

ПРОБЛЕМИ ТА ПЕРСПЕКТИВИ ІМПЛЕМЕНТАЦІЇ ВЕЛИКИХ ДАНИХ В ОФІЦІЙНУ СТАТИСТИКУ УКРАЇНИ В УМОВАХ ВОЄННОГО СТАНУ

Осауленко Олександр Григорович,

доктор наук з державного управління, професор,
член-кореспондент НАН України, заслужений економіст України,
ректор;

Горобець Олена Олександрівна,

кандидат економічних наук, доцент
кафедри статистики, інформаційних технологій
та математичних методів в економіці;
Національна академія статистики, обліку та аудиту

У зв'язку з повномасштабним вторгненням Російської Федерації на територію суверенної України Президентом було підписано Указ «Про введення воєнного стану в Україні» та низку інших законодавчих актів, у тому числі Закон України «Про захист інтересів суб'єктів подання звітності та інших документів у період воєнного стану або стану війни», яким установлюється, що на період воєнного стану та три місяці після його скасування відміняється подання будь-якої звітності в документальній і електронній формах, включаючи статистичну і фінансову звітність [1].

Тому окрім непоправних втрат для України та кожного її жителя, війна завдала також значної шкоди системі державного управління країни, в тому числі й органам офіційної статистики, які призупинили збирання більшої частини первинних статистичних даних та оприлюднення відповідної зведеної статистичної інформації.

На сьогодні офіційна статистика України практично позбавлена всіх основних джерел статистичних даних, а саме: первинних даних, отримуваних за допомогою статистичних спостережень та обстежень (від суб'єктів комерційної діяльності, фізичних осіб, домашніх господарств, а також державних структур, суспільних об'єктів, тощо), вторинних даних з адміністративних джерел, а також інформації, що збиралася спеціалізованими урядовими агенціями.

Зважаючи на те, що наявність якісних статистичних даних є необхідною складовою нормального функціонування та забезпечення прогресу у всіх сферах життєдіяльності, у тому числі в державному управлінні, а офіційна статистична інформація – частиною глобального інформаційного простору, офіційна статистика України нагально потребує залучення альтернативних джерел даних та застосування інноваційних методів для їхнього оброблення,

з метою заповнення прогалин в середовищі традиційних даних і забезпечення збереження національної статистичної системи.

Альтернативними джерелами даних для офіційної статистики є великі дані. Однак, попри свою беззаперечну своєчасність та цінність, усе ще існують деякі проблеми, які залишаються актуальними для цього середовища.

Використання великих даних для статистичних проблем потребує розробки принципово нової політики з управління інформацією, її захисту, стосунків із респондентами, а також процесу професійної підготовки статистичного персоналу.

В частині методології проблемою використання великих даних у статистиці 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. У той же час великі дані формуються спонтанно й поза межами реєстрів, їхня хаотичність, неструктурованість та динамічність ускладнює статистичні процеси, які є сталими для концепції офіційної статистики. При цьому, існуючі 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.

Варто додати, що 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.

Зараз основні зусилля статистиків України повинні зосередитися не лише на розбудові потенціалу великих даних, обміні знаннями на міжнародному рівні, але, і в першу чергу, на розвитку партнерства з приватним сектором з метою практичного використання отриманих знань та навичок у частині залучення та опрацювання цих нових видів даних. В умовах

війни, великі дані мають стати доповненням до традиційних статистичних даних або ж стати їх альтернативою ще з більшою деталізацією.

На сьогодні існує велика кількість пропозицій щодо використання великих даних в офіційній статистиці. В процесі імплементації великих даних в офіційну статистику, слід обов'язково дотримуватись Фундаментальних принципів офіційної статистики (2014) [2]. При цьому корисним буде використання принципів аналітики великих даних [3] та 12 правил Кодда [4], які уже стали класичними при побудові систем керування базами даних.

З огляду на зазначене та враховуючи необхідність використання великих даних офіційною статистикою, пропонуються наступні принципи роботи з великими даними:

1. Процес збирання великих даних повинен гарантувати безпеку та збереження їх конфіденційності, а також бути забезпеченим перевіркою на надійність та якість даних.
2. Необхідно забезпечити інтелектуальний супровід під час збирання та аналізу великих даних, оприлюднення та поширення висновків, що ґрунтуються на результатах їх обробки.
3. Норми та закони щодо збирання, оброблення, аналізу великих даних, а також оприлюднення та поширення висновків на основі результатів їх аналізу повинні бути узгоджені та оприлюднені у форматі чотирьохсторонньої групи: «держава-наука-бізнес-статистика».
4. Забезпечення гарантованого вільного доступу для членів групи «держава-наука-бізнес-статистика» до державних та приватних сховищ великих даних.
5. Напрацювання статистичними відомствами методологій та рекомендацій щодо збирання, опрацювання та аналізу великих даних, з урахуванням специфіки кожної їхньої окремої групи.
6. Забезпечення узгодження та оприлюднення алгоритмізації основних процесів збирання, оброблення та аналізу великих даних.

