

СЕКЦІЯ XIV. КОМП'ЮТЕРНА ТА ПРОГРАМНА ІНЖЕНЕРІЯ

ДОСЛІДЖЕННЯ ЗВ'ЯЗКІВ МІЖ СВІТОВИМИ ТРЕНДАМИ ТА ТЕМАТИКАМИ ВІДЕО-КОНТЕНТА

Яковлєв Денис Сергійович

здобувач вищої освіти кафедри АІС

Кременчуцький національний університет ім. Михайла Остроградського, Україна

Науковий керівник: Рилова Наталя Вікторівна

канд. техн. наук, старший викладач

Кременчуцький національний університет ім. Михайла Остроградського, Україна

Тема дослідження зв'язків між світовими трендами та тематиками відео-контенту є надзвичайно актуальною в сучасному цифровому світі, де відео-контент стає важливим і впливовим способом передачі інформації. Аналіз цих зв'язків дозволяє глибше зрозуміти, як суспільні інтереси та культурні тенденції формують контент, який споживає аудиторія, та навпаки, як відео-контент може впливати на формування суспільних настроїв і трендів. Таке дослідження також має практичне значення для маркетологів, розробників контенту та аналітиків медіа, оскільки дає можливість краще розуміти та передбачати інтереси аудиторії.

Основною метою є аналіз зв'язків між світовими трендами, що відображаються в пошукових системах, та тематиками відео-контенту. Під час дослідження потрібно виявити, як актуальні запити користувачів інтернету впливають на зміст відеоматеріалів, які стають популярними в мережі. Для досягнення цієї мети використовується метод класифікації тексту назв відео, що дозволяє точно визначити основні теми та категорії цих відеоматеріалів. На основі цих даних, можна знайти зв'язок між кількістю переглядів відео-контенту та зацікавленістю аудиторії темами, розкритими у матеріалі.

Для отримання наборів даних, необхідних для цього дослідження, було використано Youtube Data API [1], а також відео-контент популярного ютуб-каналу WatchMojo. В ході дослідження було отримано два датасети: перший містить дані про 970 найпопулярніших відео каналу, а другий датасет складається з 390 останніх відео-публікацій за останні 2 місяці. Обидва датасета досліджувались окремо: такий підхід сприяє глибшому розумінню того, як різні види відеоконтенту відображають сучасні тренди та інтереси аудиторії.

Для визначення тематик відео у цьому дослідженні була використана класифікація тегів відео, отриманих через Youtube Data API [1]. Використання методу класифікації тегів дозволило отримати більше інформації про зміст відео без необхідності аналізувати текст відеоряду. Цей підхід ефективно спрощує процес аналізу, оскільки теги, що присвоєні відеоматеріалам, надають ключову інформацію про тематику відео, їх зміст та основні ідеї.

Для визначення світових трендів був використаний ресурс Google Trends [2]. Аналіз зосереджувався на п'яти основних тематиках кожного відео із датасетів. Значущість тематики визначалася коефіцієнтом частки кількості запитів до

пошукової системи Google на цю тему у день коли було опубліковане дане відео. Коефіцієнт кожної теми визначається від 0 до 100 відповідно до її актуальності.

Після отримання всіх необхідних даних, був розрахований коефіцієнт кореляції Пірсона [3] для кожного датасету. Коефіцієнт розраховується за відповідною формулою (1).

$$r_p = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}} \quad (1)$$

де:

r - коефіцієнт Пірсона;

n - розмір датасету;

x, y - відповідні значення наборів даних.

Після отримання результатів аналізу першого датасету (рис. 1) та другого (рис. 2) можна зробити певні висновки.

Коефіцієнт кореляції Пірсона першого датасету дорівнює -0.03, що вказує на відсутність зв'язку трендів з кількістю переглядів відео. Також можна помітити, що велика кількість записів має коефіцієнт трендовості тематик більше за 100, а кількість переглядів від 1 до 35 мільйонів що вказує на те, що майже всі відеоматеріали були зроблені на популярні теми.

Коефіцієнт кореляції Пірсона другого датасету дорівнює 0.20, що вказує на слабкий зв'язок трендів з кількістю переглядів відео. Датасет включає відео з кількістю переглядів в межах від 10 до 700 тисяч і має більший розподіл значень коефіцієнта трендовості тематик. Оскільки результат другого датасету є більш значущим, то доречно надати розподіл найбільш поширених тем (рис. 3).

