**ЛАБОРАТОРНА РОБОТА № 2**

**ПОРІВНЯННЯ МЕТОДІВ КЛАСИФІКАЦІЇ ДАНИХ**

***Мета:*** використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити різні методи класифікації даних та навчитися їх порівнювати..

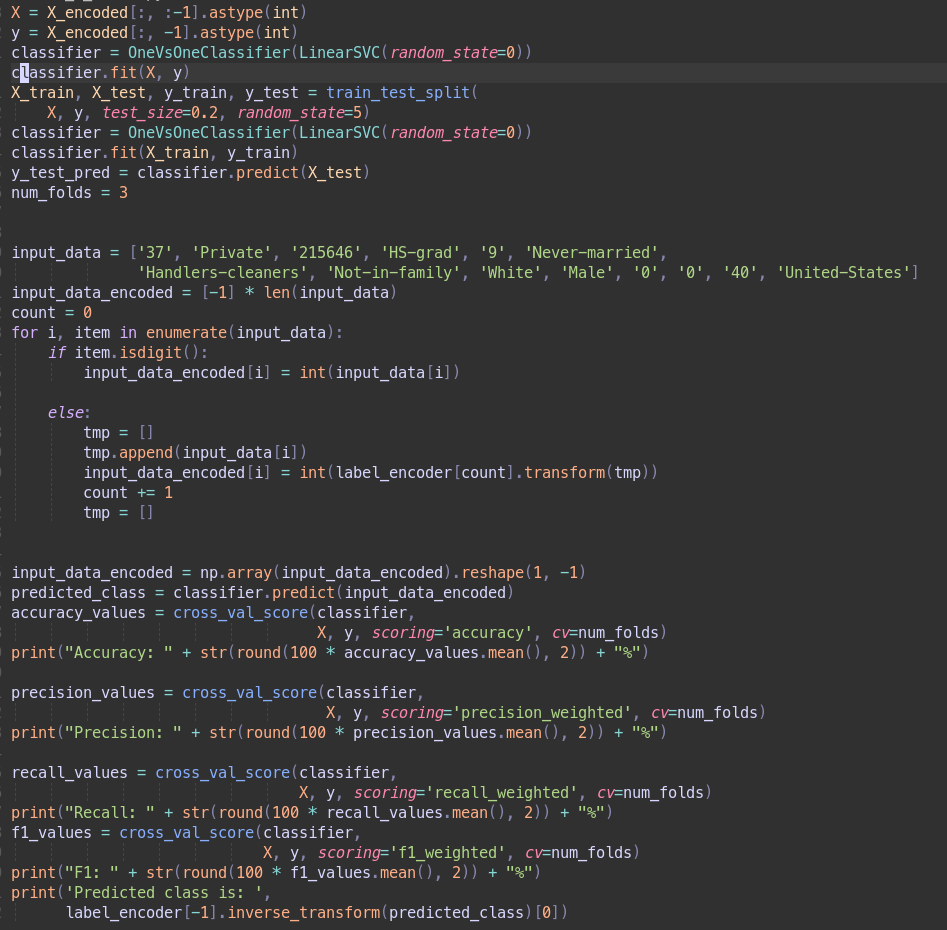
**Хід роботи:**

**Завдання 2.1. Класифікація за допомогою машин опорних векторів (SVM)**

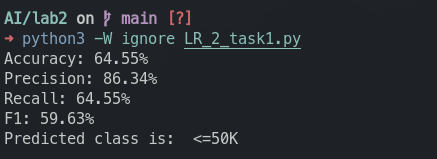
Числові ознаки: age, flnwgt, education-num, capital-gain, capital-loss, hours-per-week; категоріальні: workclass, education, marital-status, occupation, relationship, race, sex, native-country.

