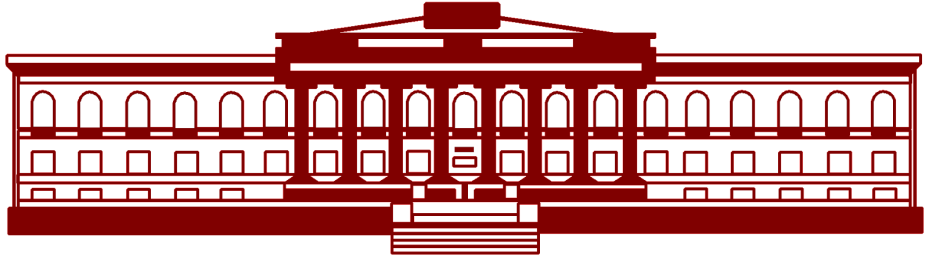
**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

****

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

**Лабораторна робота №2**

**з курсу**

**«Технології обробки даних»**

*Студентки 5 курсу*

*групи ІС-11*

*спеціальності 122 «Комп'ютерні науки»*

*ОП «Інформаційні системи»*

Швидюк Анастасії Юріївни

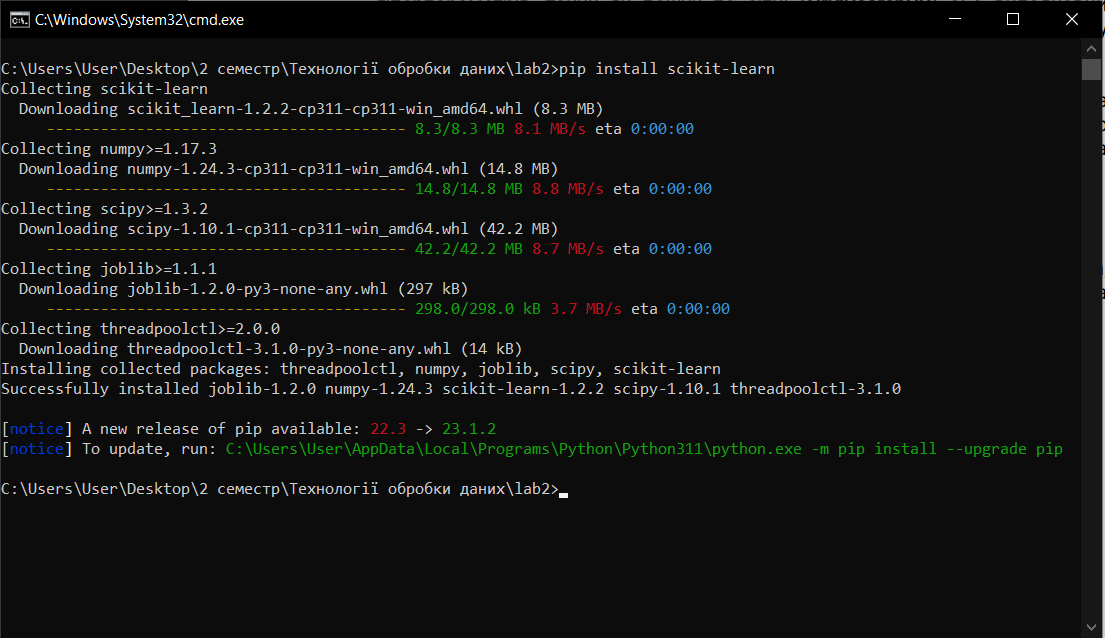
**Київ – 2023**

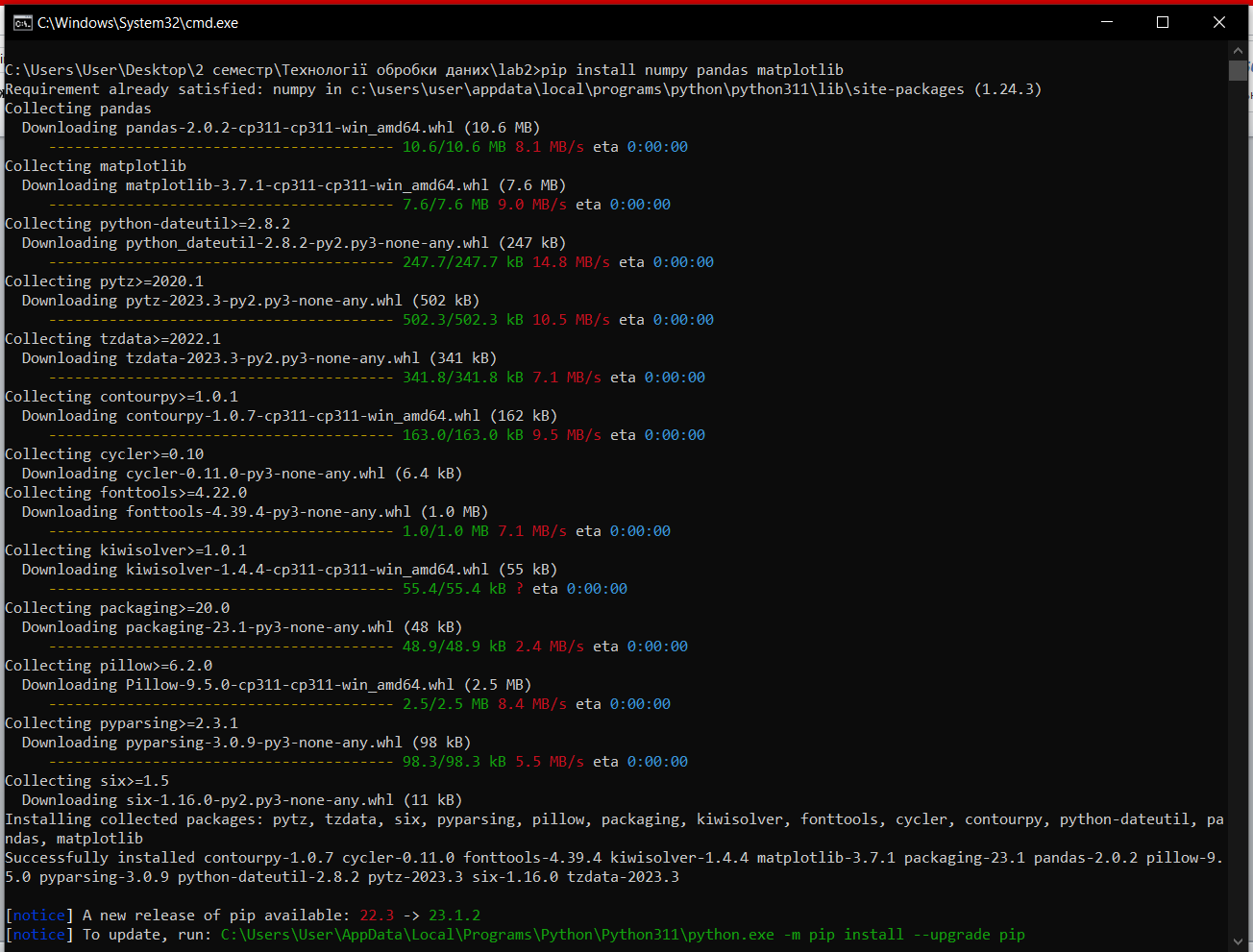
Лабораторна робота № 2

