**ЛАБОРАТОРНА РОБОТА № 1**

**ПОПЕРЕДНЯ ОБРОБКА ТА КОНТРОЛЬОВАНА КЛАСИФІКАЦІЯ**

**ДАНИХ**

***Мета:***  використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити попередню обробку та класифікацію даних.

**Хід роботи:**

**Завдання 1:**  Попередня обробка даних.

**Результат:**



Рис. 1.1. Бінаризація

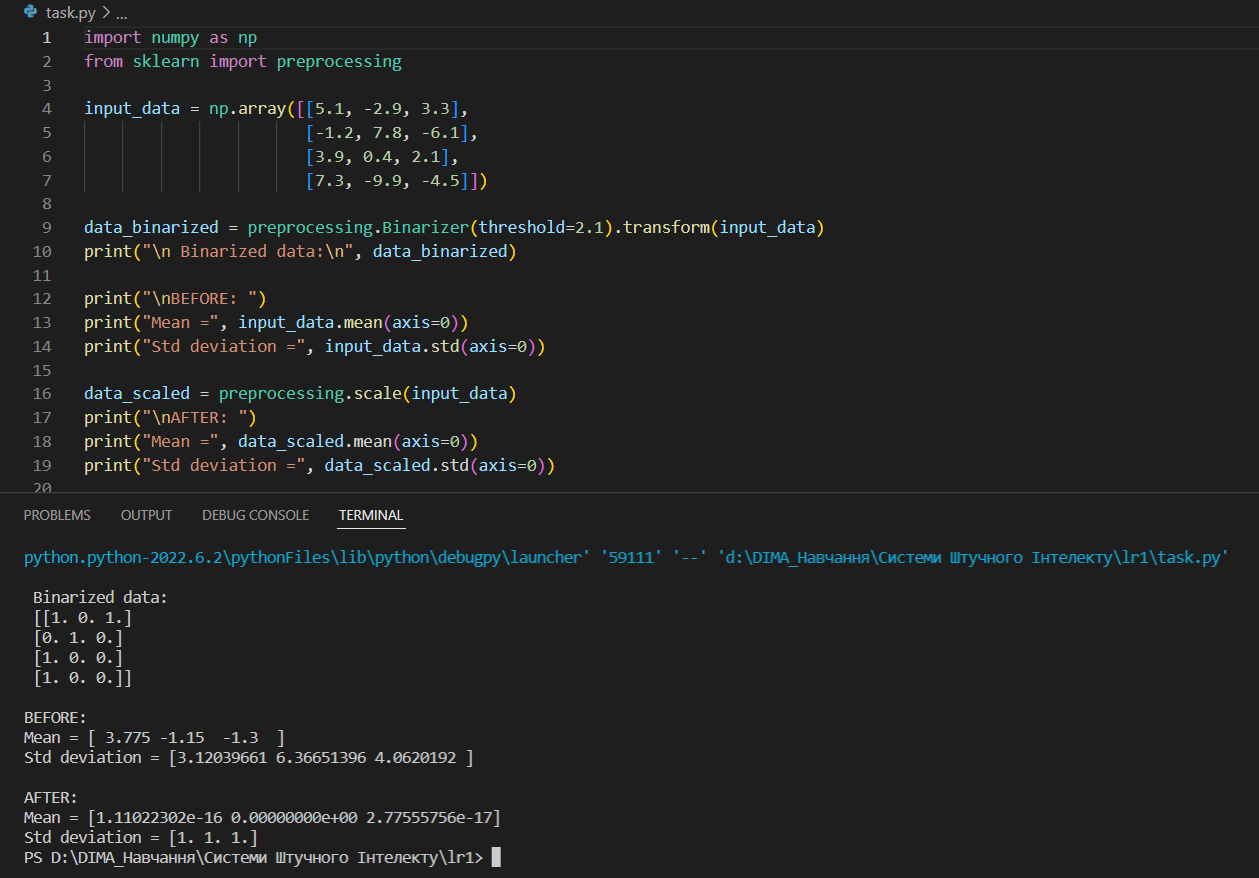


Рис. 1.2. Виключення середнього

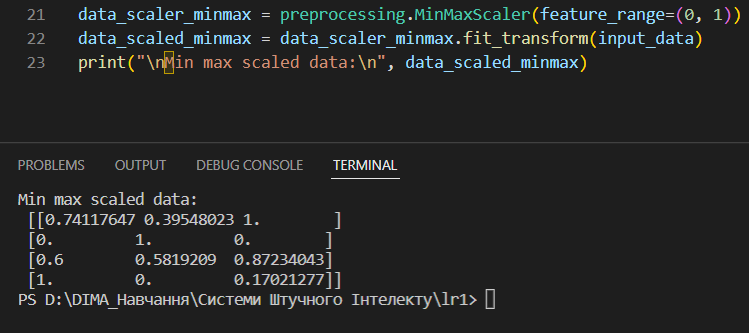


Рис. 1.3. Масштабування

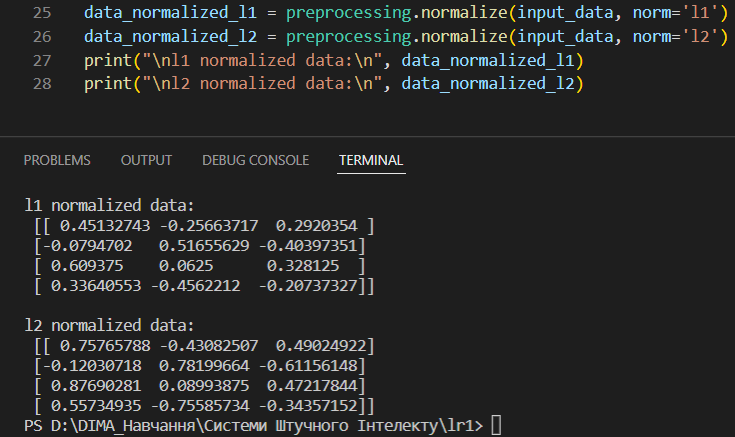


Рис. 1.4. Нормалізація даних

L1-нормалізація використовує метод найменших абсолютних відхилень (Least Absolute Deviations), що забезпечує рівність 1 суми абсолютних значень в кожному ряду.

L2-нормалізація використовує метод найменших квадратів, що забезпечує рівність 1 суми квадратів 4 значень.

Звідси можна зробити висновок, що L1-нормалізація є надійнішою за L2-нормалізацію.