Лістинг програми:





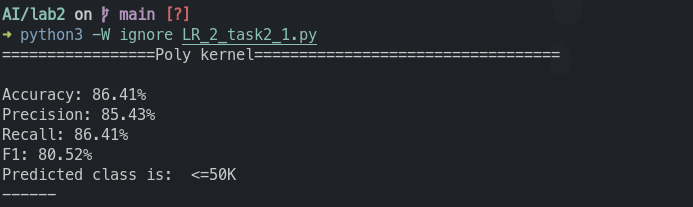
Результат роботи програми:

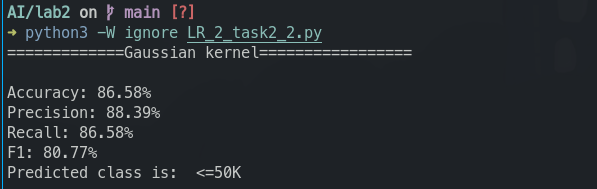


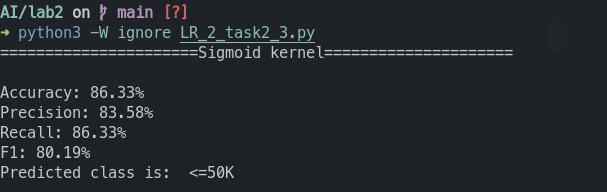
Тестова точка належить до класу ‘<=50К’.

**Завдання 2.2. Порівняння якості класифікаторів SVM з нелінійними ядрами**

Результати роботи класифікаторів:

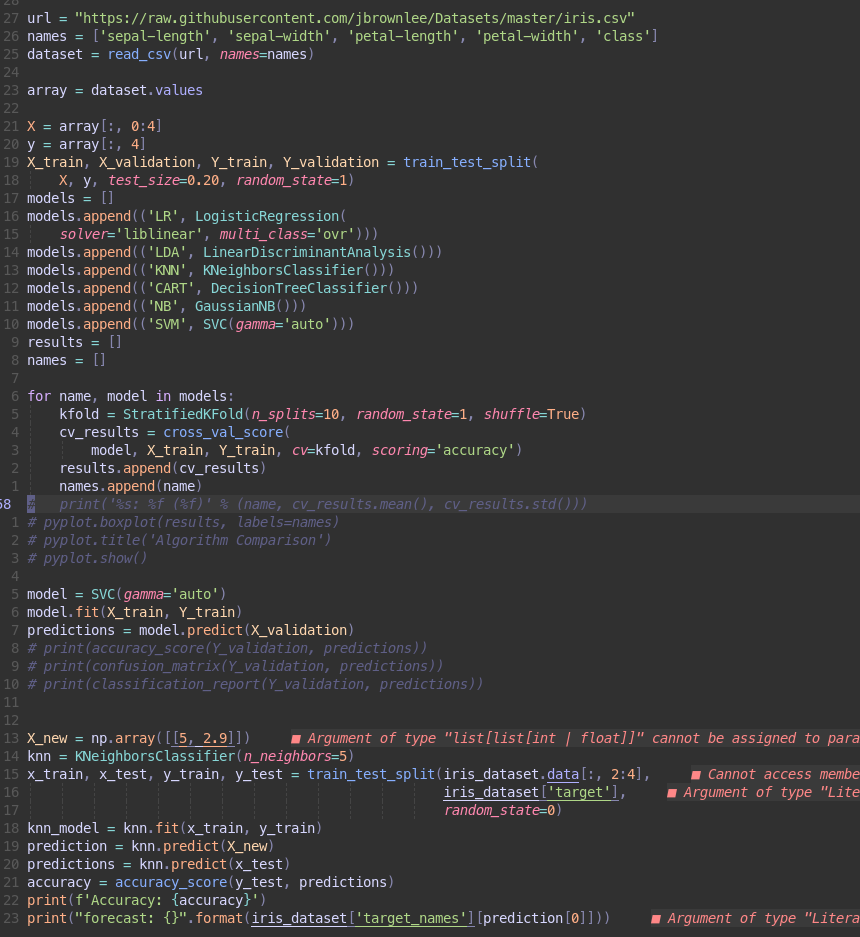




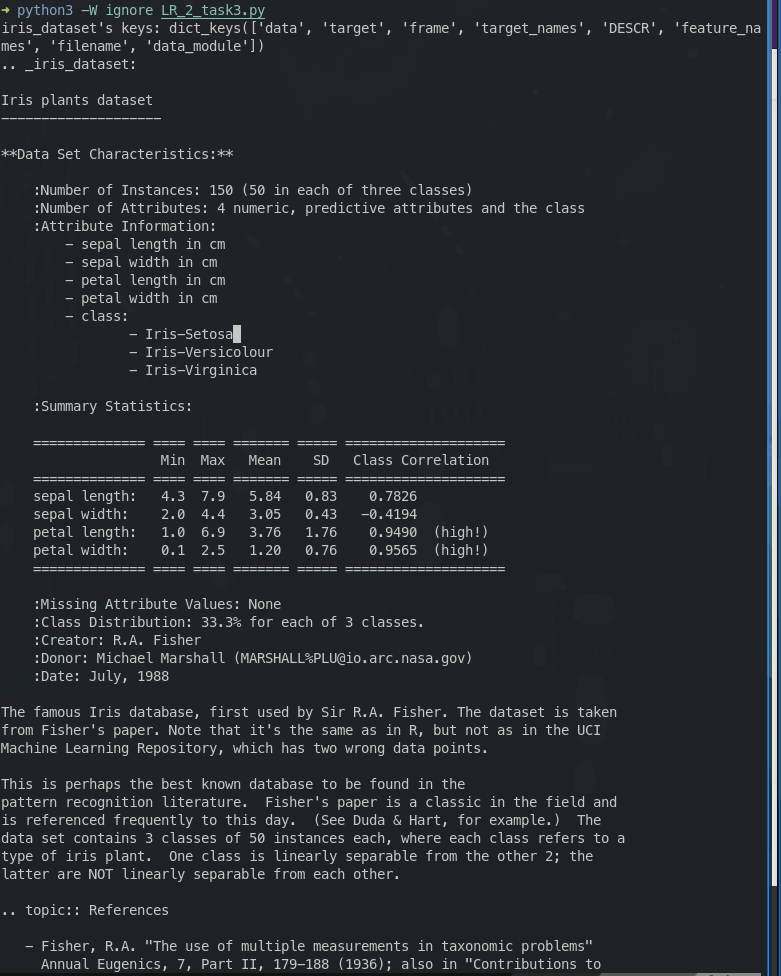


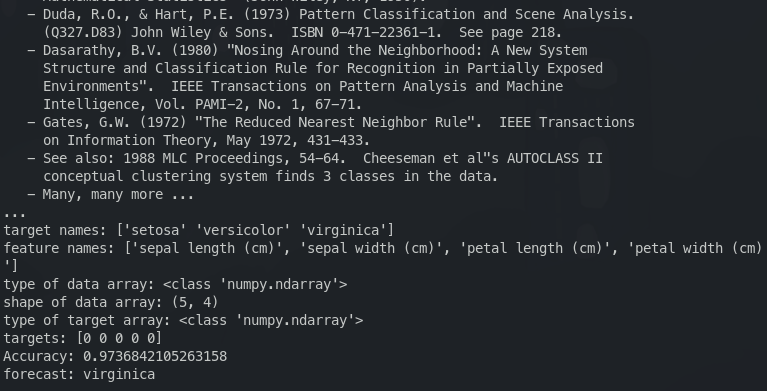
**Завдання 2.3. Порівняння якості класифікаторів на прикладі класифікації сортів ірисів**

Код програми:

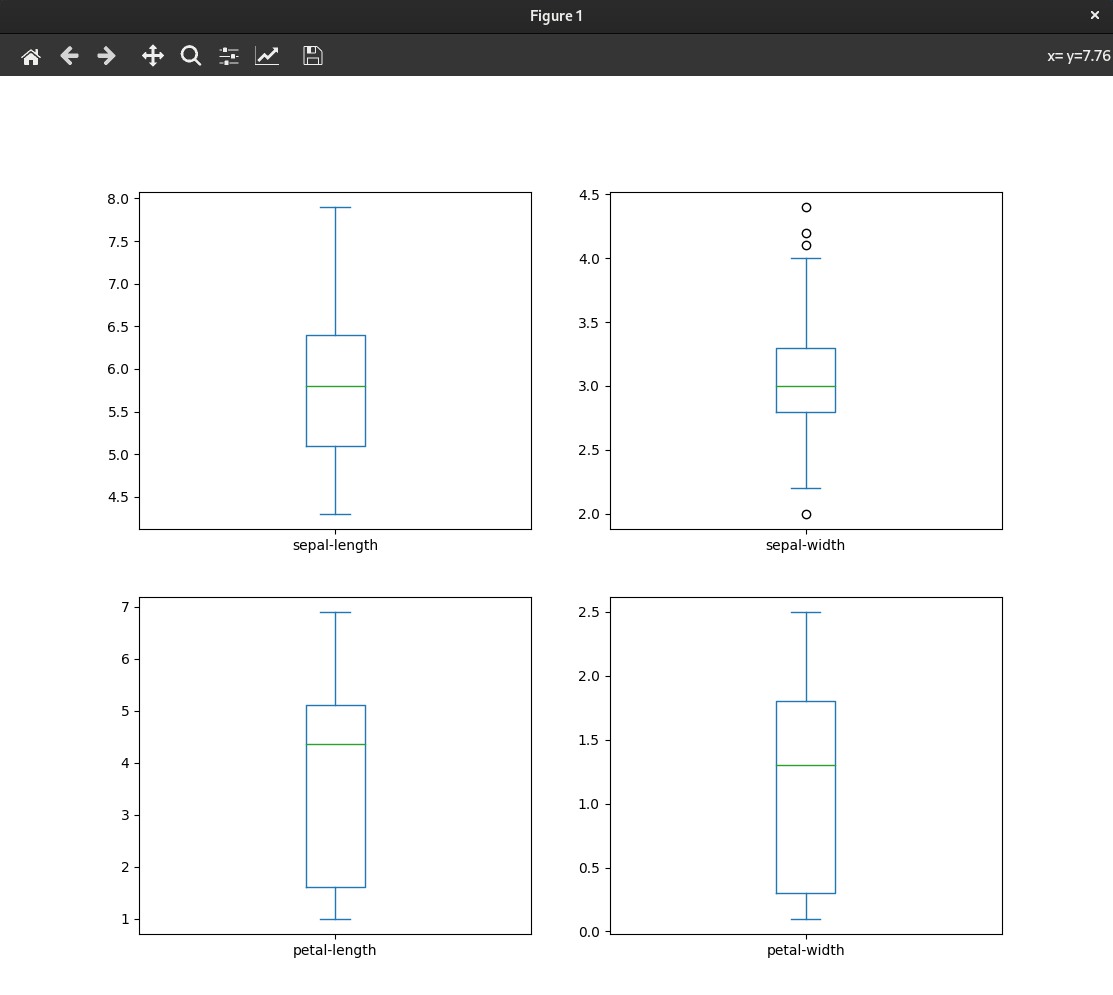


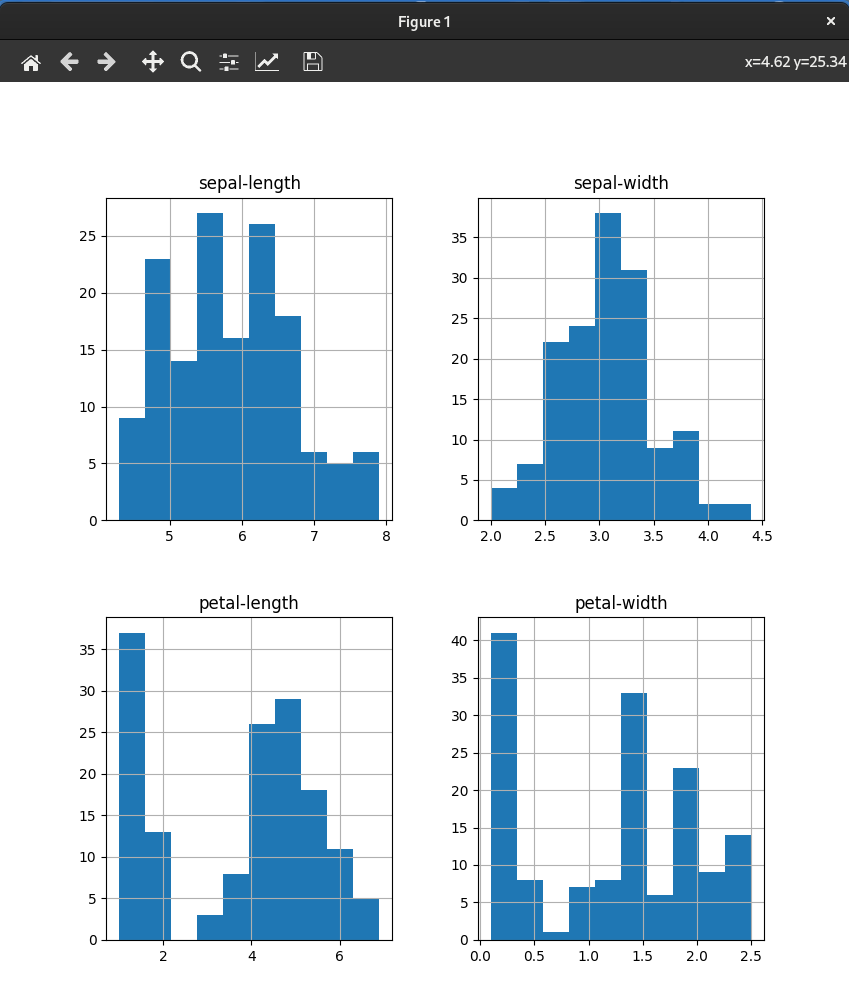
Результат виконання програми:

