Лабораторна робота #5  
Кластеризація

short line

Мета роботи

Ознайомитись з основами машинного навчання та аналізу даних без учителя для розв'язання задачі кластеризації даних, реалізувати метод, що базується на принципі навчання без учителя (К-середніх) за допомогою бібліотеки *scikit-learn*.

# **Рекомендована література**

Про роботу з методом К-середніх в Python:

<https://scikit-learn.org/stable/modules/generated/sklearn.cluster.KMeans.html>

<https://towardsdatascience.com/k-means-clustering-with-scikit-learn-6b47a369a83c>

Лекційні матеріали та російськомовна версія книги **Сабастьяна Рашки “Python и машинное обучение”** у pdf-форматі доступна для завантаження за посиланням:

<https://drive.google.com/file/d/1NOBl5mhAhlTxoxxl8aMWFQMU-MLuZ8ya/view?usp=sharing>

# **Хід роботи**

## **Дані**

В роботі використовується публічний набор даних з <https://www.kaggle.com/datasets> (для тих, хто претендує на максимальний бал).

**Примітка**: на нижчу оцінку можна використовувати типові ідеальні набори даних (типу ірисів Фішера і т.п.)

## **Підготовчий етап**

Провести аналіз вибраного набору даних, візуалізувати дані, спробувати виявити основні залежності, детектувати аномалії, неповні зразки тощо у даних. Провести підготовку даних до подальшого використання.

**Реалізація моделі**

Вивчити засади роботи з методами без учителя та основні методи для розв'язання задачі кластеризації. Написати код для кластеризації для обраного набору даних та підбір оптимальної кількості кластерів.

**Аналіз результатів**

1. Вибір оптимальної кількості кластерів
2. Оцінка результатів роботи кластеризації при різній кількості кластерів
3. Результати оформити протоколом

**Контрольні питання**

1. Як працюють методи машинного навчання без учителя?
2. Критерій оптимальної кількості кластерів?