

Список використаних джерел

1. Цілі сталого розвитку Україна – Поточний прогрес у розрізі цілей та індикаторів. Держ. служба статистики України. URL: <http://www.ukrstat.gov.ua/>
2. Організація економічного співробітництва та розвитку. URL: <https://data.oecd.org/>
3. Офіційний сайт Європейського Союзу. URL: <https://ec.europa.eu/>
4. Гендерні дослідження : прикладні аспекти : монографія. за наук. ред. В. П. Кравця. Тернопіль : Навчальна книга – Богдан, 2013. 448 с.
5. Індекс гендерної нерівності (GII). Звіти про людський розвиток. URL: <https://hdr.undp.org/data-center/thematic-composite-indices/gender-inequality-index#/indicies/GII>
6. Індекс гендерного розвитку (GDI). Звіти про людський розвиток. URL: <https://hdr.undp.org/gender-development-index#/indicies/GDI>

РОЛЬ ТЕХНОЛОГІЙ BIG DATA В СУЧАСНІЙ СТАТИСТИЦІ

Єлісєєва Оксана Костянтинівна,

доктор економічних наук, професор,
професор кафедри економічного моделювання, обліку та статистики;

Перетяцько Андрій Ігорович

аспірант,

факультет економіки;

Дніпровський національний університет ім. О. Гончара

Big Data – це серія підходів, інструментів і методів опрацювання структурованих і неструктурованих величезних обсягів і значного розмаїття даних для одержання результатів, які людина сприймає і які є ефективними за умов невинного приросту, розподілу по численних вузлах обчислювальної мережі. Цей напрям сформувався наприкінці 2000-х рр. і є альтернативним традиційним системам управління базами даних і рішенням класу Business Intelligence. До цієї серії включають засоби масового опрацювання невизначено структурованих даних, передусім, рішеннями категорії NoSQL, алгоритмами MapReduce, програмними каркасами та бібліотеками проекту Hadoop.

Термін "великі дані" існує вже майже двадцять років, ставши за цей час всесвітньо обговорюваним. За історію свого недовгого існування він встиг здобути широку популярність – когось ці технології вже встигли розчарувати, а якісь компанії вже впровадили Big Data у свою діяльність, що дало змогу оптимізувати роботу з даними. Під великими даними розуміють сукупність даних з можливим експоненціальним зростанням, які є занадто великими, занадто неформатованими або занадто неструктурованими для аналізу традиційними методами [8]. Інструментарій і методи обробки великих даних докорінно відрізняються від роботи зі звичайними базами даних. На

сьогоднішній день, розробкою інструментів для роботи з Big Data займаються всесвітньо відомі компанії - Microsoft, Oracle, IBM, SAP [1].

Описуючи характеристики, що визначають великі дані такими, відзначають так звані "три V": Volume – фізичний об'єм, наявна база даних має являти собою великий об'єм інформації, для якого було б надто трудомістко застосовувати традиційні методи опрацювання та зберігання даних; Velocity – швидкість, не лише накопичення даних, а й необхідна швидкість їх опрацювання; Variety – різноманіття, що дає змогу одночасного опрацювання як структурованої, так і неструктурованої різноформатної інформації. Неструктурована інформація, на відміну від структурованої, яку можна класифікувати, потребує значно ретельнішого та комплекснішого аналізу для отримання можливості її якісного подальшого опрацювання. Першим критерієм порівняння є обсяг інформації: якщо традиційні дані займають від кількох гігабайт (10⁹ байт) до терабайт (10¹² байт), то обсяг бази BigData сягає від кількох петабайт (10¹⁵ байт) до кількох ексабайт (10¹⁸ байт). Можна сказати, що реалізується принцип "N=All", який говорить про те, що людині доступні абсолютно всі дані. Другим порівняльним критерієм є спосіб зберігання та обробки інформації. Традиційному підходу притаманний централізований спосіб зберігання та оброблення, який передбачає наявність обчислювального центру (ОЦ), 84 на який передається вихідна інформація, і звідки отримують результати оброблення користувачі.[2] Своєю чергою такий метод видається доволі трудомістким і складним процесом при спробі налагодження зв'язків, а також становить загрозу перевантаженості комп'ютера. Для Великих Даних характерний децентралізований спосіб зберігання інформації, який зумовлений появою персональних комп'ютерів (ПК). Такий підхід дає можливість розбити інформаційну базу на кілька розділених і дає змогу кожному клієнту користуватися своєю базою даних, яка може бути або частиною загальної інформаційної бази, або копією інформаційної бази в цілому. Напівструктурованість або неструктурованість бази Big Data є ще однією відмінністю від традиційних даних, для яких притаманна структурованість. Під структурованістю розуміється наявність встановлених зв'язків і відносин між елементами всередині системи, розподіл елементів системи за рівнями ієрархії. Big Data частково або повністю не володіє перерахованими вище ознаками. Ще одним критерієм порівняння Big Data і традиційних даних може бути генерація даних. Традиційні джерела даних завжди передбачають присутність людини, Великі Дані часто генеруються автоматично, без участі людини. Таким чином, Великі дані мають низку принципових відмінностей від традиційних даних, завдяки яким вони мають як низку переваг, так і містять певні труднощі[3].