Дотримання вищезазначених принципів дозволить напрацьовувати та надалі дотримуватися єдиних підходів на шляху від «неординарних пропозицій щодо використання великих даних» до «адекватних управлінських рішень щодо імплементації великих даних у різні сфери життєдіяльності».

При пошуку рішень щодо використання великих даних у статистиці важливе значення має вибір правильних інструментів для роботи з ними, перш за все у сфері ІТ-технологій із використанням методів математичного моделювання.

Досить ефективним засобом для процесу вилучення потрібних даних є веб-скрайпінг (Web scraping). Доджусі М. та Сетінкайя-Рандел М. зазначають, що «веб-скрайпінг – це процес програмного вилучення даних з Інтернету та перетворення їх у структурований набір даних, який дозволяє збирати більші обсяги даних за коротший проміжок часу та в автоматизований спосіб з максимальною мінімізацією помилок» [5]. Веб-скрайпінг доцільно застосовувати для формування індексу споживчих цін (використовуючи сайти магазинів), аналізу ринку працевлаштування (сайти вакансій), аналізу ринку нерухомості (сайти оренди) та ін. Однак використання цього засобу має свої недоліки, так, Сумма Д. та ін. відзначають деякі проблеми, які можуть спіткати статистиків під час веб-скрайпінгу. Зокрема, онлайн-дані цін на товари із платформ електронної комерції можуть не охоплювати повного переліку товарів або послуг. Технічні проблеми включають часті зміни в структурі веб-сайту, необхідність в оновленнях сканерів або розробленні окремих сканерів для різних веб-сайтів, можливість автоматичного блокування високочастотного веб-скрайпінгу, що вимагає співпраці та партнерства з власниками веб-сайтів. Також автори особливу увагу у процесі веб-скрайпінгу акцентують на якості даних [6].

Незважаючи на це, досвід Національного інституту статистики в Італії у використанні Web scraping для розробки нових стратегій інтеграції даних традиційних опитувань, адміністративних органів та інноваційних джерел, таких як великі дані, свідчить про позитивні результати у зменшенні статистичного навантаження на респондентів, дозволяючи при цьому формувати релевантні бази даних та створювати нові форми статистичних реєстрів [7].

Варто зазначити, що завдяки стрімкому розвитку цифровізаційних процесів з'явилося багато різноманітних програмних продуктів для аналізу даних, які можуть бути використані для потреб статистики, зокрема: Actian (засіб який дозволяє зберігати дані та забезпечує їхню підготовку для подальшого аналізу), Ambari (інструмент для управління кластерами), Avro (система серіалізації даних), Apache Kafka (платформа для опрацювання даних в режимі реального часу), Hadoop (набір утиліт, бібліотек та фреймворків для ефективного обробки великих даних), HIVE (інфраструктура сховища даних, яка забезпечує їхню агрегацію) та ін.

У контексті цієї роботи, зазначимо, що війна в Україні характеризується активним залученням інформаційно-комунікативних технологій, серед яких варто виокремити технології штучного інтелекту, які функціонують за допомогою вдало розроблених та налаштованих алгоритмів обробки великих даних.

З огляду на зазначене, варто акцентувати увагу на тому, що в Україні популярністю користуються розумні технології на базі штучного інтелекту,

дані яких входять до однієї із зазначених вище глобальних груп екосистеми великих даних. Здебільшого розумні технології застосовуються українцями на побутовому рівні – ними оснащені будинки, авто, прибудинкові території. За допомогою таких технологій Міністерством цифрової трансформації України було зібрано безліч відео- та фото-доказів порушення законів та звичаїв війни, закріплених договорами Женевської конвенції [8].

Ще одним невід’ємним джерелом великих даних є соціальні мережі. Їхня популярність сумісно з технологіями штучного інтелекту дозволяють знаходити та ідентифікувати осіб із ворожої армії, які знищували мирне населення в українських містах [9].

