

можливості для особистої аналітики. Для покращення роботи додатку, будь-хто може стати “донором” детальної інформації про себе заради науково-дослідницьких цілей. Тож цей інструмент буде давати ще більше уявлення про те, як користувач проводить свій час в мережі Інтернет.

Список літературних джерел

1. Wolfram|Alpha: веб-сайт. URL: <https://uk.wikipedia.org/wiki/WolframAlpha>
2. Анализ данных мира Facebook. вебсайт. URL: <https://habr.com/ru/company/wolfram/blog/262445/>

УДК 004.82:004:85

*Войтко Б. С., студент 1 курсу спеціальності
122 «Комп'ютерні науки»*

*Нескородєва Т. В., канд. техн. наук, доцент,
завідувач кафедри комп'ютерних наук та
інформаційних технологій*

ДЖЕРЕЛА ДАНИХ ДЛЯ ШТУЧНОГО ІНТЕЛЕКТУ ТА ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ

Донецький національний університет імені Василя Стуса, м. Вінниця, Україна

Рішення на основі статистичного навчання поступово починають входити в наше повсякденне життя, кардинально змінюючи звичні дії та речі. Вже сьогодні статистичне навчання застосовують у таких областях як комп'ютерний зір, розпізнавання мовлення, біоінформатика, бейсбол, підприємництво та інші [1]. Тисячі компаній використовують дані можливості для покращення продажів товарів. При цьому цілком очевидно, що на горизонті найближчих п'яти років статистичне навчання з використанням штучного інтелекту і ML виведуть на принципово новий рівень автоматизацію бізнес-процесів у великих і середніх компаніях. Але ці інновації неможливі без використання датасетів.

Датасет – колекція з логічних записів, що зберігаються у вигляді кортежу. Дані можуть бути представлені в різній формі. Корисно мати базову класифікацію різних типів даних, щоб відразу ж визначати тип нових даних і використовувати відповідний метод аналізу. Набір даних складається з результатів спостережень об'єктів, які зазвичай включають реєстрацію однієї і тієї ж інформації для кожного об'єкта. Ми визначаємо елементарні одиниці як самі об'єкти (наприклад: компанії, люди, домогосподарства, міста, телевізори), щоб відрізнити їх від результатів вимірювань або спостережень (наприклад: обсяги продажів, вага, дохід, населення, розмір). Можна вказати чотири основних способи класифікації наборів даних [2]:

- За кількістю порцій інформації (змінних) для кожної елементарної одиниці.
- За типом вимірювання (числа або категорії) для кожного спостереження.
- По тому, важлива чи ні впорядкованість в часі записів про результати вимірювань.
- По тому, чи збиралася інформація спеціально для цього аналізу або дані збиралися раніше кимось іншим для своїх потреб.

Набір даних можна порівняти з файлом, але на відміну від файлу набір даних є одночасно і каталогом, і файлом файлової системи, і не може містити в собі інші набори. Існує близько 72 форматів збереження наборів даних, але найбільшою популярністю користуються: RDF*, XML*, JSON*, CSV*, XLS(X), ODS*, YAML*[3]. Датасет може містити у собі від одного до декількох сотень ознак та до декількох мільйонів об'єктів. Правильно підібраний датасет – це запорука успішного результату. Можна було б збирати дані самостійно, але це займає багато часу і у поєднанні з великим обсягом роботи, втомлює. Тут нам на допомогу приходять готові датасети із найрізноманітніших категорій. Із найвідоміших ресурсів можна виділити [4]:

- Google Dataset Search. Dataset Search дозволяє за ключовим словом шукати датасети по всій Мережі.
- Kaggle. Майданчик для змагань по машинному навчанню з безліччю цікавих датасетів. У списку датасета можна знайти різні нішеві екземпляри – від оцінок рамена до баскетбольних даних NCAA і бази ліцензій на домашніх тварин в Сіетлі.
- UCI Machine Learning Repository. Один з найстаріших джерел датасета в Мережі і перше місце, куди варто зазирнути в пошуку цікавих датасетів. Хоча вони додаються користувачами і тому мають різну ступінь «чистоти», більшість з них очищені. Дані можна завантажувати відразу без реєстрації.
- VisualData. Датасети для комп'ютерного зору, розбиті за категоріями. Доступний пошук.
- Find Datasets | CMU Libraries. Колекція датасетів, надана університетом Карнегі Меллон.
- MIMIC-III. Датасети з безособовими даними про стан здоров'я ~ 40 000 пацієнтів, що знаходяться на інтенсивній терапії. Він включає демографічні дані, показники життєдіяльності, лабораторні аналізи, ліки та багато іншого.
- Quandl. Платформа з фінансовими і економічними даними, що допомагають в розробці data-driven стратегій.
- Academic Torrents. Розподілена система, яка містить понад 45 ТБ даних для досліджень. Потрібно звернути увагу на умови ліцензійної угоди – більшість наборів даних дозволено використовувати в некомерційних та освітніх цілях.

- Data.gov. Репозиторій відкритих наборів даних від урядових установ США з таких тем, як клімат, споживачі, освіта, екосистеми, енергетика, фінанси, виробництво, наука.

Серед великого різноманіття джерел датасетів також можна виділити українські, хоча їх невелика кількість, але вони користуються популярністю. Одним із найвідоміших українських джерел наборів даних є Єдиний державний веб-портал відкритих даних [5]. Він містить 35 тисяч наборів даних, що розбиті на групи. Будь-яка особа може вільно копіювати, публікувати, поширювати, використовувати їх, у тому числі в комерційних цілях, у поєднанні з іншою інформацією або шляхом включення до складу власного продукту, публічну інформацію у формі відкритих даних з обов'язковим посиланням на джерело отримання такої інформації. Структура набору відкритих даних оприлюднюється у форматах XSD, JSON, CSV або інших аналогічних форматах.

Підводячи підсумок, можна сказати, що завдяки стрімкому розвитку інформаційних технологій створюється все більша кількість наборів даних, що провокує збільшення джерел для їх зберігання. Набори даних все частіше використовують для проведення data-driven досліджень, задоволення інтересу до тенденцій в певних сферах нашого життя та у вирішенні низки проблем у різних областях.

Список літературних джерел

1. Neskorodieva T., Fedorov E. Method of spectral clustering of payments and raw materials supply for the compliance audit planning. *Радіоелектроніка, інформатика, управління*. 2021. №1. pp.127-135. doi: 10/15588/1607-3274-2021-1-13
2. Vunivere.ru [Електронний ресурс]. Режим доступу – <https://vunivere.ru/work35003>
3. Loc.gov[Електронний ресурс]. Режим доступу – https://www.loc.gov/preservation/digital/formats/fdd/dataset_fdd.shtml
4. Nuancesprog.ru [Електронний ресурс]. Режим доступу – <https://nuancesprog.ru/p/4258/>
5. Gov.ua [Електронний ресурс]. Режим доступу – <https://data.gov.ua/en/pages/about>

УДК 004.82:004:85

*Грущенко В. Ю., студент 1 курсу
СО Магістр*

*Нескородєва Т. В., канд. техн. наук, доцент,
завідувач кафедри комп'ютерних наук та
інформаційних технологій*

АНАЛІЗ ДАНИХ РИНКУ ПЕРСОНАЛЬНИХ КОМП'ЮТЕРІВ

Донецький національний університет імені Василя Стуса, м. Вінниця

Аналіз даних про ринок персональних комп'ютерів, дозволяє ознайомитись з тим, які пропозиції на ньому переважають, та зробити висновок