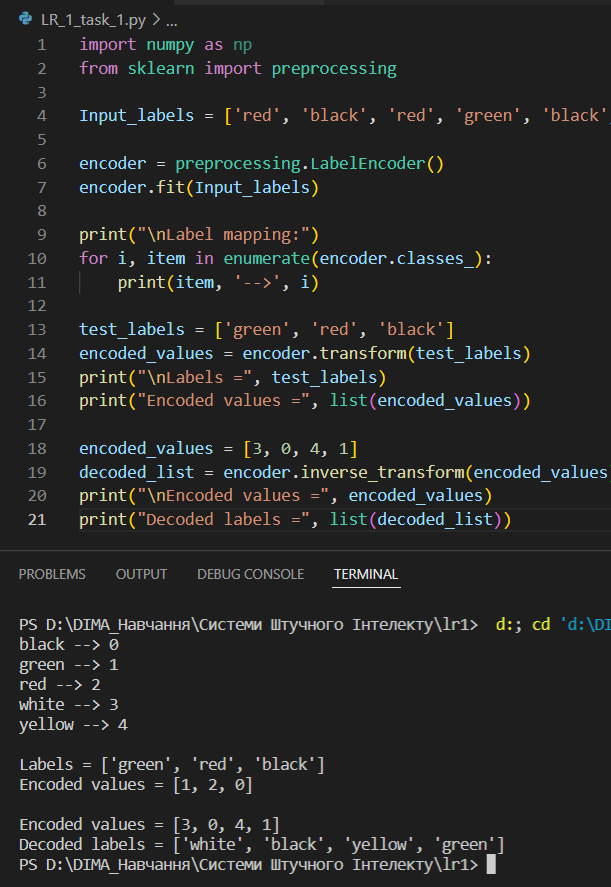


Рис. 1.5. Кодування міток

**Завдання 2:**  Попередня обробка нових даних.



**Результат:**

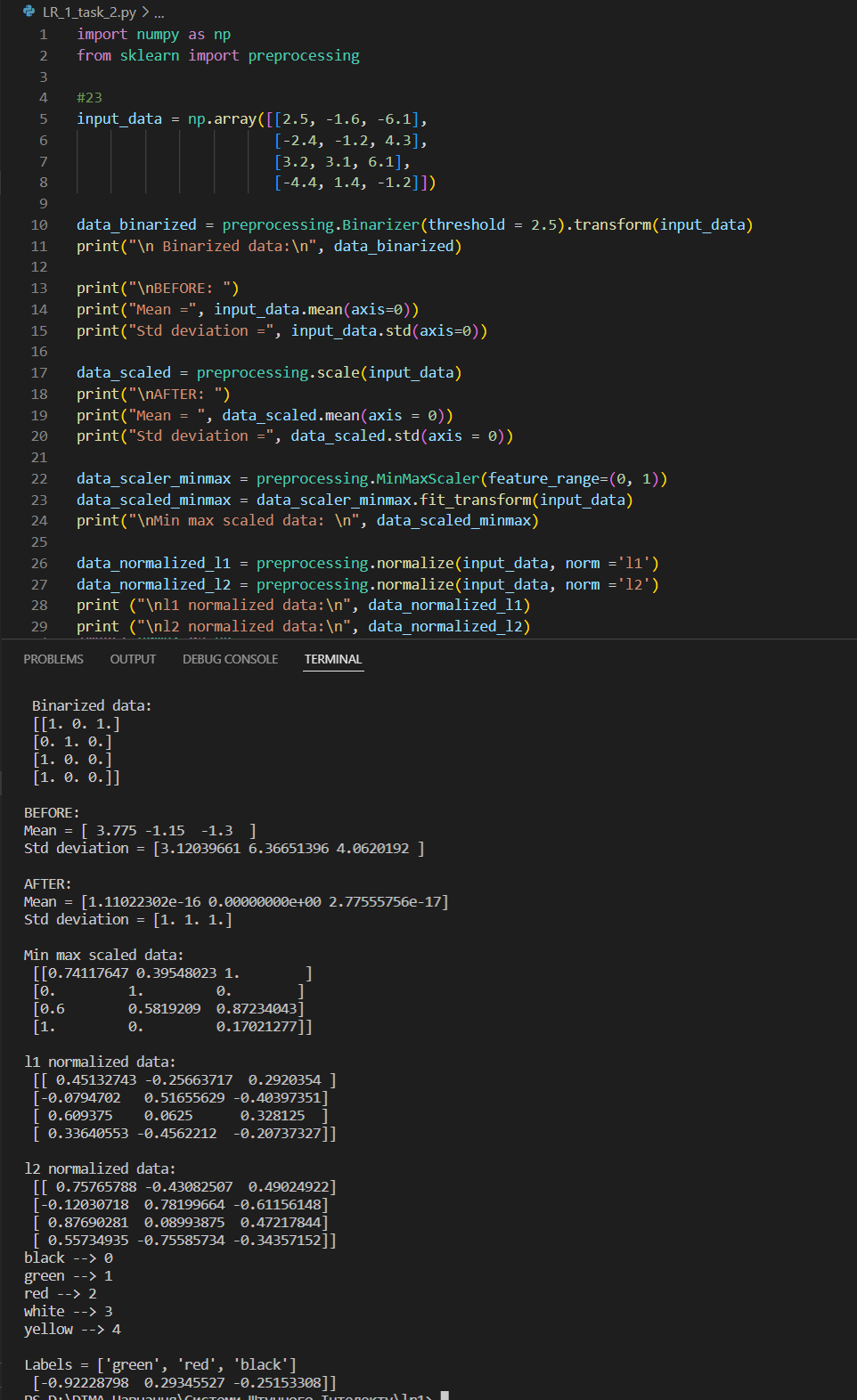


Рис. 2.1. Результат виконання

**Завдання 3:**  Класифікація логістичною регресією або логістичний класифікатор.

