

Список використаних джерел

1. Osaulenko O., Horobets O. (2023). Using Big Data by Ukrainian official statistics when martial law applies: problems and solutions. *Statistics in Transitions. New series*. vol. 24, 1, pp. 29-43.
2. Індекс споживчих цін. URL: https://ukrstat.gov.ua/operativ/prostat/pererahunki/ind_cin/isc.pdf
3. What is Web Scraping and How Can It Benefit Your Business? (2023). Retrieved from: <https://dataforest.ai/blog/what-is-web-scraping-and-how-can-it-benefit-your-business>

КРАУДСОРСИНГ ДАНИХ – СУТНІСТЬ ТА ПЕРЕВАГИ ЗАСТОСУВАННЯ

Гринчак Наталія Анатоліївна,
кандидат економічних наук, доцент,
доцент кафедри менеджменту,
маркетингу та публічного управління;
Національна академія статистики, обліку та аудиту

Краудсорсинг стає все більш популярним методом збору даних. Використовуючи можливість залучення великої групи людей до вирішення певної проблеми, компанії та організації можуть швидко й дешево збирати великі обсяги даних. Як правило, краудсорсинг реалізується через онлайн-спільноти, що дозволяє збирати інформації від великої спільноти користувачів, незалежно від їх місця розташування.

Краудсорсинг даних — це процес отримання даних із великої кількості джерел. Прикладом краудсорсингу даних є використання онлайн-анкет для збору відгуків від клієнтів. Краудсорсинг даних можна використовувати для покращення обслуговування клієнтів, розуміння потреб клієнтів і прийняття кращих бізнес-рішень.

Стрімке зростання популярності краудсорсингу пояснюється його численними перевагами [1]:

– Краудсорсинг даних може надати підприємствам точні та своєчасні дані. Дані є гнучкими та можуть бути змінені відповідно до потреб бізнесу. Підприємство може платити за використання даних або отримувати сповіщення в реальному часі, коли трафік перевантажений.

– Краудсорсинг даних може допомогти пришвидшити процес пошуку потрібних даних, дозволяючи великій кількості людей швидко та дешево надавати дані. Це гарантує швидке виконання завдань з обробки даних із дотриманням високих стандартів якості.

– Краудсорсинг дає компаніям доступ до великої кількості кваліфікованих збирачів даних з усього світу. Високоякісний набір даних важливий для успіху

моделі штучного інтелекту, а дані можна легко та дешево збирати за допомогою краудсорсингу.

– Дані краудсорсингу можуть допомогти швидше приймати рішення, надаючи гнучкий спосіб збору даних у режимі реального часу. Це можна використовувати для легшого та своєчасного виявлення помилок. Наприклад, сповіщення про дорожній рух можна надсилати в режимі реального часу на основі попередньо вибраних порогових значень або історичних тенденцій.

– Краудсорсинг даних можна використовувати для отримання нових ідей щодо економічно ефективних рішень. Це дешевший і доступніший спосіб отримати рішення складних проблем, ніж традиційні методи. Краудсорсинг не обмежується виключно технічними та складними проблемами – його також можна використовувати для досліджень і розробок (НДДКР).

– Дозволяє пришвидшити розробку продукту, дозволяючи швидше отримувати зворотний зв'язок і краще розуміти потреби користувачів. Завдяки краудсорсингу даних компанії можуть отримати відгуки та вхідні дані від великої кількості користувачів за короткий проміжок часу. Це можна використовувати, щоб покращити продукти та зробити їх більш зручними для користувача. Крім того, краудсорсинг даних можна використовувати для розуміння настроїв клієнтів і відстеження ефективності продукту.

– Дозволяє покращити обслуговування клієнтів, збираючи відгуки клієнтів про їхній досвід. Краудсорсинг даних також може допомогти визначити закономірності та тенденції у взаємодії з клієнтами, що може допомогти покращити якість обслуговування клієнтів.

Платформи краудсорсингу даних зазвичай дозволяють користувачам реєструватися та виконувати прості завдання в обмін на винагороду. Ці завдання можуть включати відповіді на запитання, надання відгуків або оцінювання продуктів. Потім зібрані дані використовуються компанією чи організацією, яка керує платформою, щоб покращити розуміння певної теми чи проблеми.

Яскравим прикладом застосування краудсорсингу в проведенні статистичних досліджень є досвід Статистичного управління Канади. У рамках модернізації Статистичного управління Канади краудсорсинг став інноваційним способом збору цінної інформації для статистичних цілей. Так як, краудсорсинг передбачає збір інформації від великої спільноти користувачів, він спирається на принцип, що окремі громадяни є експертами у своєму місцевому середовищі. Крім того, краудсорсингові опитування дозволяють порівнювати та перевіряти дані з іншими джерелами додаткових даних, щоб гарантувати якість результатів.

Статистичне управління Канади нещодавно приступило до інноваційного проекту з краудсорсингу цін на сектор канабісу до легалізації. Ця ініціатива продовжує збирати інформацію про відносно новий ринок і допомагає контролювати ціни в конфіденційний та ненав'язливий спосіб. Статистичне управління Канади попросило широку громадськість повідомити ціну їхньої останньої покупки сушеного канабісу; ця інформація збирається анонімно та поширюється серед громадськості через спеціальну програму та веб-сторінку [2].