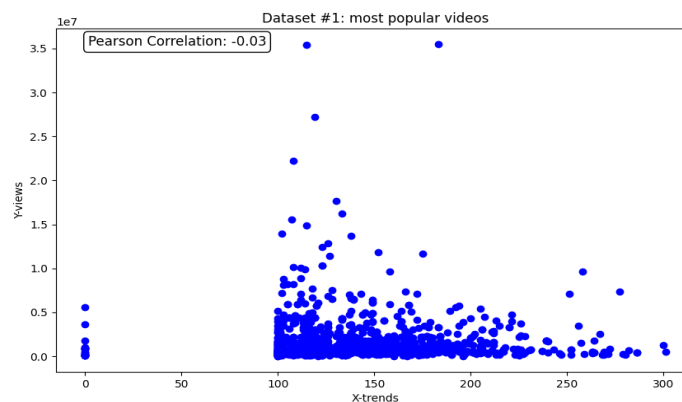


Рис. 1. Аналіз першого датасету

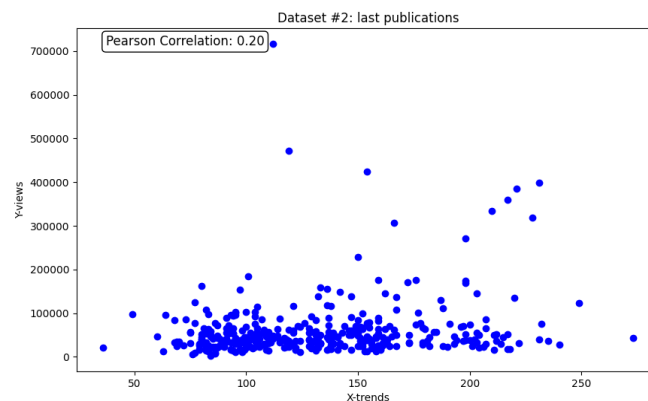


Рис. 2. Аналіз другого датасету

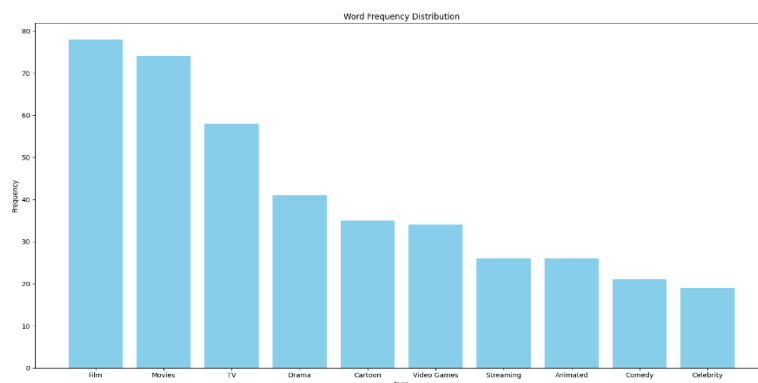


Рис. 3. Розподіл найбільш поширених тем другого датасету

Висновки: Проаналізувавши перший датасет можна відмітити, що використання датасету із популярних відео у аналізі трендів пошукових систем не є ефективним при створенні моделей для передбачення кількості переглядів або іншого аналізу відео-контента через відсутність будь-якого зв'язку. Проаналізувавши другий датасет можна відмітити, що навіть у випадку низької кореляції, отриманий датасет може бути корисним для побудови нейронних мереж. Це можна використовувати для тренування моделей машинного навчання, які можуть виявляти складніші зв'язки та закономірності, які не завжди відображаються в простих кореляційних залежностях.

Список використаних джерел:

1. YouTube Data API. Офіційна документація. [Електронний ресурс] – Режим доступу: <https://developers.google.com/youtube/v3> (дата звернення 25.01.2024).
2. Google Trends. [Електронний ресурс] – Режим доступу: <https://trends.google.com.ua/trends/?geo=UA> (дата звернення 26.01.2024).
3. Коефіцієнт кореляції Пірсона. Вікіпедія. [Електронний ресурс] – Режим доступу: https://uk.wikipedia.org/wiki/Коефіцієнт_кореляції_Пірсона (дата звернення 27.01.2024).