtoday.in.ua/vodafone-news/yak-vodafone-ta-inshi-veliki-kompaniyi-vikoristovuyut-nashi-dani-60797.html>