**Результат:**

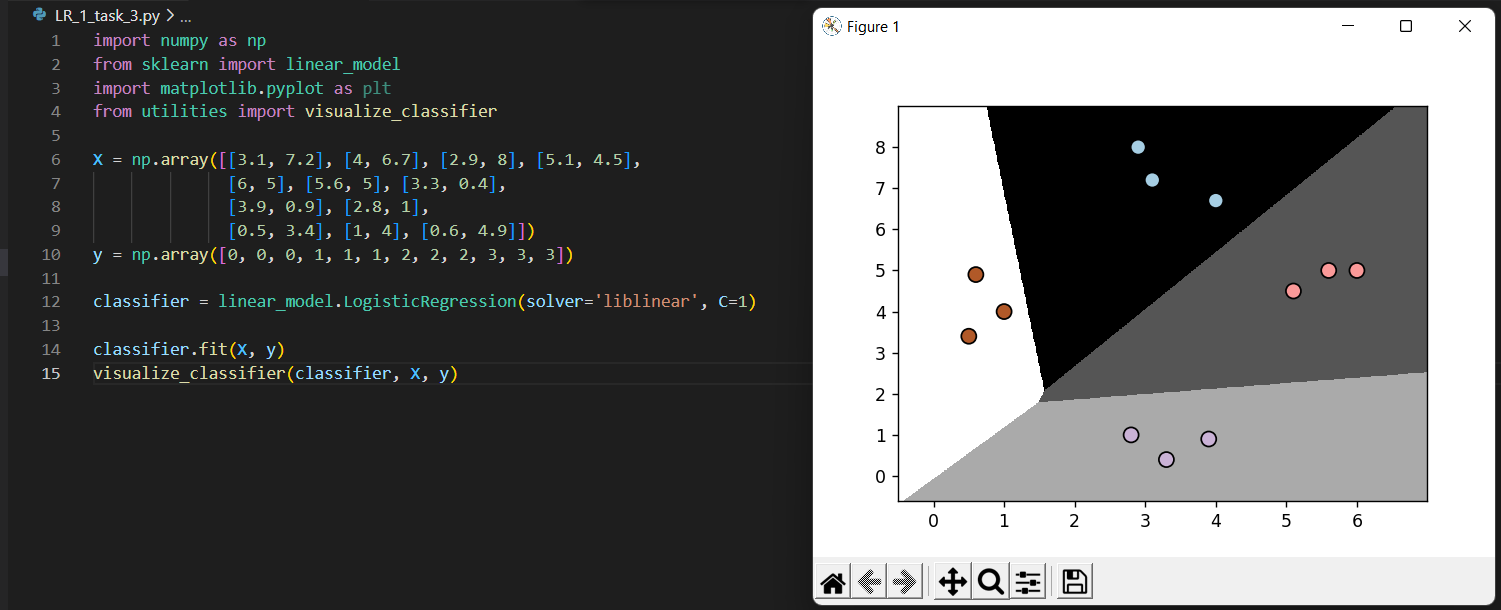


Рис. 3.1. Результат виконання

**Завдання 4:**  Класифікація наївним байєсовським класифікатором.

**Результат:**

Обидва прогона дали однаковий результат, тому як генерувались однакові набори даних для навчань і тестувань.