Нині великі дані стали розглядатися як ефективний інструмент ухвалення державних рішень. Одним зі способів оперувати великими даними для регулювання соціально-економічних і політичних процесів є складання та аналіз офіційної статистики виключно на їхній основі та в комбінації з традиційними джерелами: реєстрами, опитуваннями, обстеженнями тощо. Таким чином, в останні кілька років на світовій арені спостерігається зростання інтересу до

застосування великих даних у державній статистиці. Головною перевагою великих даних у статистичних дослідженнях є своєчасне отримання об'ємних масивів інформації з найменшими фінансовими та часовими витратами. Великі дані здатні надати широкий спектр інформації з різних аспектів, які не зачіпаються традиційними джерелами. Крім цього висока частотність отримання інформації порівняно зі звичайним обстеженням забезпечує детальний розгляд процесів і розв'язання проблем на стадії їх зародження. Однак існують значні проблеми, що перешкоджають використанню великих даних у статистичних цілях. Нині актуальними є питання методології, якості, доступу до даних, законодавства, недоторканності приватного життя, управління та фінансування. Скрутними, зважаючи на технічні можливості, на поточний момент видаються збір і зберігання великих даних на серверах, тому дедалі більшої популярності в статистичних організаціях набувають хмарні технології. Для забезпечення доступності технологій великих даних широкому колу користувачів, необхідно забезпечити довіру суспільства до використання персональних даних. На сьогодні лише деякі країни розробили довгострокову стратегію використання великих даних. [4]Щоб звести до мінімуму ризику під час освоєння нових технологій, держави взаємодіють у рамках провідних світових дослідницьких центрів, таких як Статистична комісія ООН і Глобальна робоча група з великих даних в офіційній статистиці. У низці країн створено лабораторії та робочі групи для здійснення пілотних проектів, мета яких - визначити, наскільки великі дані придатні для використання як джерело офіційної статистики. Більшість проектів пов'язані з отриманням економічної та фінансової статистики, демографічної та соціальної статистики і статистики цін. Як головні джерела, що формують великі дані для подальшого їх застосування в статистиці, розглядають мобільний зв'язок і глобальну систему визначення координат (GPS), геопросторову інформацію та соціальні мережі. Частина цих даних перебуває не у відкритому доступі, а є власністю приватного сектору, тому постає потреба налагодити взаємодію між органами статистичних досліджень і компаніями. У деяких країнах мобільний зв'язок і GPS використовуються для збору 86 відомостей про мобільність населення впродовж дня та під час епідеміологічних спалахів, статистику туризму, допомагають при проведенні переписів населення, а також у статистиці транспорту. У Колумбії широко використовуються реєструвальні пристрої, які зберігають інформацію про транспортний засіб і відстежують його місце розташування. Практичне застосування зображень із супутників в Австралії забезпечує сільськогосподарську статистику використання землі та врожайності; поряд із цим у Китаї, Колумбії та Мексиці інформація супутників оптимізує екосистемний облік, а в Бразилії великі дані метеорологічних станцій є фундаментом для статистики водних ресурсів. На додаток до цього, на застосуванні супутникових знімків базується моніторинг посівів незаконних культур. Соціальні мережі, вебфоруми та блоги використовуються для соціальної статистики: прикладом слугують Нідерланди, які практикують застосування масивів даних з публічних повідомлень користувачів 'Twitter' і

'Facebook' для оцінки настрою споживачів, а також Італія та Китай, які використовують мережу Інтернет для статистики праці.[5]

Таким чином, уже є низка позитивних прикладів використання великих даних в офіційній статистиці, їхнє застосування призводить до ефективного і своєчасного розв'язання актуальних завдань економіки, політики та соціальної сфери. Однак слід прискорити процес упровадження нових технологій і методів збору та обробки даних на основі модернізації всієї статистичної системи, науково-методологічної, технічної та нормативно-правової бази.

Список використаних джерел

1. Hammer C. L., Kostroch D. C., Quirós G. STA Internal Group. Big Data: Potential, Challenges, and Statistical Implications. IMF Staff Discussion Note no. 2017. 17/06. September
2. A Suggested Framework for the Quality of Big Data. Deliverables of the UNECE Big Data Quality Task Team. December 2014. URL: <http://www1.unece.org/stat/platform/display/bigdata>.
3. Vale S. UNECE Task Team on Big Data. Classification of Types of Big Data. 2013. 27 June. URL: <https://statswiki.unece.org/display/bigdata/Classification+of+Types+of+Big+Data>.
4. Eurostat Big Data Task Force. Item 8 of the agenda. ESS Big Data Action Plan and Roadmap
5. Work Programme Objective 11.1. ESSC 2014/22/8/EN26/09/2014. 22nd Meeting of the European Statistical System Committee, 26 September 2014, Riga (Latvia).