1. На основі будь-якого access.log сформувати датасет, що надав би інформацію про користувачів веб-ресурсу, а потім виконати наступні кроки:
   1. Визначити кількість користувачів за днями
   2. Ранжувати користувачів за User-Agent
   3. Ранжувати користувачів за операційними системами
   4. Ранжувати користувачів за країною запиту
   5. Виокремити пошукових ботів
   6. Детектувати аномалії (якщо такі є)
2. На основі даних з [Google Analytics Demo](https://analytics.google.com/analytics/web/?utm_source=demoaccount&utm_medium=demoaccount&utm_campaign=demoaccount#/p213025502) account вибрати ТРИ різні часові ряди і на їх основі шляхом застосування методів визначення аномалій визначити аномалії в поведінці системи.
3. Використовуючи API Twitter, написати скрипт, що буде автентифікуватися в соціальній мережі та забирати N>1000 останніх записів певного акаунту (акаунт має бути публічним), та створити датасет. Опціонально – сформувати систему сповіщення (пошта, телефон, будь-що) про знаходження певного ключового слова.
4. На основі наданого викладачем датасету (або будь-якого, знайденого самостійно) збудувати систему прогнозування / класифікації (в залежності від датасету та типу даних) із застосуванням методів машинного навчання.

Повідомлення з ТГ:

1. Відновлення повноти даних про відвідуваність веб-ресурсу шляхом аналізу логів доступу (визначення аномалій у логах веб-сайтів, access.log, error.log\*).

2. Визначення аномальної поведінки користувачів веб-ресурсу із використанням Google Analytics (не менше трьох показників, або часових рядів).

3. Збір великих даних із соціальних мереж. Автентифікація користувача. Збір інформації із джерел Deep Web.

4. Збір великих даних із новинних ресурсів. Формування датасетів. Можна один ресурс не менше 1000 записів.