Такий успішний досвід використання великих даних в умовах війни підкреслює цінність та своєчасність таких даних, хоча попри це ще залишаються відкритими зазначені вище проблемні питання щодо їхньої імплементації, зокрема й в офіційну статистику.

Список використаних джерел

1. Про захист інтересів суб’єктів подання звітності та інших документів у період воєнного стану або стану війни: Закон України. URL: <https://zakon.rada.gov.ua/laws/show/2115-20#Text>

2. Fundamental Principles of Official Statistics (A/RES/68/261 from 29 January 2014). Retrieved from: <https://unstats.un.org/unsd/dnss/gp/fundprinciples.aspx>

3. Chen, C.L. Philip and Zhang, C.-Ya. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Science*, 275, pp. 314-347.

4. Codd E. F., (1985). Is Your DBMS Really Relational? *Computer World*.

5. Dogucu, M., & Çetinkaya-Rundel, M., (2021). Web Scraping in the Statistics and Data Science Curriculum: Challenges and Opportunities. *Journal of Statistics and Data Science Education*. Vol. 29. Pp. 112-122.

6. Summa, D., Bianchi, G., Consalvi, M., Gentili, B., Pancella, F. and Scalfati, F., (2019). Using Big Data for Official Statistics: Web Scraping as a Data Source for Statistical Business Registers (SBRs). *NTTS2019*. Retrieved from: https://coms.events/ntts2019/data/abstracts/en/abstract_0007.html

7. Polidoro, F., Giannini, R., Conte R. L., Mosca, S. & Rossetti, F., (2015). Web scraping techniques to collect data on consumer electronics and airfares for Italian HICP compilation. *Statistical Journal of the IAOS*, 31(2). pp. 165-176.

8. Geneva Convention (IV) on Civilians, 1949. *Treaties, States parties, and Commentaries*. 23 March 2010. Retrieved from: <https://ihl-databases.icrc.org/applic/ihl/ihl.nsf/Treaty.xsp?documentId=AE2D398352C5B028C12563CD002D6B5C&action=openDocument>

9. Михайло Федоров розповів про розпізнавання облич російських окупантів. URL: <https://www.kmu.gov.ua/news/mihajlo-fedorov-pro-rozpoznavannya-oblich-rosijskih-okupantiv>

РОЗВИТОК ОСВІТНЬОЇ СФЕРИ УКРАЇНИ В УМОВАХ ВОЄННОГО СТАНУ

Панасенко Інна Вікторівна,

кандидат економічних наук, доцент,

доцент кафедри математичного моделювання та статистики;

Київський національний економічний університет імені Вадима Гетьмана

Роль освіти для соціально-економічного розвитку держави важко переоцінити. Факт позитивного впливу є доведеним і, враховуючи це, багато країн світу здійснюють пріоритетне фінансування сфери освіти, збільшують частку ВВП країни на утримання та розвиток освітньої сфери, підвищують рівень матеріального заохочення працівників сфери освіти, реалізують програми розвитку сфери освіти на різних рівнях. Розвиток науки та освіти в Україні наразі характеризується уповільненими темпами, що обумовлено загальносвітовою тенденцією через пандемією COVID-19, воєнними діями та відповідним падінням національної економіки.

Сучасна система освіти в Україні характеризується низкою проблем:

- ❑ Низьким рівнем фінансування та відсутністю механізмів впровадження альтернативних джерел фінансування сфери освіти. Розрахунки свідчать, що в середньому щорічно за період 2011-2020 рр. витрати зведеного бюджету України на освіту скорочувались на 146,7 млн. дол. США, що у відносному вираженні склало 1,44 % падіння щорічно. Загальноприйняте світове порівняння витрат бюджету на освіту та величини ВВП свідчить, що їх питома становила у 2011 р. – 6,6 %, у 2019 р. – 6,0 %, тобто скорочення склало 0,6 % [1]. Відповідно, фінансування сфери освіти було на досить низькому рівні і прогнозувати його збільшення зараз неможливо.
- ❑ Високим ступенем зносу основних засобів у сфері освіти, який за період 2017 – 2020 рр. збільшився на 10,3 % та у 2020 р. становив 52,7% [2]. Напрямок покращення ситуації може стати злиття або укрупнення закладів однакового спрямування, що дозволить акумулювати матеріальні й фінансові ресурси. Також можливе запровадження схем державно-приватного партнерства в управлінні закладами освіти, які є визнаними у світовій практиці [3, 4]. Незважаючи на