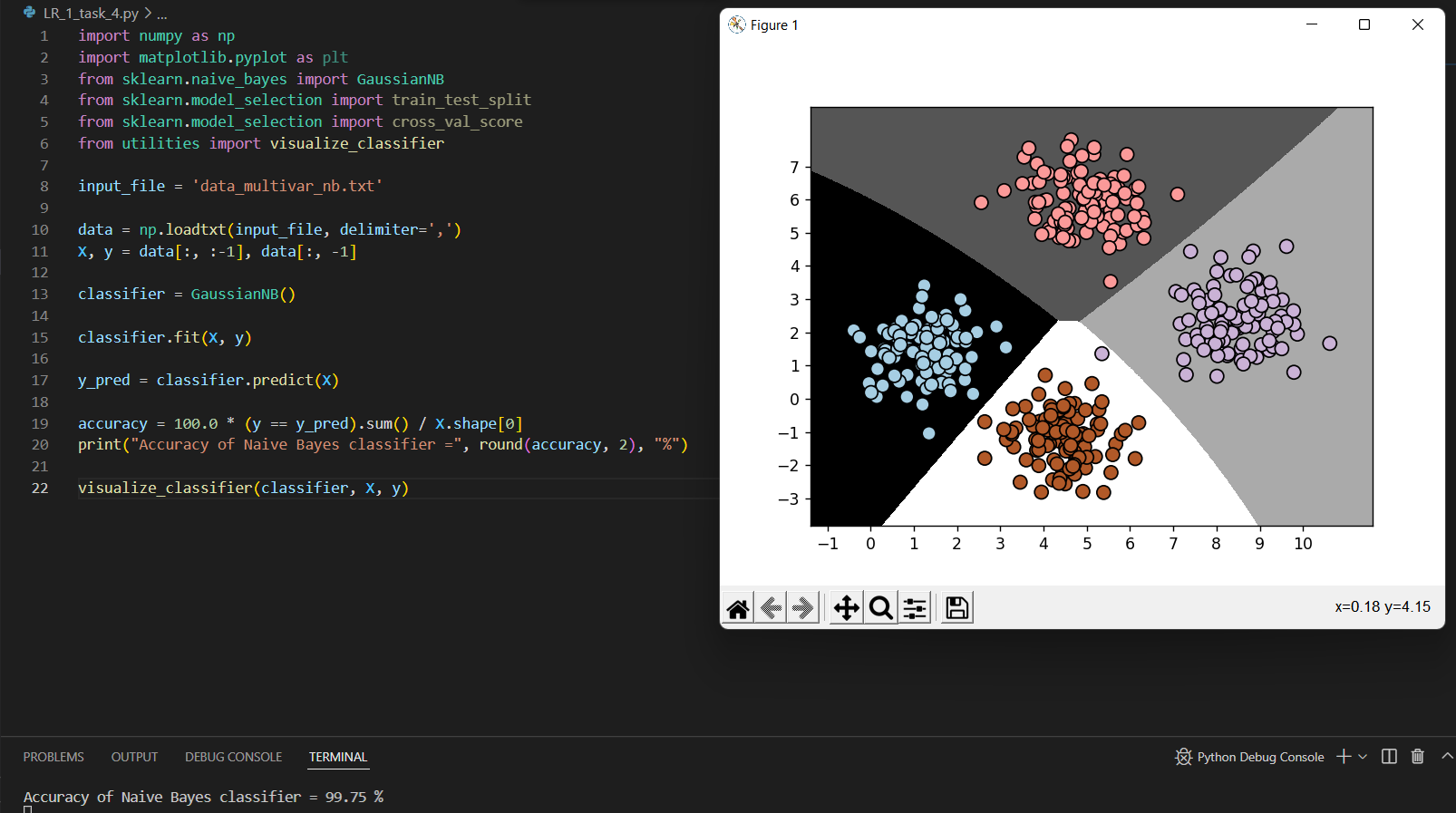


Рис. 4.1. Класифікація наївним байєсовським класифікатором

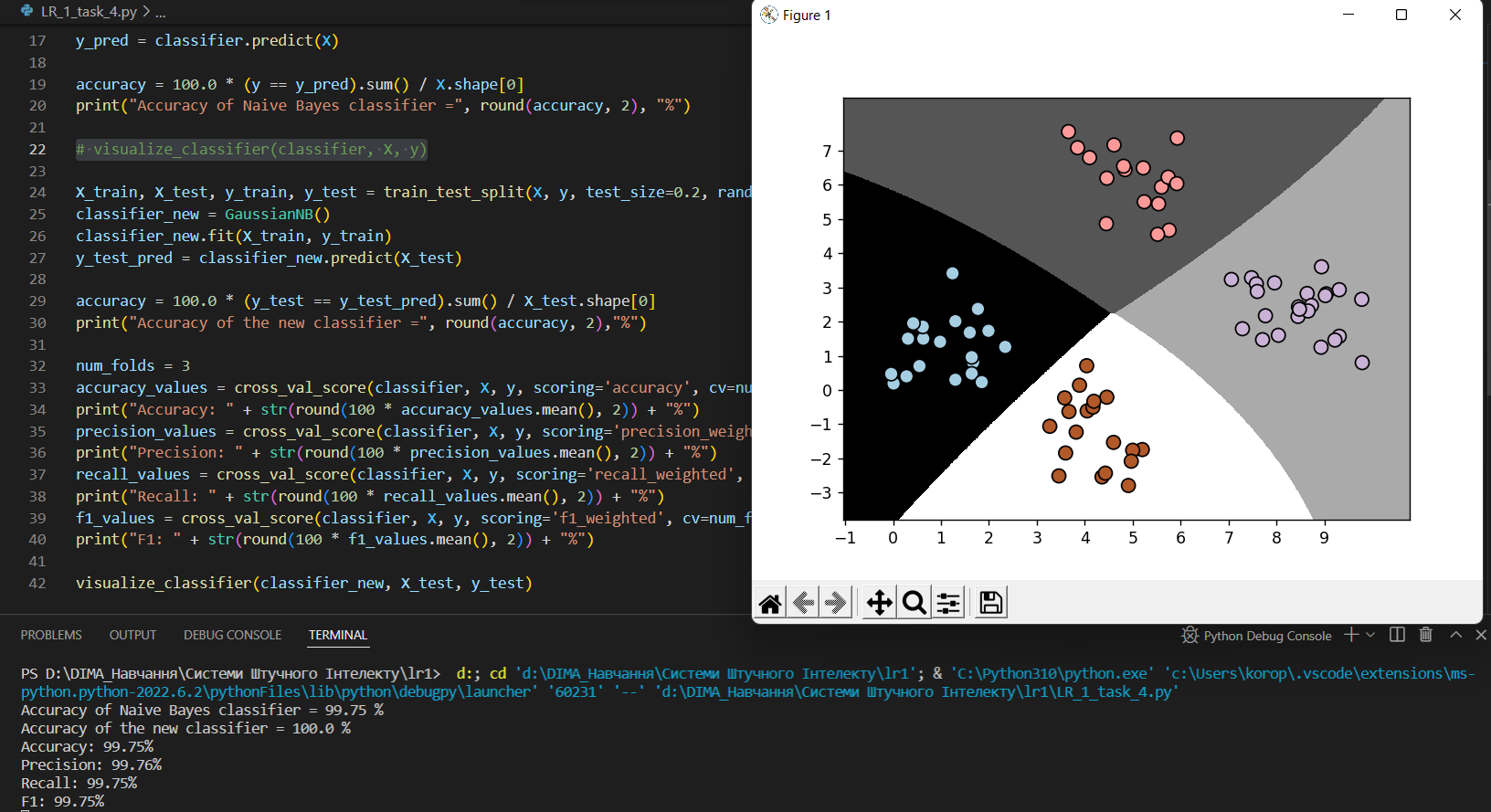


Рис. 4.1. Класифікація наївним байєсовським