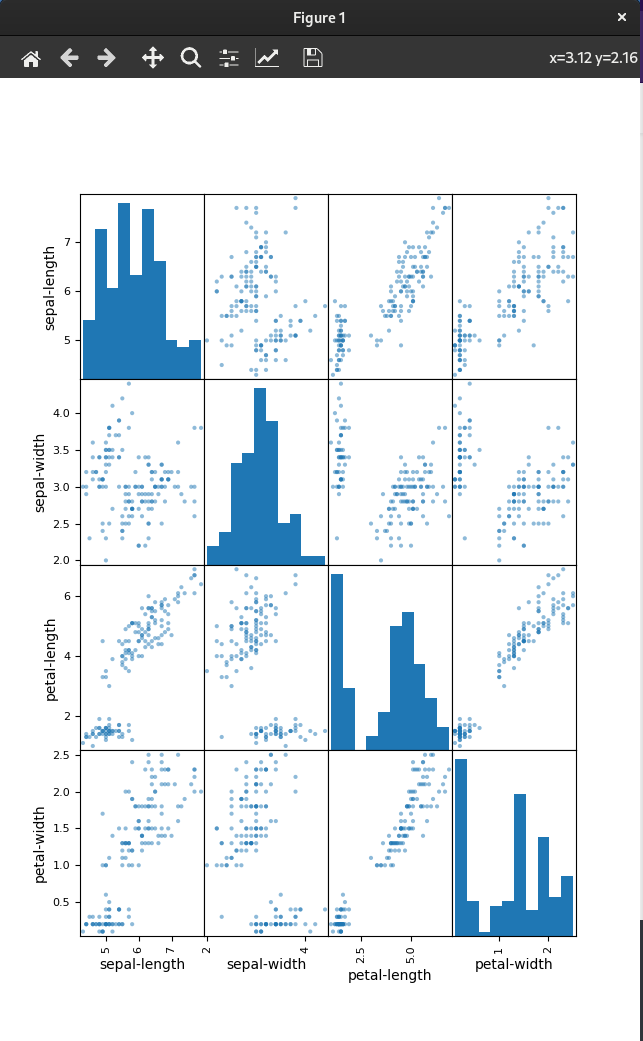


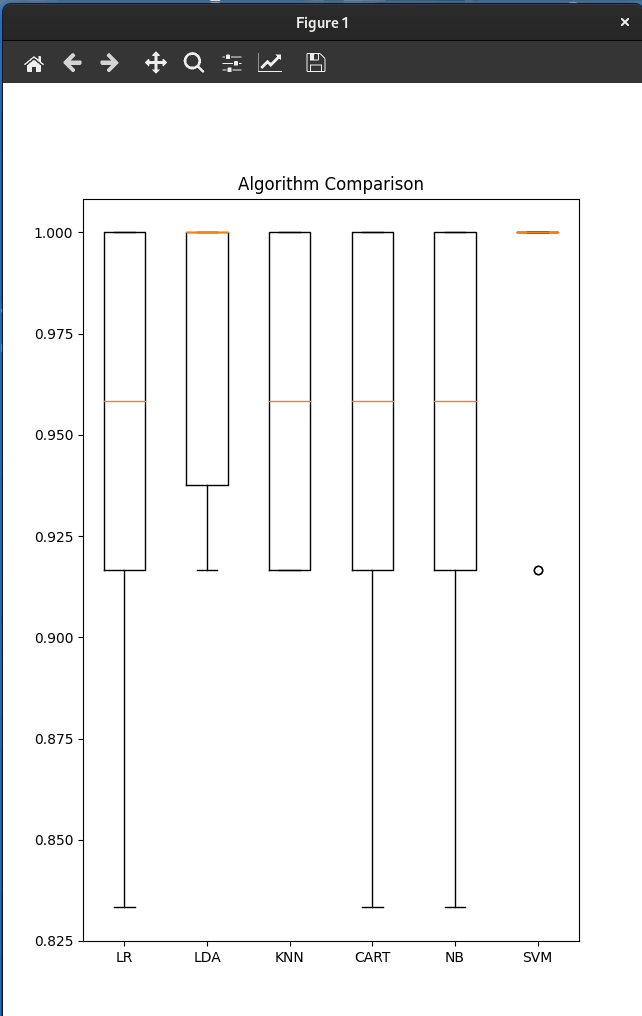


Побудовані графіки:

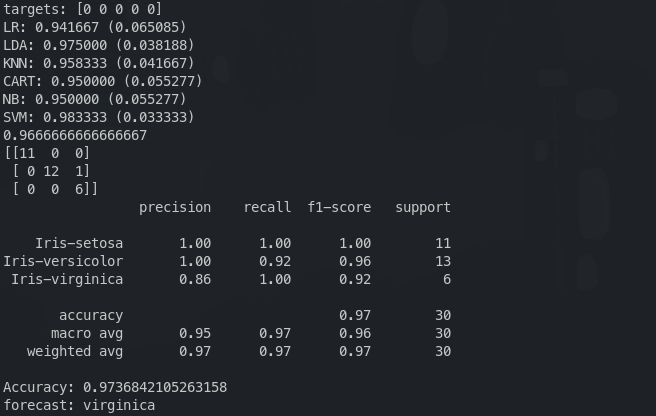








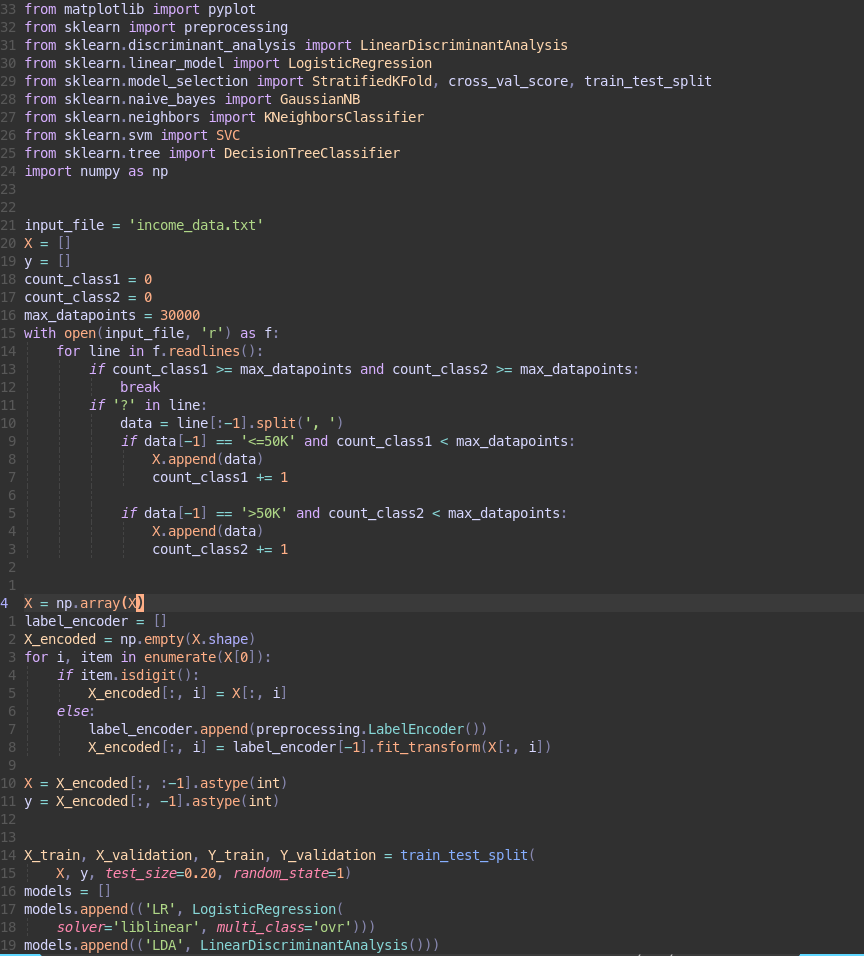
Результат роботи програми:

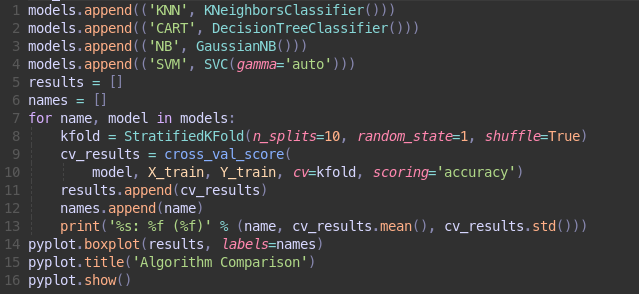


За результатами тренування вдалося досягти точності 0.97 та встановити, що квітка належить до класу virginica.

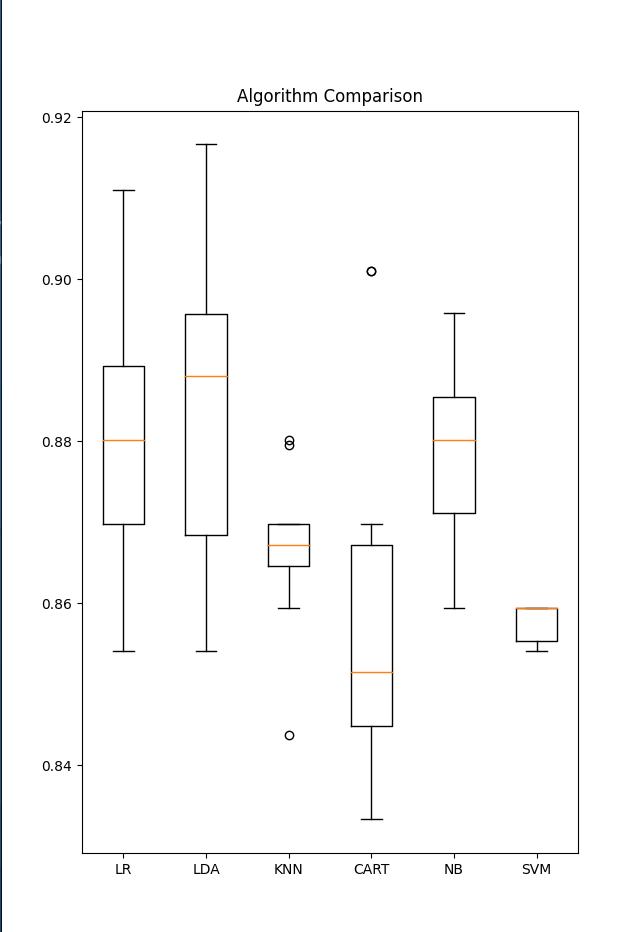
**Завдання 2.4. Порівняння якості класифікаторів для набору даних завдання 2.1**

Код програми:





Результати роботи програми:

