

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. Отримати десять варіантів перемішування набору даних та розділення його на навчальну (тренувальну) та тестову вибірки, використовуючи функцію *ShuffleSplit*. Сформуванати початкову та тестові вибірки на основі восьмого варіанту. З'ясувати збалансованість набору даних.
5. Використовуючи функцію *KNeighborsClassifier* бібліотеки *scikit-learn*, збудувати класифікаційну модель на основі методу k найближчих сусідів (значення всіх параметрів залишити за замовчуванням) та навчити її на тренувальній вибірці, вважаючи, що цільова характеристика визначається стовпчиком *quality*, а всі інші виступають в ролі вихідних аргументів.
6. Обчислити класифікаційні метрики збудованої моделі для тренувальної та тестової вибірки. Представити результати роботи моделі на тестовій вибірці графічно.
7. З'ясувати вплив кількості сусідів (від 1 до 20) на результати класифікації. Результати представити графічно.