Тема: Пайплайни в scikit-learn як засіб автоматизації дослідження

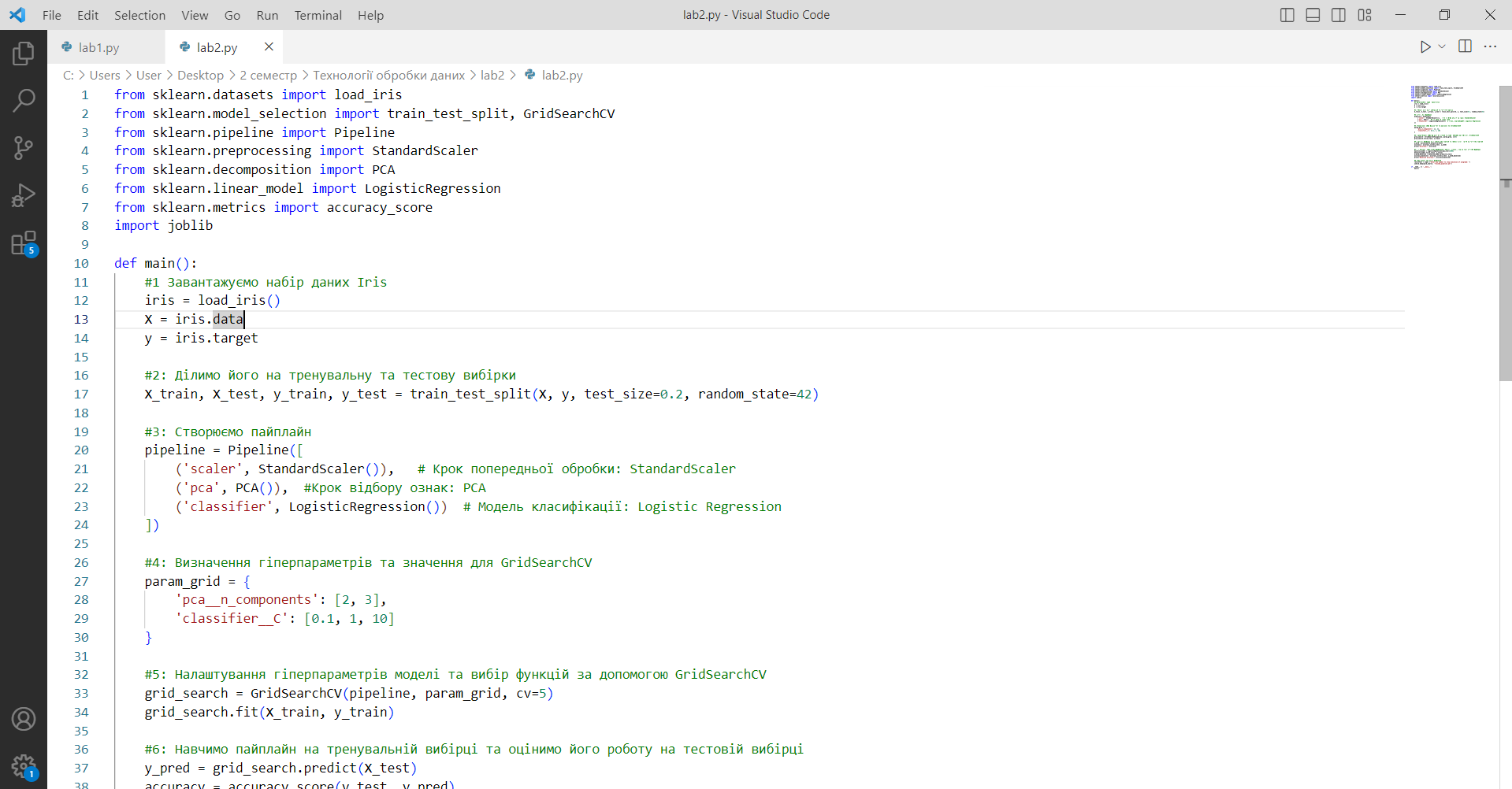
Мета роботи: розуміння та застосування пайплайнів у scikit-learn для автоматизації процесу дослідження в машинному навчанні.