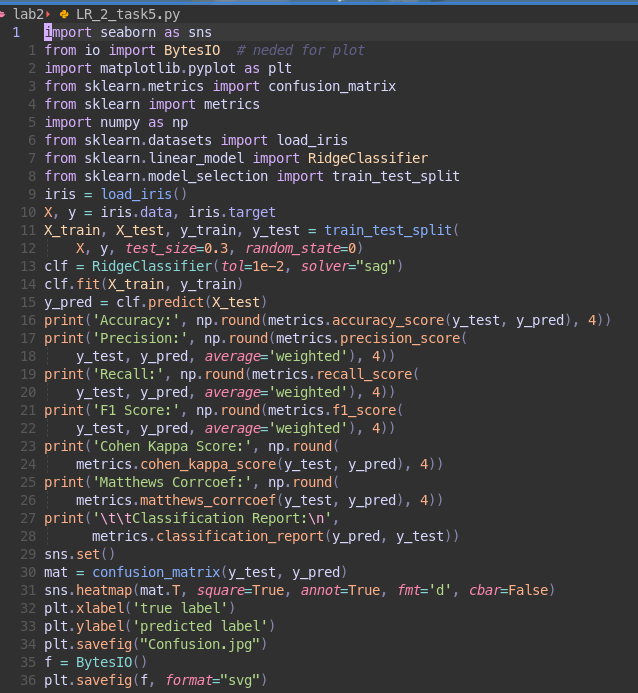




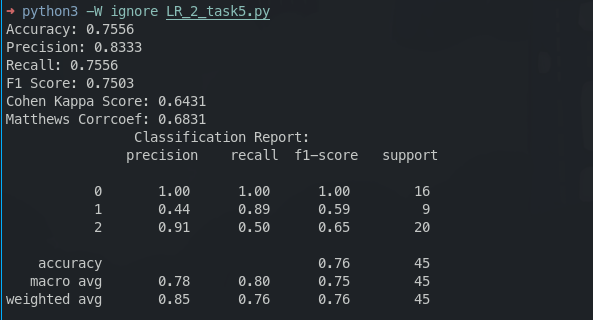
Порівнявши алгоритми за показником точності можна дійти висновку, що найкращим є алгоритм NB.

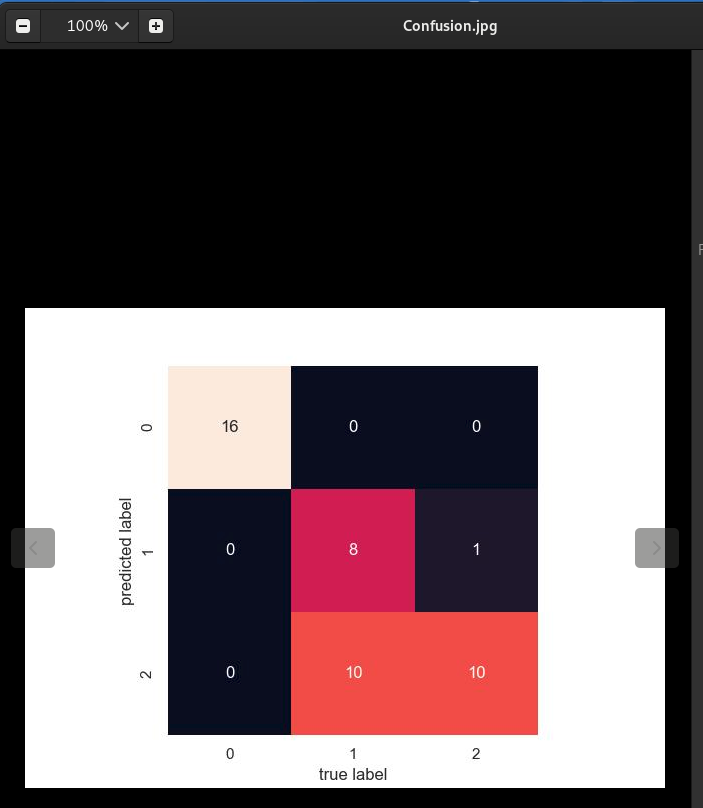
**Завдання 2.5. Класифікація даних лінійним класифікатором Ridge**

Код програми:



Результат роботи програми:





При налаштуванні класифікатора було використано такі параметри: tol (точність рішення) та solver (рішення для обрахунків). Коефіцієнт Коена Каппа є мірою узгодженості між двома спостерігачами, які оцінюють якісну змінну.

Коефіцієнт кореляції Метьюза використовується для оцінки зв'язку між результатами діагностики та станом пацієнта. Він розраховується як відношення кількості правильних діагнозів до загальної кількості діагнозів, враховуючи відсутність зв'язку випадкових результатів.

***Висновки***: у ході виконання даної лабораторної роботи я, використовуючи спеціалізовані бібліотеки та мову програмування Python, дослідив класифікацію даних різними методами та за допомогою метрик якості класифікації оцінив їх ефективності.