

ВИКОРИСТАННЯ МЕТОДІВ DATA SCIENCE ДЛЯ ПРОЕКТІВ СОЦІАЛЬНОГО СПРЯМУВАННЯ

Актуальність теми дослідження. В епоху цифрових технологій з прогресивним накопиченням та розвитком інформаційних структур та пов'язаних з ними соціально-гуманітарних преображень, постає закономірне питання про вплив великих потоків даних на методологію та інструментарій продукування знання. Все швидшими темпами початковий фокус сучасних медіа-технологій спрямовується на першу частину спектра дані-інформація-знання. Наука, що оперувала величезним арсеналом інструментарію змушена переорієнтуватись на новий формат структури науки та виробництва знання. Наука під натиском величезних потоків інформації перетворюється на своєрідний цифровий конвеєр продукування знання та технологій. Для оцінки масштабу обсягів цифрових даних Р. Вільямс склав таблицю відповідності між одиницями виміру даних і звичними для людського сприйняття об'єктами, переведеними в цифрову форму. До 2010 р. обсяг всіх створених у світі цифрових даних становить 1,2 зетабайт, що можна представити як стопку DVD-дисків від Землі до Місяця і назад [1].

Мета роботи – Розробка застосунку, який на основі реальних даних зможе візуалізувати, оброблювати, сортувати та визначати проблемні місця в різних частинах міста. Програмний продукт повинен, оброблювати різного роду виклики від реальних людей, які зареєструвалися в даній мережі. Користувачі будуть повідомляти про небезпечні місця, випадки аварій, насильства тощо. Сервіс буде оброблювати ці дані і вже з певною базою викликів та повідомлень візуалізувати карту небезпек міста. Що дасть змогу соціальним структурам знайти більш точну причину небезпек в даній частині міста та незабаром вирішити ці проблеми.

Проблемна ситуація. Відкритий обмін інформацією, доступний в мережі між дослідниками, має основоположне значення для науки, для збільшення швидкості досліджень і зростання визнання вчених, стає можливим обмін інформацією між дослідниками, який охоплює ряд ініціатив з метою видалення перешкод для доступу до даних і опублікованих документів. Медіа-інфраструктурі характерні підходи, парадигма, моделі, орієнтовані на цифрові дані. Так як даних багато, то існує проблематика правильної обробки цих даних з різних джерел.

Результат дослідження. Аналіз даних є невід'ємною частиною всіх прикладних досліджень та вирішення проблем в промисловості. Найбільш фундаментальні підходи аналізу даних — візуалізація (гістограми, точкові ділянки, ділянки поверхні, дерево карт, паралельно координовані ділянки і т. д.), статистика (гіпотеза тест, регресія, СПС і ін), видобуток даних (асоціації гірничодобувної промисловості, і т. д.) і методи машинного навчання (кластеризація, класифікація, дерева рішень, і т. д.). Серед усіх цих підходів, візуалізації інформації, або, іншими словами, візуального аналізу даних, є той, який спирається в основному на пізнавальні навички аналітиків, а також дозволяє розкриття неструктурованих дієвих ідей, які обмежені тільки людською фантазією та творчістю. Аналітик не повинен застосовувати різні витончені методи, щоб мати можливість інтерпретувати візуалізацію даних. Візуалізація інформації — це також схема гіпотез, які можуть бути, і, як правило, є попередниками більш аналітичного або формального аналізу на кшталт статистичних гіпотез. [2]

Висновки. Будь-який аналіз впливу «великих даних» необхідно розпочати з визначення того, що власне означає цей термін, який часто використовується, але зазвичай не всі його розуміють. Він стосується насамперед величезної кількості даних, які постійно збираються за допомогою пристроїв і технологій, таких як платіжні картки та картки лояльності клієнтів, Інтернету та соціальних медіа і все частіше, через датчики Wi-Fi та електронні мітки. Велика

частина цієї інформації є неструктурованою – тобто це дані, які не відповідають певній, заздалегідь визначеній, послідовності. У свою чергу, дані (Data), згідно з визначенням «Енциклопедії кібернетики» за редакцією академіка В. Глушкова [18, с. 230], – це факти та ідеї, подані у формалізованому вигляді, завдяки чому їх можна передавати чи обробляти за допомогою певного процесу (й відповідних технічних засобів). Тобто «великі дані» – це наступний етап розвитку даних. Забезпечення якості даних є ключовим фактором. Процес рівно настільки якісний, наскільки якісні дані, які в ньому використовуються. Оскільки автоматична обробка даних є однією з основних прикладних задач кібернетики, то прописування кодів команд із набором даних повинно перетворюватися у візуальне відображення зрозумілого змісту та форми для пересічного користувача статистичної інформації.

Ключові слова: аналіз даних, візуалізація даних, небезпечні місця.

Література:

1. *Наука про дані* - [Електронний ресурс] / Режим доступу до даних: https://uk.wikipedia.org/wiki/Наука_про_дані – 23.05.2018 р. – Заголовок з екрану.
2. *Візуалізація інформації* - [Електронний ресурс] / Режим доступу до даних: https://uk.wikipedia.org/wiki/Візуалізація_інформації – 23.05.2018 р. – Заголовок з екрану.