**КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ**

**імені ТАРАСА ШЕВЧЕНКА**

****

**ФАКУЛЬТЕТ ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ**

**Кафедра прикладних інформаційних систем**

**Звіт до лабораторної роботи №2**

**з дисципліни**

**«Технології обробки даних»**

*Студента 1 курсу магістрів*

*групи* ***ІС-11***

*спеціальності 122 «Комп'ютерні науки»*

*ОП «Прикладне програмування»*

**Кучера Романа Володимировича**

**Київ – 2023**

**Тема:** Пайплайни в scikit-learn як засіб автоматизації дослідження.

**Мета:** навчитися розуміти та застосувати пайплайни у scikit-learn для автоматизації процесу дослідження в машинному навчанні.

**Завдання:**

1. Завантажте датасет (наприклад, Iris, Wine або Breast Cancer) і поділіть його на тренувальну та тестову вибірки. Для цього використайте функцію train\_test\_split з бібліотеки scikit-learn.

2. Створіть пайплайн, який включає предобработку даних (наприклад, масштабування за допомогою StandardScaler) та відбір ознак (наприклад, використовуючи метод головних компонент, PCA).

3. Додайте до пайплайну модель класифікації або регресії (залежно від обраного датасету), таку як LogisticRegression, RandomForestClassifier чи LinearRegression.

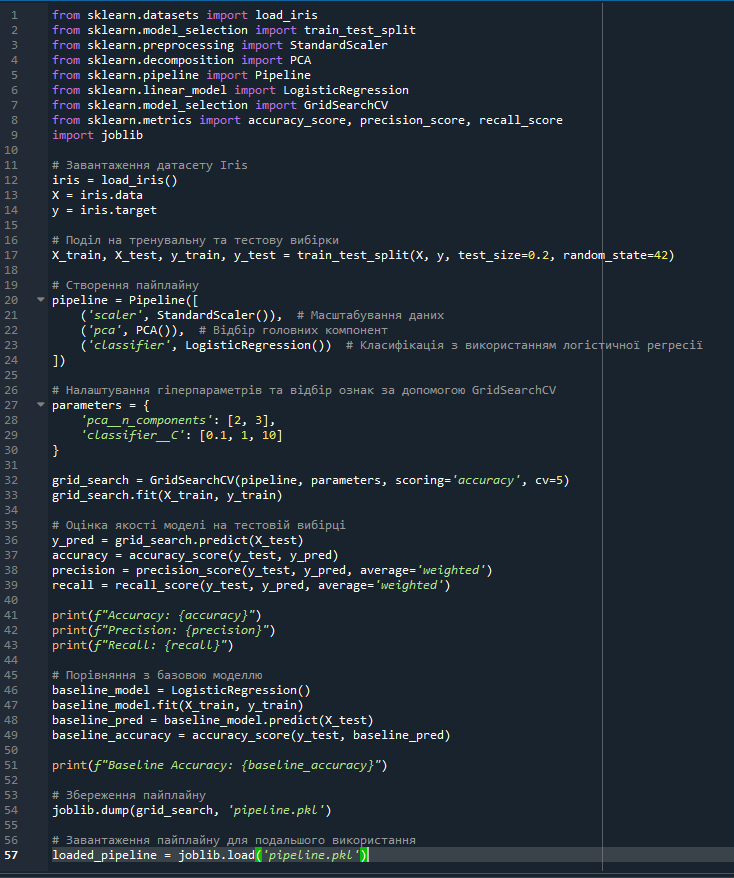
4. Використайте GridSearchCV або RandomizedSearchCV для налаштування гіперпараметрів моделі та відбору ознак. Визначте список гіперпараметрів, які ви хочете налаштувати, та відповідні значення, які ви хочете перевірити.

5. Навчіть пайплайн на тренувальній вибірці та оцініть його якість на тестовій вибірці, використовуючи метрики, такі як точність (accuracy), точність (precision), повнота (recall) або коефіцієнт детермінації (R^2), залежно від типу задачі.

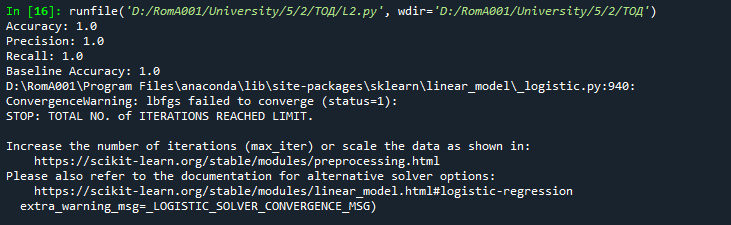
6. Порівняйте результати пайплайну з базовою моделлю, яка не використовує пайплайн. Це допоможе вам оцінити ефективність використання пайплайнів у вашому дослідженні.

7. Збережіть навчений пайплайн за допомогою бібліотеки joblib або pickle та завантажте його для подальшого використання.

**Код програми:**



**Результат виконання коду програми:**



**Висновок:** виконуючи лабораторну роботу, я навчився розуміти та застосувати пайплайни у scikit-learn для автоматизації процесу дослідження в машинному навчанні.