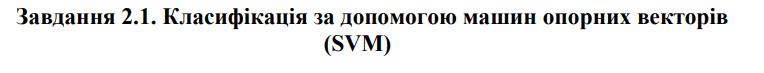
**ЛАБОРАТОРНА РОБОТА № 2**

**Тема**: ПОРІВНЯННЯ МЕТОДІВ КЛАСИФІКАЦІЇ ДАНИХ.

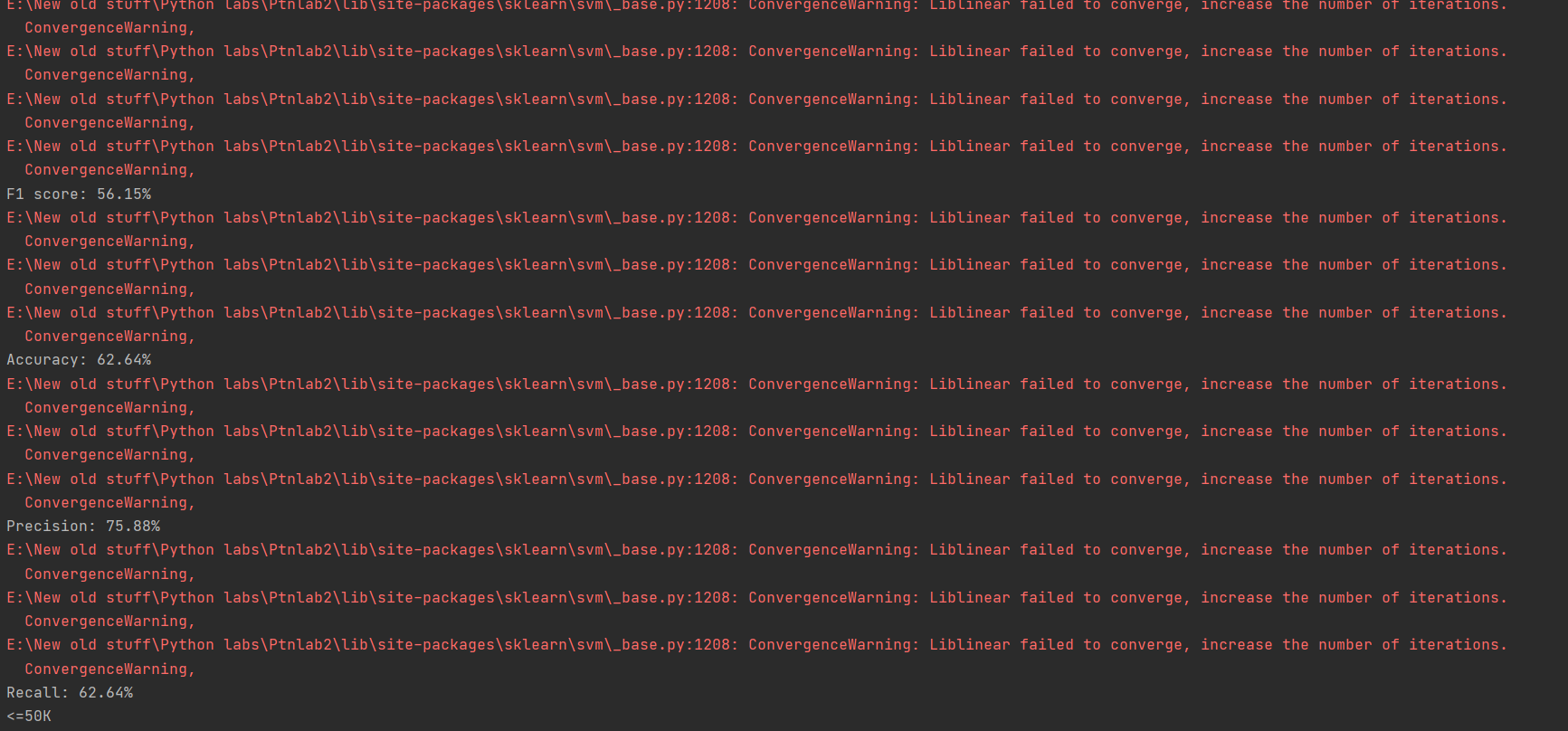
**Мета**: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити різні методи класифікації даних та навчитися їх порівнювати.

