

Аналіз великих даних (big data) для підвищення ефективності регіонального управління

Олексій Туболєв

*здобувач освітньої програми «Комп'ютерні Науки», гр. КН-25,
ВНЗ «Університет економіки та права «КРОК», м. Київ, Україна,
e-mail: TubolievOY@krok.edu.ua*

Денис Балдик

*кандидат економічних наук, доцент,
завідувач кафедри інформаційного менеджменту,
математики та статистики,
ВНЗ «Університет економіки та права «КРОК», м. Київ, Україна,
e-mail: baldykd@krok.edu.ua,
ORCID: 0000-0003-2670-1190*

У сучасних умовах цифрової трансформації державного управління обсяги інформації, що накопичуються на регіональному рівні, зростають із безпрецедентною швидкістю. Традиційні інструменти обробки даних уже не дозволяють ефективно виявляти закономірності та приймати обґрунтовані управлінські рішення. Саме тому аналіз великих даних (Big Data) стає ключовим інструментом підтримки регіонального розвитку, прогнозування соціально-економічних процесів та оптимізації державних сервісів. Використання Big Data здатне забезпечити більш точне розуміння потреб населення та підвищити ефективність управлінських стратегій, проте водночас створює низку методологічних та обчислювальних викликів, які потребують системного вирішення.

«Big Data» це термін, який використовується для позначення наборів даних, що є надто великими або складними для ефективного обробки традиційним програмним забезпеченням. Дані з великою кількістю записів (рядків) забезпечують вищу статистичну потужність, тоді як дані з більшою складністю (більшою кількістю атрибутів або стовпців) можуть призводити до вищого рівня хибних відкриттів [1].

Актуальність проблеми обумовлена постійним зростанням обсягу даних. Цей ріст викликаний вибуховим збільшенням як машинно-генерованих даних (оскільки все більше датчиків та пристроїв збирають і передають інформацію), так і високою залученістю людини до соціальних мереж. Особливо цей процес прискорився останнім часом, коли пандемія змусила перевести більшість видів діяльності в онлайн-режим. У наші дні майже кожна дія, слово чи клік – створює дані, які необхідно збирати, зберігати та досліджувати [2].

Аналіз великих даних (Big Data) відкриває для регіонального управління значні можливості, зокрема для виявлення прихованих закономірностей у житті населення та врахування інтересів окремих субпопуляцій, що було неможливо при використанні малих вибірок. Однак впровадження Big Data створює унікальні статистичні та обчислювальні виклики. Головна проблема полягає в

тому, що традиційні методи аналізу, які добре працюють із даними малої чи середньої розмірності, не здатні впоратися з масивними розмірами вибірки та високою розмірністю сучасних даних. Це вимагає повної зміни парадигми як в обчисленнях, так і в статистичному підході [3].

Ключова проблема при аналізі великих даних у регіональному управлінні – це висока розмірність, яка породжує деякі специфічні виклики.

По-перше виникає «накопичення шуму» (noise accumulation). Коли управлінська модель враховує тисячі показників, похибки оцінювання кожного з них підсумовуються. Це може призвести до ситуації, коли традиційні класифікаційні моделі (наприклад, для прогнозування соціальних потреб) працюють не краще, ніж випадкове вгадування, оскільки шум від нерелевантних показників домінує над корисним сигналом.

По-друге, висока розмірність створює «хибні кореляції» (spurious correlations). Це феномен, при якому два показники, що науково ніяк не пов'язані (наприклад дані з соціальних мереж та рівень споживання комунальних послуг), можуть демонструвати високу вибіркочну кореляцію виключно через велику кількість змінних. Опора на такі хибні зв'язки призводить до невірних статистичних висновків та, як наслідок, до прийняття помилкових науково необґрунтованих управлінських рішень.

По-третє, виникає проблема «випадкової ендогенності» (incidental endogeneity). Це ситуація, коли деякі змінні в моделі корелюють із похибкою через приховані або не враховані фактори. У високорозмірних наборах даних це трапляється випадково, оскільки кількість змінних настільки велика, що ймовірність появи таких кореляцій зростає. Внаслідок цього статистичні оцінки стають зміщеними, а прогнози – ненадійними. Наприклад, якщо модель прогнозує ефективність субсидій, але не враховує приховані фактори (неформальні доходи населення тощо), то деякі показники, як-от активність у соцмережах, можуть випадково корелювати з похибкою.

Таким чином, стрімке зростання обсягів гетерогенних даних, що генеруються як машинами, так і людьми в соціальних мережах, робить традиційні методи аналізу недієздатними. Спроби застосувати старі підходи до Big Data в регіональному управлінні створюють серйозні статистичні виклики, зокрема накопичення шуму (noise accumulation), хибні кореляції (spurious correlations) та випадкову ендогенність (incidental endogeneity). Ігнорування цих специфічних проблем, а також пов'язаних з ними обчислювальних труднощів, неминуче призводить до невірних статистичних висновків і, як наслідок, до прийняття помилкових та науково необґрунтованих управлінських рішень.

Ключові слова: Big Data; регіональне управління; високорозмірні дані; статистичні виклики; цифрова трансформація; аналіз даних.

Список використаних джерел

1. *Big Data analysis methods and technics. (n.d.). IT-UA. Retrieved November 11, 2025, from <https://www.it.ua/en/knowledge-base/technology-innovation/big-data-bolshie-dannye>.*
2. *Barila, A., Danubianu, M., & Turcu, C. O. (2021). Big data: Actuality and challenges. Journal of Danubian Studies and Research, 11(1). Retrieved from <https://dj.univ-danubius.ro/index.php/JDSR/article/view/1271>.*
3. *Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. National Science Review, 1(2), 293–314. <https://doi.org/10.1093/nsr/nwu002>.*