**ДОМАШНЄ ЗАВДАННЯ 2**

**Мета**: навчитися візуалізувати дані, знаходити та оброблювати викиди, обирати найкращу модель та валідувати.

**Кроки виконання завдання:**

1. В завданні можна використовувати або датасет з попереднього завдання, або обрати новий.
2. Використовуючи бібліотеки pandas або numpy, завантажити дані в jupyter notebook.
3. Дослідіть дані:
   1. Візуалізуйте їх за допомогою пакетів matplotlib, seaborn або будь-якого іншого. Побудуйте не менше трьох різних видів графіків для ваших даних.
   2. Окремо до цих графіків побудуйте графік «ящик з вусами» та дослідіть за допомогою нього, чи є в ваших даних викиди.
4. Якщо виявилося, що ваші дані мають викиди, видаліть такі точки. Використовуйте для цього підходи, які було розглянуто на лекції.
5. Оберіть метрику, яка найбільше підходить для вашої задачі та даних. Візьміть до уваги той факт, чи ваші дані збалансовані, чи ні.
6. Якщо дані незбалансовані, використайте або один з підходів до ребалансування даних, або оберіть відповідну для таких випадків метрику.
7. Розділіть ваші дані на дві підвибірки – тренувальну та валідаційну.
8. Оберіть декілька моделей, які найбільше підходять для вирішення вашої задачі.
9. Натренуйте ваш зоопарк моделей за допомогою крос-валідації на тренувальних даних.
10. (**update**) В кінці отримайте результат вашої моделі – значення метрики, яку ви обрали раніше на валідаційній вибірці даних для моделі, що дала найкращий результат при крос-валідації.

Виконане завдання - jupyter notebook та модуль з функціями обробки даних закомітити на ваш github. Відмітитись в наступному файлі – записати ім'я/прізвище та посилання на виконане завдання github: [гугл док](https://docs.google.com/spreadsheets/d/1Uw7prdKOwOd1qMdjHHUM_cWVzwZNjfxzDY-FP4_IDaY/edit#gid=642196238) – сторінка HW\_2. В описі до репозиторію залишити посилання на дані, які ви використали в домашній роботі.

Дедлайн здачі роботи – 31 січня до кінця.

**Зверніть увагу**: робота оцінюється!