**Хід роботи:**

**Task1**

****

****

****

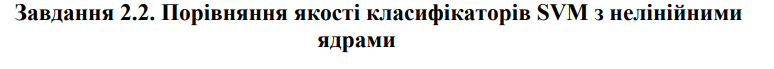
**F1 score: 56.15%**

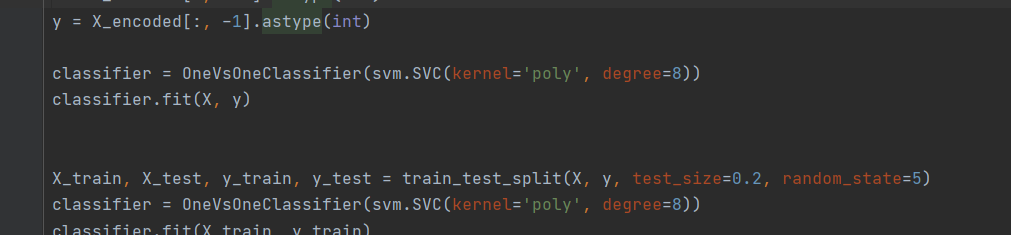
**Accuracy: 62.64%**

**Precision: 75.88%**

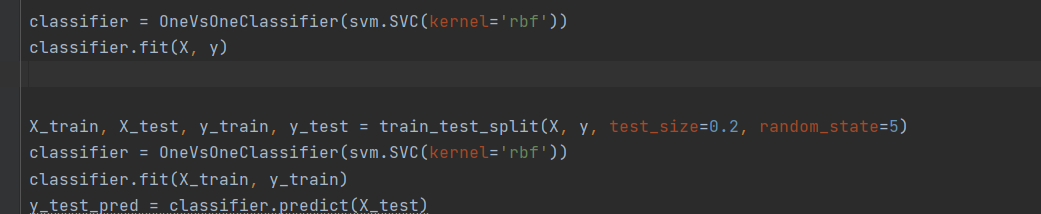
