Міністерство освіти і науки України

Західноукраїнський національний університет

Факультет комп’ютерних інформаційних технологій

Кафедра інформаційно-обчислювальних систем і управління

Звіт

Про виконання лабораторної роботи №3

з дисципліни « Методи та системи штучного інтелекту»

Виконав:

Студент групи КНШІ-31

Семенюк Руслан

Тернопіль 2025

**Мета:** набути навики щодо можливостей аналізу даних з використанням класифкатораметодом к-найближчих сусідів (k-nn).

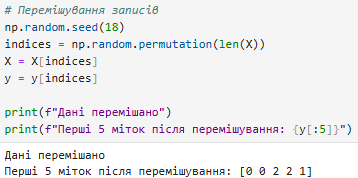
**Хід роботи**

**Крок1**. Завантажити базу параметрів квітів



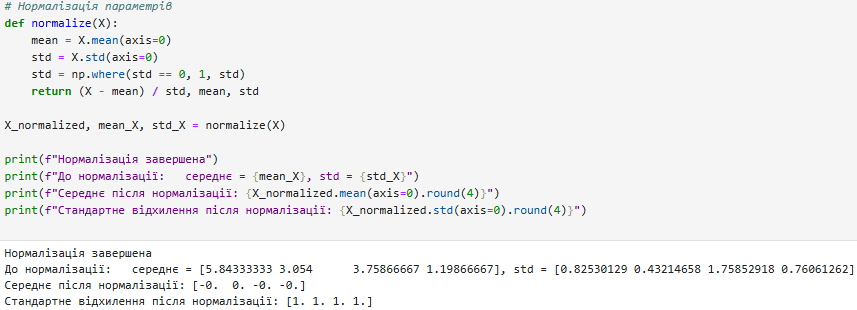
Читаємо CSV-файл з даними про іриси. Перші 4 колонки - це числові параметри квітів, остання колонка - текстова назва виду. Перетворюємо текстові назви видів ("Iris-setosa", "Iris-versicolor", "Iris-virginica") у числові коди (0, 1, 2)

**Крок 2**. Перемішуємо записи



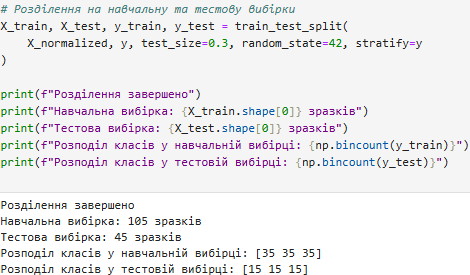
Випадковим чином змішуємо всі записи, щоб у навчальній і тестовій вибірках були різні види ірисів у випадковому порядку.

**Крок 3**: Нормалізуємо параметри квітів



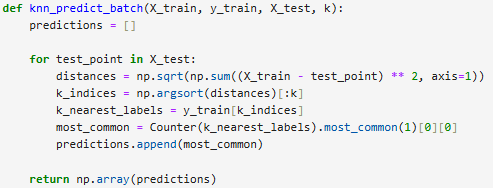
Приводимо всі параметри до одного масштабу. Наприклад, довжина чашолистка може бути 4-8 см, а ширина пелюстки 0.1-2.5 см. Нормалізація робить так, щоб усі параметри мали середнє значення 0 і стандартне відхилення 1.

**Крок 4**: Розділяємо вибірку



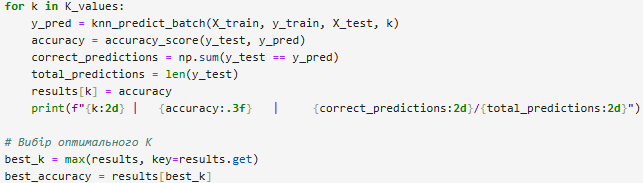
Ділимо дані на дві частини: навчальну(70%) та тестову(30%) вибірки

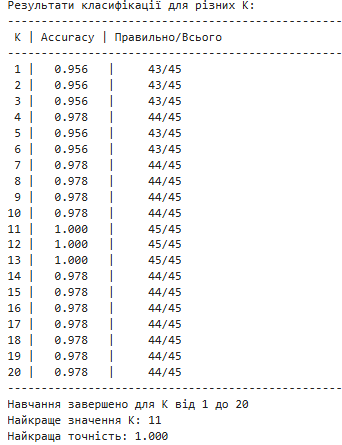
**Крок 5**: Навчаємо KNN-класифікатор із різними значеннями К та Вибираємо найкращий параметр К



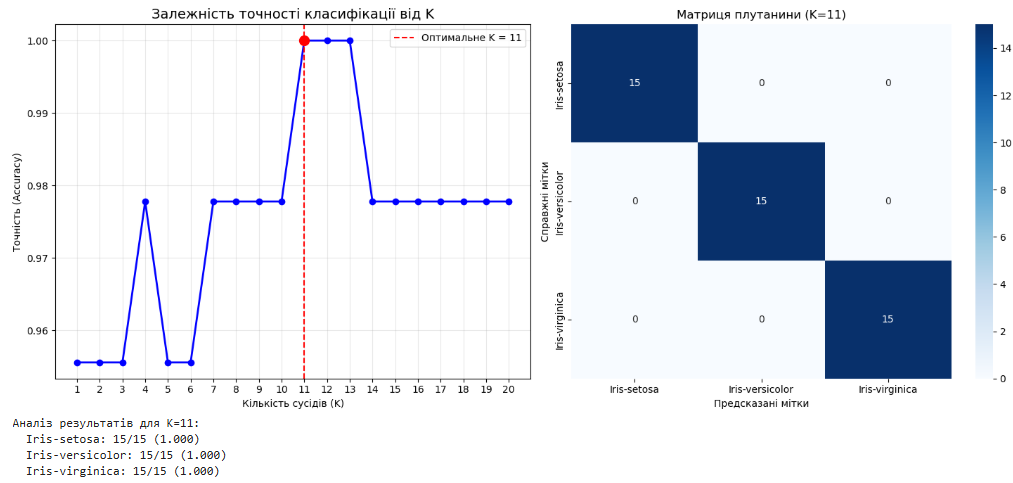
Для кожного тестового квіта:

1. Обчислюємо відстані до всіх навчальних квітів
2. Знаходимо K найближчих сусідів
3. Дивимося, які види переважають серед цих сусідів
4. Призначаємо тестовому квіту найпоширеніший вид





**Крок 6**: Візуалізуємо результати



1. Модель з K=11 досягла повної точності (100%)
2. Всі 45 тестових квітів правильно класифіковані
3. Жодних помилок між видами
4. K=11 є оптимальним параметром для цього датасету

Висновок: В ході виконання роботи було реалізовано алгоритм k-найближчих сусідів (KNN) для класифікації видів квітів ірису. На основі наданого датасету проведено повний цикл обробки даних, що включав завантаження, перемішування, нормалізацію параметрів та розділення на навчальну і тестову вибірки.

Шляхом експериментального дослідження впливу параметра K на якість класифікації було встановлено, що оптимальне значення K=11 забезпечує максимальну точність розпізнавання. Аналіз матриці плутанини підтвердив повну відсутність помилок класифікації - всі 45 тестових зразків були правильно ідентифіковані належними до відповідних видів.

**GitHub**[https://github.com/s1pliy/AI/blob/main/Lab3/Lab3\_MSAI.ipynb](https://github.com/s1pliy/AI/blob/main/Lab3/Lab3_MSAI.ipynb%20)