

## Лабораторна робота №4 Метод $k$ найближчих сусідів

### Підгрупа №2

1. Відкрити та зчитати наданий файл з даними.
2. Визначити та вивести кількість записів.
3. Вивести атрибути набору даних.
4. З'ясувати збалансованість набору даних.
5. Отримати двадцять варіантів перемішування набору даних та розділення його на навчальну (тренувальну) та тестову вибірки, використовуючи функцію *ShuffleSplit*. Сформувати початкову та тестові вибірки на основі обраного користувачем варіанту .
6. Використовуючи функцію *KNeighborsClassifier* бібліотеки *scikit-learn*, збудувати класифікаційну модель на основі методу  $k$  найближчих сусідів (кількість сусідів обрати самостійно, вибір аргументувати) та навчити її на тренувальній вибірці, вважаючи, що цільова характеристика визначається стовпчиком *Class*, а всі інші виступають в ролі вихідних аргументів.
7. Обчислити класифікаційні метрики збудованої моделі для тренувальної та тестової вибірки. Представити результати роботи моделі на тестовій вибірці графічно.
8. Обрати алгоритм *KDTree* та з'ясувати вплив розміру листа (від 20 до 200 з кроком 5) на результати класифікації. Результати представити графічно.