**Recall: 62.64%**

**Task2**

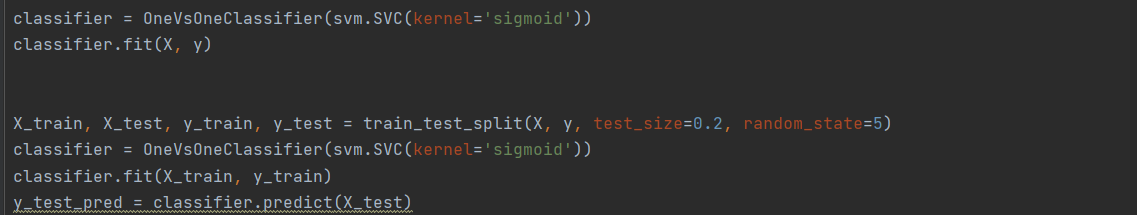
****

****

Поліномінальне ядро

****

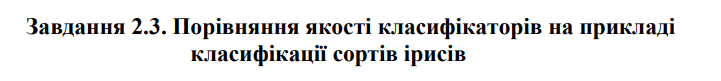
Гаусового ядро

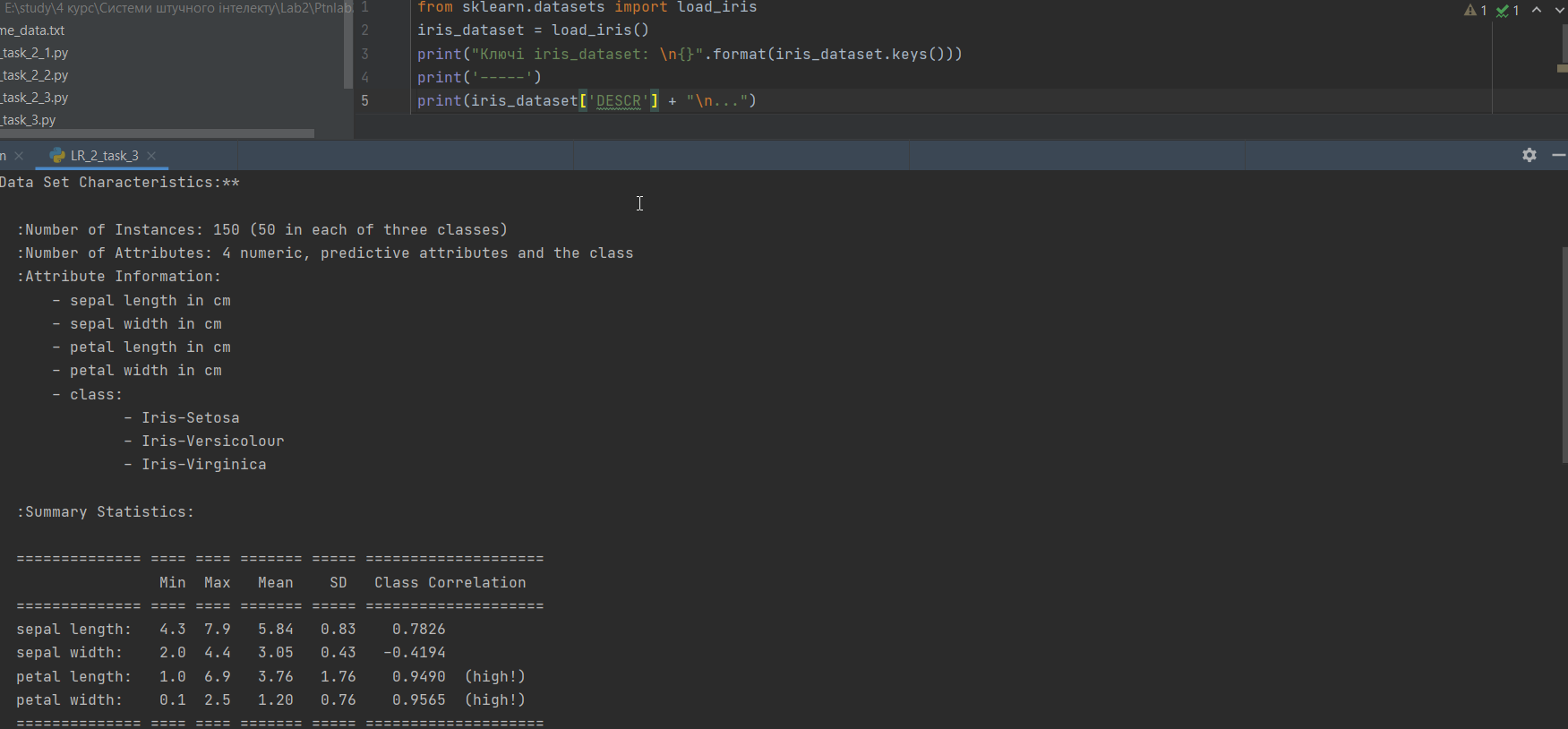
****

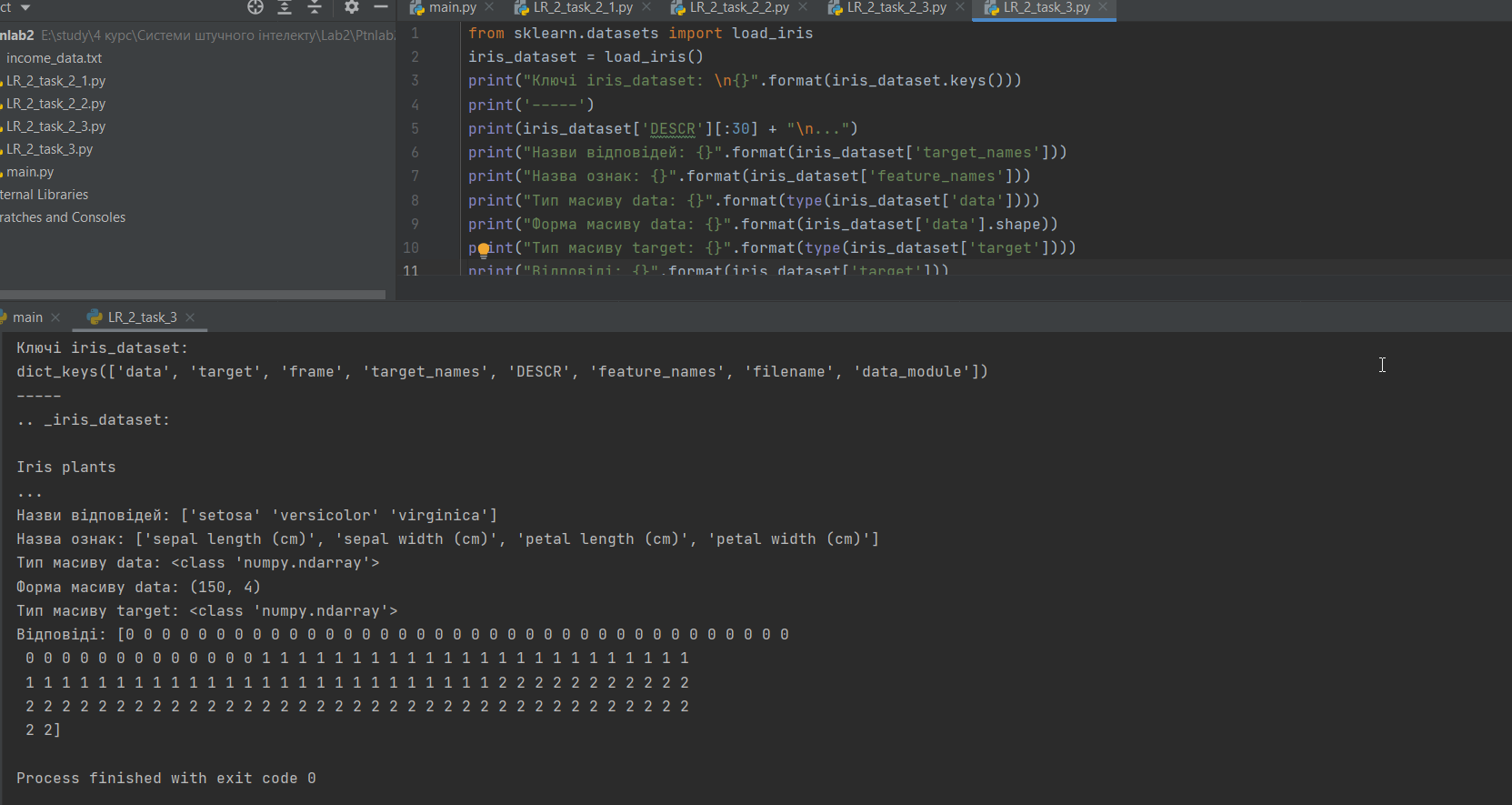
Сигмоїдальне ядро

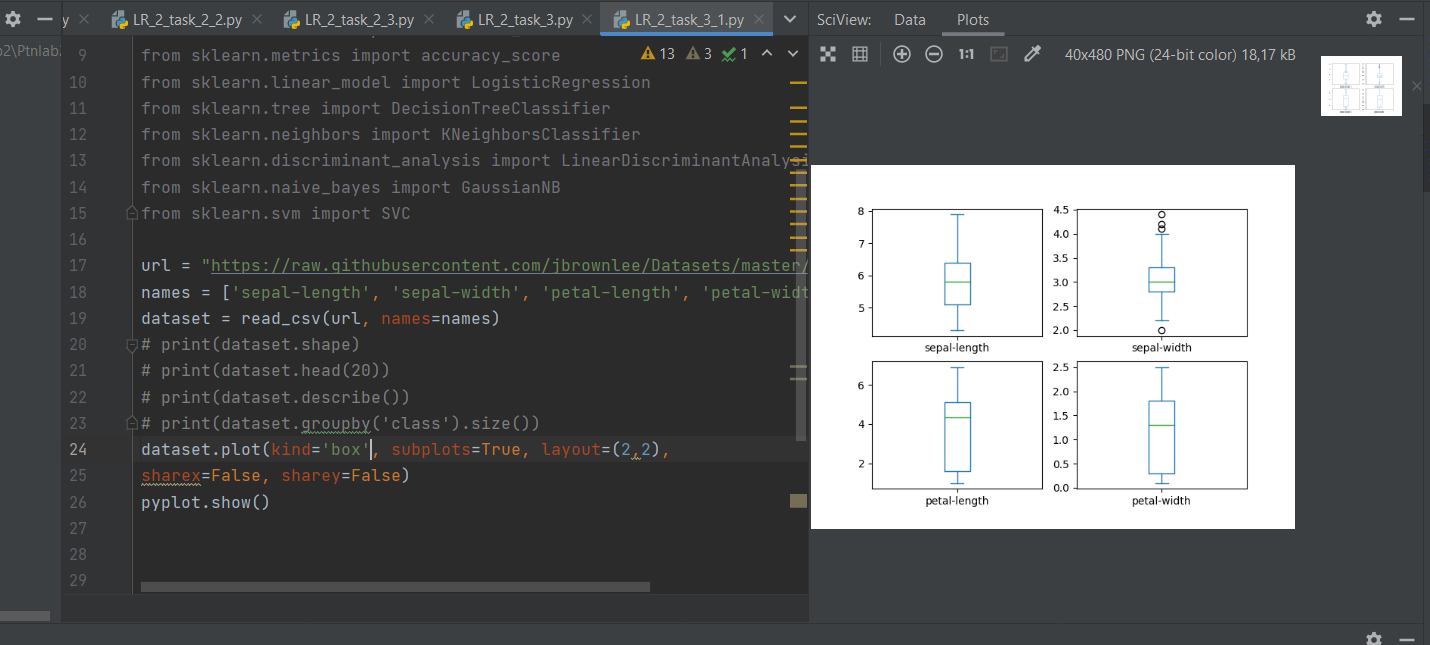
За даних умов нерівномірності використання даних, найкращий результат надає гаусове ядро. За використання всіх наявних даних, результати можуть бути іншими, проте для їх отримання необхідно надати дуже багато часу на навчання поліномінальому ядру.

