

Лазарєв І. І., студент гр. 126м-21з-1

Науковий керівник: В.Ю. Каштан, к.т.н., доцент кафедри інформаційних технологій та комп'ютерної інженерії

(Національний технічний університет "Дніпровська політехніка", м. Дніпро, Україна)

## АВТОМАТИЗОВАНА МЕТОДИКА АНАЛІЗУ ПОСТІВ В СОЦІАЛЬНІЙ МЕРЕЖІ INSTAGRAM

Соціальні медіа, як платформи для обміну контентом, поширення інформації та публікації думок є предметом широкого спектру досліджень, які вивчали та вивчають онлайн-активність у сферах від політики та медіа до охорони здоров'я та комунікації у кризових ситуаціях [1]. У цій статті ми обговоримо наші методи та результати на прикладі публікацій в Instagram. Сьогодні в Instagram понад 1 мільярд користувачів. Це масштабний ринок, який можна оптимізувати, щоб збільшити популярність і залученість публікацій в Instagram. Тому, метою даної роботи є розробка методу, який автоматично аналізує публікації на основі нейронної мережі. Це дозволить визначити, які зміни допомагають збільшити охоплення контенту через лайки публікацій. Використовуючи цей метод, можна згенерувати оцінку для отримання найбільшої видимості та залучення Instagram підписників.

Щоб мати можливість для аналізу або навіть передбачити популярність публікації в Instagram, перше, що нам потрібно, це багато даних, на яких можна навчати наші моделі. Тому, перший крок роботи полягав у створенні набору даних, а саме збору списку впливових людей (блогерів) Instagram, тобто користувачів, які за своєю професією публікують рекламні публікації в Instagram. Для цього ми скористались даними з Inconosquare Index Influencers [2], який містить список із 2000+ впливових осіб Instagram. Наступним кроком буде аналіз профілів Instagram впливових осіб, отриманих на попередньому етапі. Для цього використаємо функцію posts для читання вибраних змінних із кожного профілю через метадані JSON та сканування останніх 12 публікацій (рис.1.).

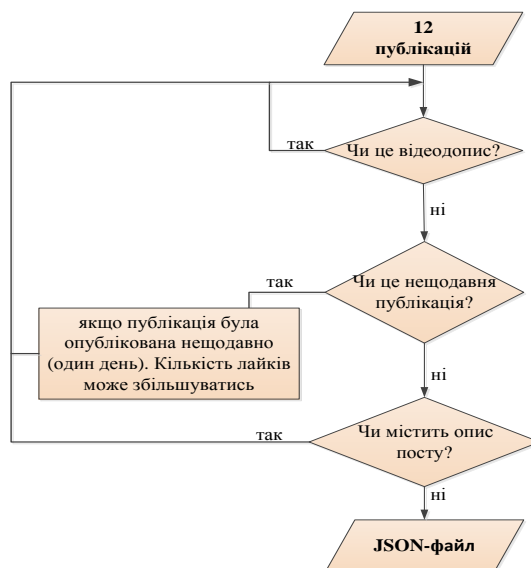


Рисунок 1 – Алгоритм роботи функції posts

Функція не скануватиме публікацію, якщо не виконується будь-яка з цих умов:

- чи це відеодопис: якщо допис є відео, функція не зацікавлена в його скануванні.

Ми шукаємо дописи, які містять зображення, оскільки згодом згортова нейронна мережа (CNN) буде використана для вилучення деяких характеристик зображення;

- чи це нещодавня публікація: якщо публікація була опублікована нещодавно (один день), ми не зацікавлені в цьому, оскільки кількість лайків, наша цільова змінна, може бути не оптимальним, він продовжуватиме зростати;

- чи містить підпису: якщо публікація не має підпису, функція не скануватиме її. Ми витягнемо текстові елементи для кожного підпису публікації за допомогою методів обробки природної мови.

Далі ми пропонуємо використати нейронну мережу для автоматичного аналізу публікацій в Instagram. Так як, досліджувана задача містить змішані дані (показники, текст і зображення), то запропонована модель повинна мати можливість використовувати всі вищезазначені дані. Тому, пропонуємо використати нейронну мережу з трьома гілками. Крім того, пропонуємо використати коефіцієнт залучення. Це кількість загальних залучень, отриманих обліковим записом Instagram за певною публікацією, виражена у відсотках підписників. Залучення визначається як:

$$K_z = 100 \cdot \frac{\text{кількість\_лайків}}{\text{підписники}_i} \quad (1)$$

Запропонована методика була протестована на публікаціях блогерів та проведено аналіз даних залежно від дня тижня (рис.2). Так, день, коли завантажується найбільша кількість постів, – неділя (18%), потім субота та четвер (15%), п’ятниця та середа (14%) і, нарешті, понеділок (11%); найбільш активними днями проставляння лайків є неділя та понеділок. Це впливає на активність аудиторії, просування посту та потрапляння його до топ перегляду.

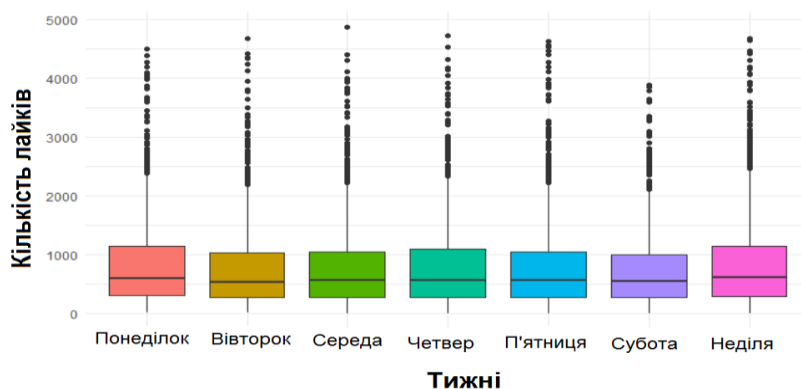


Рисунок 2 – Графік залежності кількості лайків від дня тижня

Отже, в даній роботі запропоновано методику на основі нейронних мереж. Це дозволило автоматично провести аналіз даних та зробити висновок, що на середню кількість лайків не завжди впливає кількість підписників у користувача. Неактивні та «фальшиві» (боти) підписники впливають на результати просування постів, кращим показником виявилось середнє число лайків кожного користувача.

#### ПЕРЕЛІК ПОСИЛАНЬ

4. Most used social media platforms. Statista. [Електронний ресурс] - Режим доступу. - URL:<https://www.statista.com/statistics/272014/global-social-networksranked-by-number-of-users/>
5. Instagram Business: Marketing on Instagram. [Електронний ресурс] - Режим доступу. - URL:[https://business.instagram.com/?locale=en\\_EN](https://business.instagram.com/?locale=en_EN).