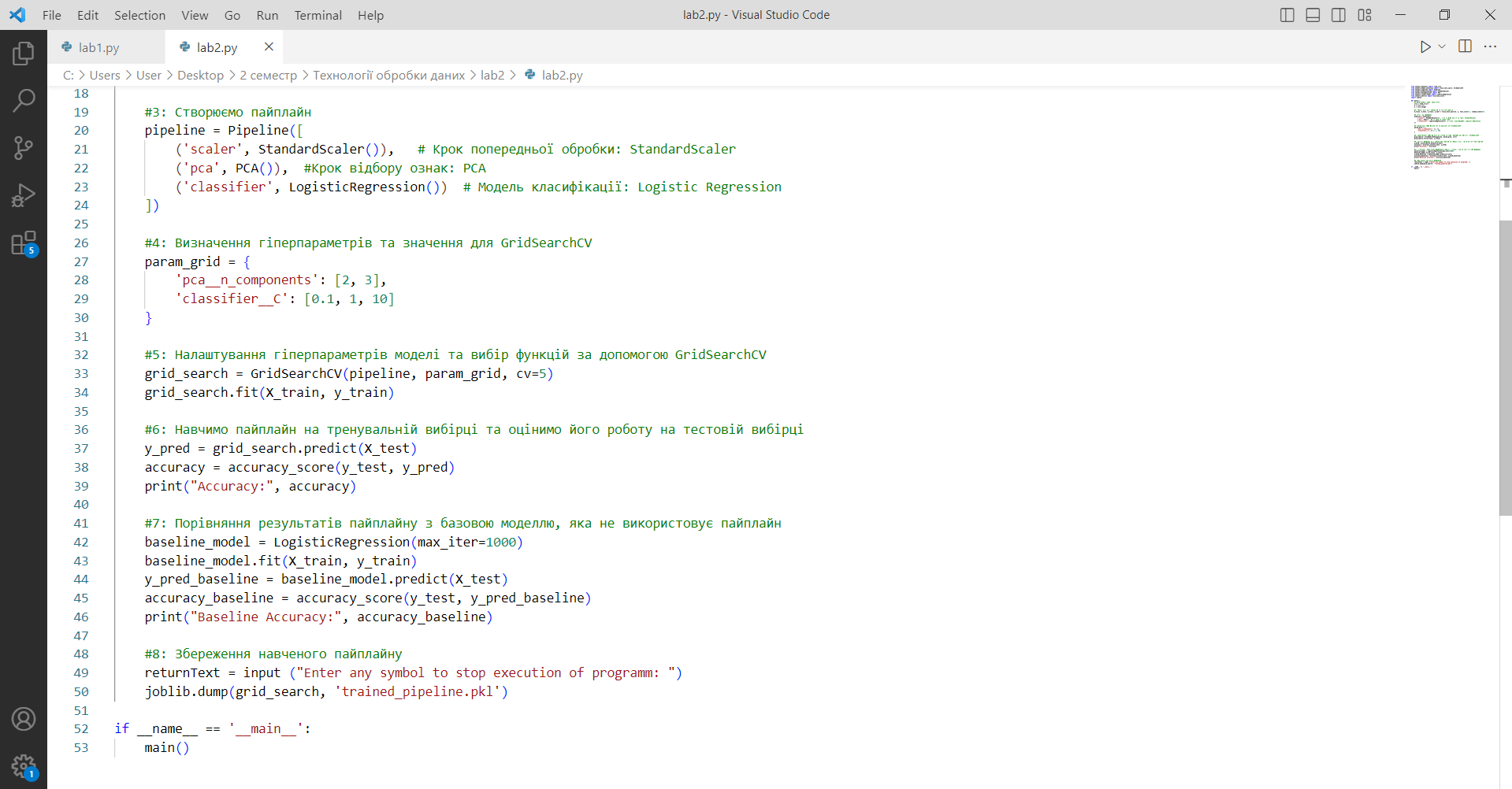
Встановимо усі необхідні пакети для виконання завдань





Код програми





Коментарі до коду програми:  
  
Набір даних Iris завантажується за допомогою функції load\_iris() з файлу sklearn.datasets, а ознаки та мітки зберігаються у змінних X та y відповідно.

Набір даних розбивається на навчальну та тестову вибірки за допомогою train\_test\_split з sklearn.model\_selection.

Пайплайн створюється за допомогою Pipeline з sklearn.pipeline. Він включає StandardScaler для попередньої обробки даних, PCA для вибору ознак і LogisticRegression як модель класифікації.

Гіперпараметри та відповідні значення для PCA та LogisticRegression визначені у словнику param\_grid для GridSearchCV.

GridSearchCV використовується для налаштування гіперпараметрів моделі та вибору ознак шляхом пошуку найкращої комбінації гіперпараметрів на основі наданих значень у param\_grid. Пайплайн підганяється до навчальних даних за допомогою функції fit.

Пайплайн використовується для прогнозування на тестовому наборі, а точність обчислюється за допомогою accuracy\_score. Продуктивність порівнюється з базовою моделлю, яка використовує LogisticRegression без жодних налаштувань гіперпараметрів чи функцій

Результат виконання