класифікатором з обчисленням якості точності і повноти

**Завдання 5:**  Вивчити метрики якості класифікації.

**Результат:**

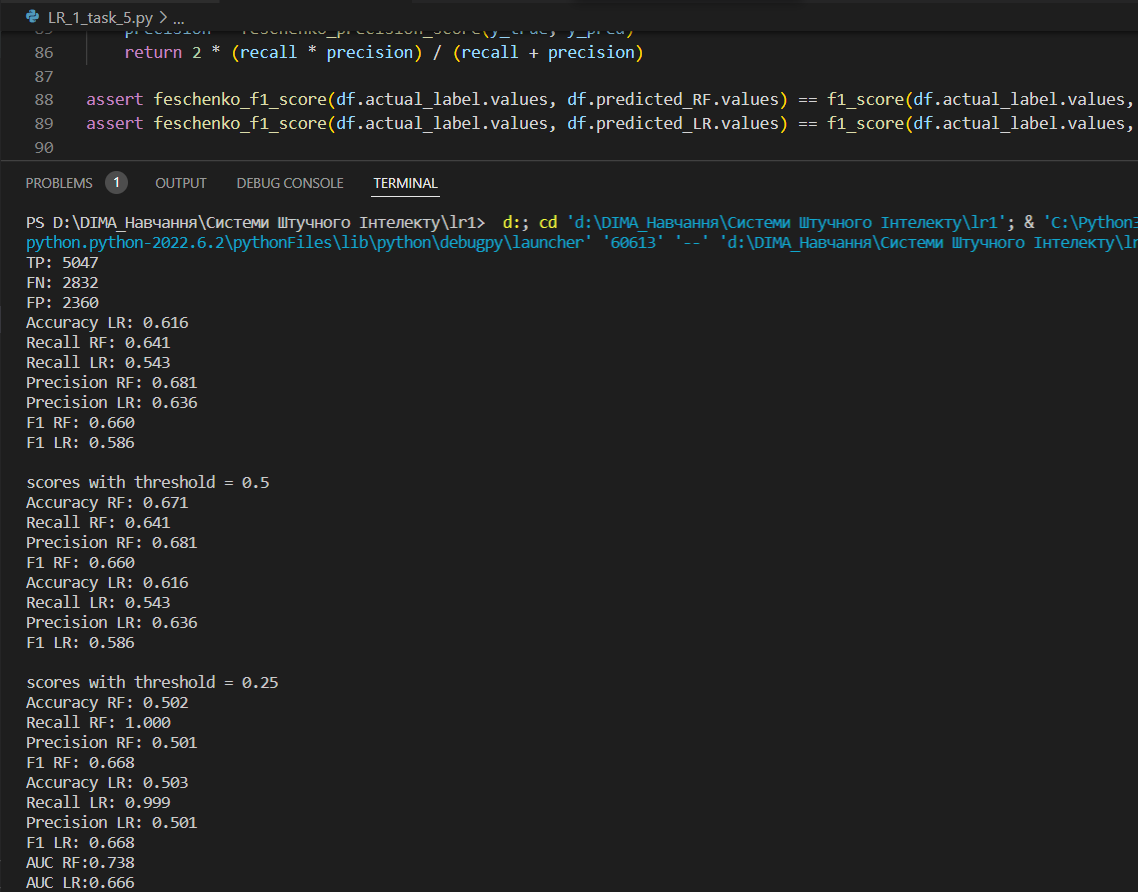


Рис. 5.1. Порівняння моделей RF та LF на кроках 0.25 і 0.5

Порівнюючи моделі при порозі 0.25 якість та точність вищі (RF), при порозі 0.5 модель справляється краще (LF). Тому залежно від вхідних даних необхідно робити вибір.

