

Завдання для категорії Data Science

Task 1 (ML). У цьому завданні необхідно побудувати модель класифікації текстових коментарів.

Це задача бінарної класифікації:

«0» – користувач написав негативний коментар,

«1» – користувач написав позитивний коментар.

Файл train.csv – тренувальний набір коментарів.

Файл test.csv – коментарі, для яких необхідно зробити класифікацію, за якими й буде оцінюватись якість ваших моделей.

Таким чином, потрібно побудувати модель по даним, які містяться у файлі train.csv. Після цього, використовуючи вашу модель, потрібно для коментарів з файлу test.csv обрахувати колонку sentiment.

Для виконання тестового завдання ви можете використовувати будь-які інструменти та алгоритми.

Першу частину завдання знаходьте за [посиланням](#).

Task 2 (Analytical). Ви - аналітик в стартапі, вам потрібно допомогти маркетинговій команді розрахувати ключові показники. Ваш стартап - додаток у якого модель монетизації є підписка. Спочатку юзер бере тріал (він триває 7 днів), після кінця тріалу він може відписатись, або ж продовжити платити за тижневу підписку. Підписка коштує 9.99 за тиждень. Юзер може відписатись у будь-який момент, після цього він не буде більше платити.

Ваше завдання - розрахувати LTV за допомогою Python. Будемо оцінювати ваше загальне розуміння завдання та результат. У файлі [data_analytics.csv](#) ви знайдете фрагмент бази даних, що вам потрібна, де:

- Event Date – дата коли відбулась подія (тріал чи продовження підписки).
- Subscription Offer Type – тип підписки (якщо вказано Free Trial - значить це початок тріалу, якщо поле пусте - це підписка).
- LTV – lifetime value. У цьому випадку рахується через конверсії з тріалу в оплату і потім в наступні оплати.
$$LTV = ([1] + [2] + [3] + [4] + [5]), \text{ де } [1] = (\text{dev proceeds} * \text{Conversion to trial}), \text{ а } [2] = [1] * 1\text{st purchase to 2nd}, [3] = [2] * 2\text{nd to 3rd і тд.}$$

Важливо! Apple Store забирає собі 30% від вартості підписки. LTV в цьому випадку ми будемо рахувати на 6 тижнів (1 тиждень тріалу та 5 підписки).

Код розмістити у відкритому репозиторії на Github. У файлі Readme додати інструкцію як запустити проект (завдання без опису перевірятись не буде).

Щодо будь-яких питань пиши нам в Telegram: @int20h2021_bot

Розв'язок першого завдання потрібно завантажити за [посиланням](#).

Розв'язок другого завдання потрібно завантажити у [форму](#).

Дедлайн: 26 січня 23:59