Ще одним цікави прикладом є пілотний краудсорсинговий проект OpenStreetMap (OSM), який був завершений у березні 2018 року, збирав географічну інформацію шляхом нанесення на карту контурів будівель у районах Оттави, Онтаріо та Гатіно, Квебек. Мережа та досвід цього пілотного проекту допомогли запуснути ініціативу Building Canada 2020 (BC2020), спрямовану на відображення всіх слідів будівель Канади на OSM до 2020 року [2].

OpenStreetMap використовує топологічну структуру даних із чотирма основними елементами (також відомими як примітиви даних):

Вузли — це точки з географічним положенням, які зберігаються як координати (пари широти та довготи) відповідно до WGS 84. За межами їхнього використання способами, вони використовуються для представлення об'єктів карти без розміру, наприклад визначних місць або гірських вершин.

Шляхи — це впорядковані списки вузлів, що представляють полілінію або, можливо, багатокутник, якщо вони утворюють замкнутий цикл. Вони використовуються як для представлення лінійних об'єктів, таких як вулиці та річки, так і для території, таких як ліси, парки, стоянки та озера.

Зв'язки — це впорядковані списки вузлів, шляхів і зв'язків (які разом називаються «членами»), де кожен учасник може додатково мати «роль» (рядок). Зв'язки використовуються для представлення зв'язку існуючих вузлів і шляхів. Приклади включають обмеження поворотів на дорогах, маршрути, які охоплюють кілька існуючих шляхів (наприклад, міжміську автомагістраль), а також ділянки з ямами.

Теги є парами ключ-значення (обидва довільні рядки). Вони використовуються для зберігання метаданих про об'єкти карти (таких як їх тип, ім'я та фізичні властивості). Теги не є окремими, але завжди прикріплені до об'єкта: до вузла, шляху чи відношення. [3] Рекомендована онтологія функцій карти (значення тегів) підтримується у вікі. Нові схеми тегів завжди можна запропонувати шляхом всенародного голосування за письмову пропозицію у вікі OpenStreetMap, проте немає вимоги дотримуватися цього процесу. Станом на червень 2017 року використовується понад 89 мільйонів різних типів тегів.

Примітиви даних OpenStreetMap зберігаються та обробляються в різних форматах. Сервер OpenStreetMap використовує базу даних PostgreSQL з однією таблицею для кожного примітиву даних, а окремі об'єкти зберігаються у вигляді рядків. [4] З цього створюється декілька дамів бази даних, які доступні для завантаження. Повний дамп називається planet.osm. Ці дампи існують у двох форматах: один використовує XML, а другий – двійковий формат буфера протоколу (PBF).

Список використаних джерел

1. Data Crowdsourcing – How to Crowdfund Data Successfully. URL: <https://www.clickworker.com/customer-blog/data-crowdsourcing/>
2. Crowdsourcing. URL: <https://www.statcan.gc.ca/en/our-data/where/crowdsourcing>

3. Foody G., See L., Fritz, S., Mooney P., Olteanu-Raimond A.-M., Fonte C. C., Antoniou V. Mapping and the Citizen Sensor. Ubiquity Press. URL: <http://www.jstor.org/stable/j.ctv3t5qzc>

4. Official website OpenStreetMap. URL: <https://www.openstreetmap.org/>

ВИБІР ФАКТОРІВ ПРИ СТАТИСТИЧНОМУ ДОСЛІДЖЕННІ РИЗИКУ КІБЕРАТАК

Густера Олег Михайлович,

кандидат економічних наук, асистент кафедри статистики,
інформаційно-аналітичних систем та демографії;
Київський національний університет імені Тараса Шевченка

Управління ризиком може базуватись на використанні даних з різних інформаційних джерел, одним з найбільш поширених з яких є статистичний аналіз що дозволяє отримати відомості про фактори що впливають на ризик, ступінь на напрямок зв'язку між різними факторами, тісноту зв'язку між різними факторами та можливість побудування математичної моделі що буде використовуватись для отримання достовірного прогнозу.

Основні параметри ризикової ситуації які можуть бути корисними для прийняття рішення та управління ризиком на прикладі ризику кібератаки для підприємства – кількість атак (загальна або у розрізі конкретної галузі підприємства) та відповідно до цього масиву кількості наслідки що можуть бути при настанні ризикової події (максимальний, мінімальний та середній збиток). Використання лише середнього значення збитків від кібератаки не може повністю охоплювати ступінь ризику для даної події. Так, наприклад, деякі атаки не приносять суттєвих збитків тому що направлені на виявлення незакритих портів або інших незахищених місць та використовуються зловмисниками не для комерційних цілей. В той же час, інші атаки при їх невеликій кількості можуть призводити до суттєвих збитків – видалення або спотворення інформації, викрадання конфіденційних даних. Найчастіше для розділення на більш або менш суттєві ризики використовують категоризацію ризиків за ступенем потенційних збитків або не враховують атаки які не призводять до збитків.

При цьому також потрібно враховувати загальну кількість пристроїв що підключені до мережі та потенційно можуть бути атакованими, так як вони визначають генеральну сукупність та відносно до неї може бути визначена імовірність атаки з урахуванням фактичної кількості кібератак.

До параметрів що можуть бути використані при статистичному дослідженні оцінки ризику можна віднести:

- загальна кількість кібератак,
- сфера діяльності підприємства,
- розміщення інформаційної інфраструктури підприємства,
- наявність захисту від найбільш поширених атак,