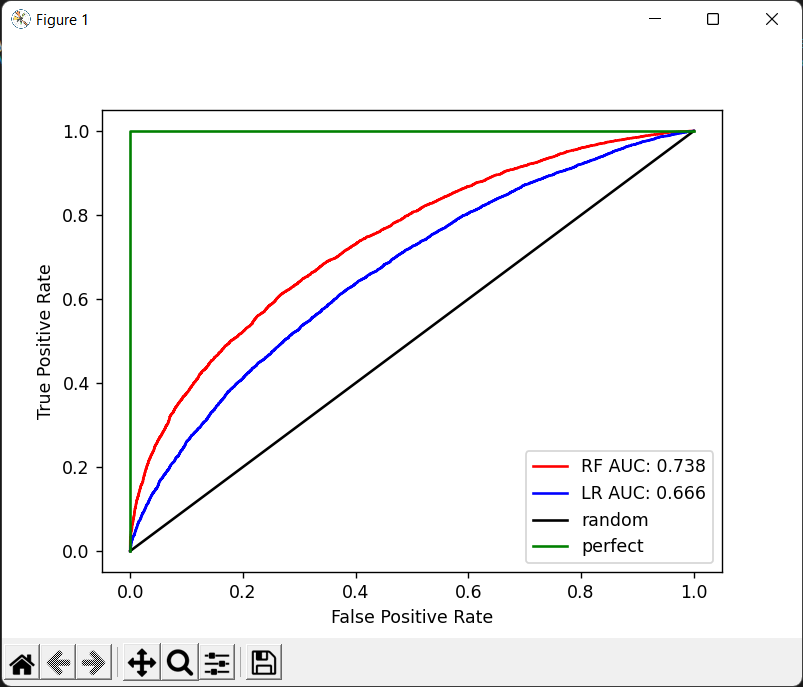
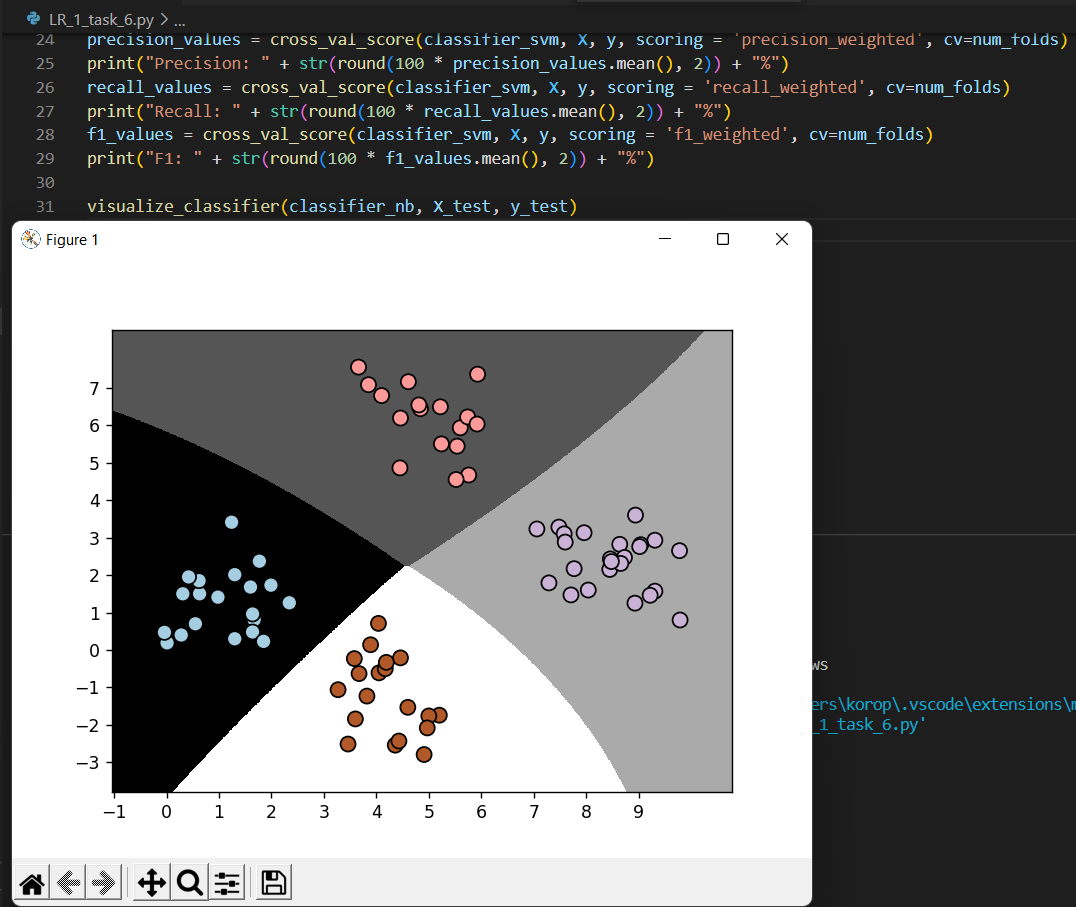


Рис. 5.2. Порівняння моделей за допомогою кривих ROC

**Завдання 6:**  Розробіть програму класифікації даних в файлі data\_multivar\_nb.txt за допомогою машини опорних векторів (Support Vector Machine - SVМ). Розрахуйте показники якості класифікації. Порівняйте їх з показниками наївного байєсівського класифікатора. Зробіть висновки яку модель класифікації краще обрати і чому.

**Результат:**



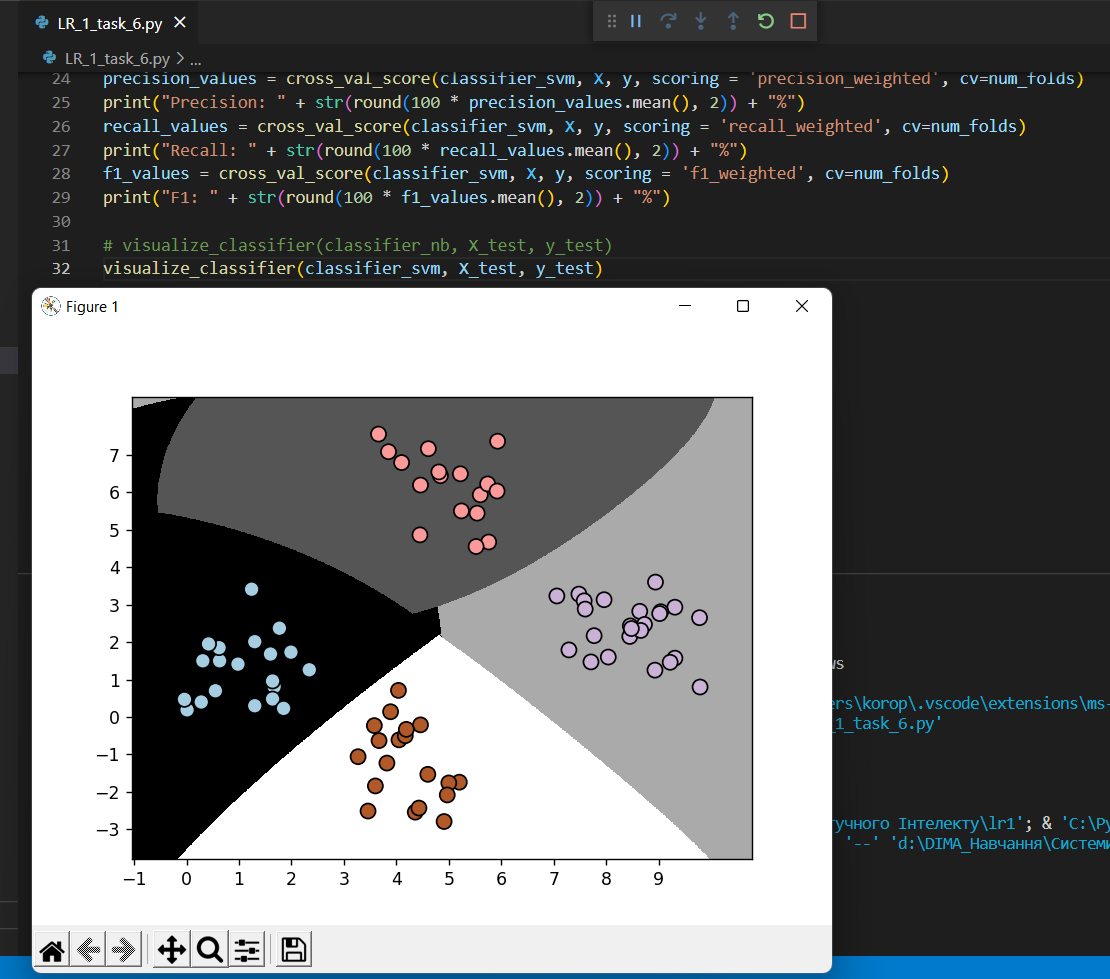


Рис. 6.1. Порівняння класифікаторів наївного байєса та SVM

**Висновок:** на цій лабораторній роботі ми використовуючи спеціалізовані бібліотеки та мову програмування Python дослідили попередню обробку та класифікацію даних.