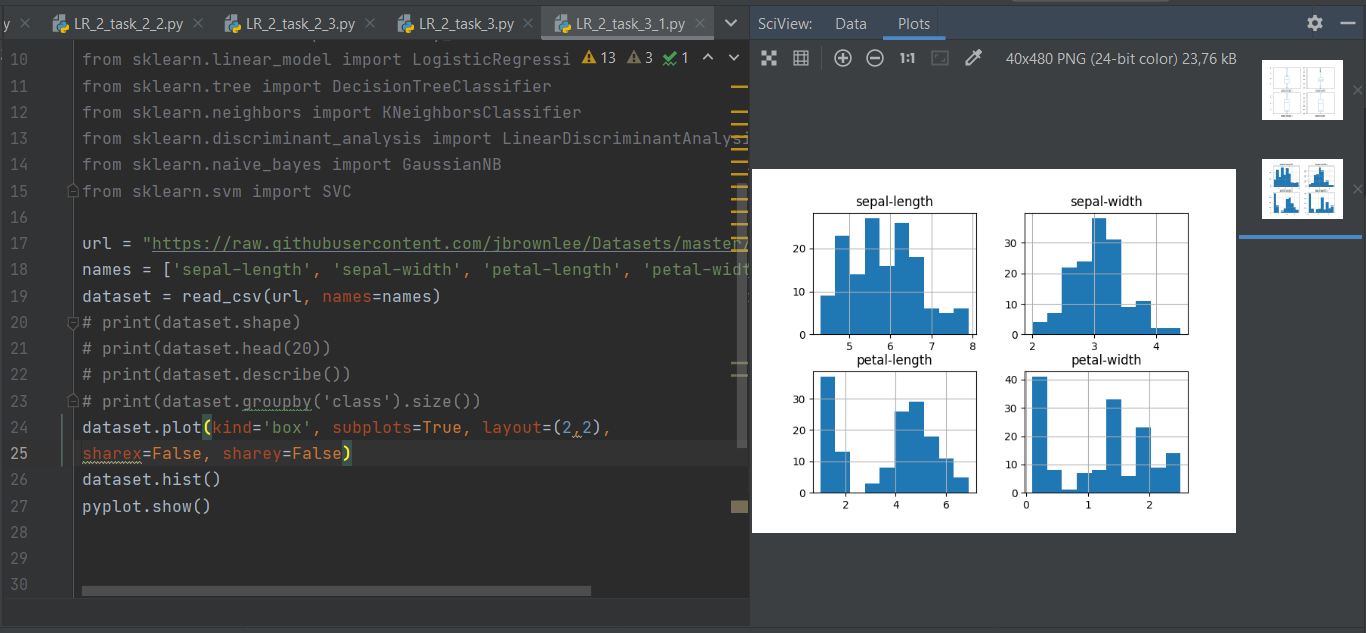
**Task3**

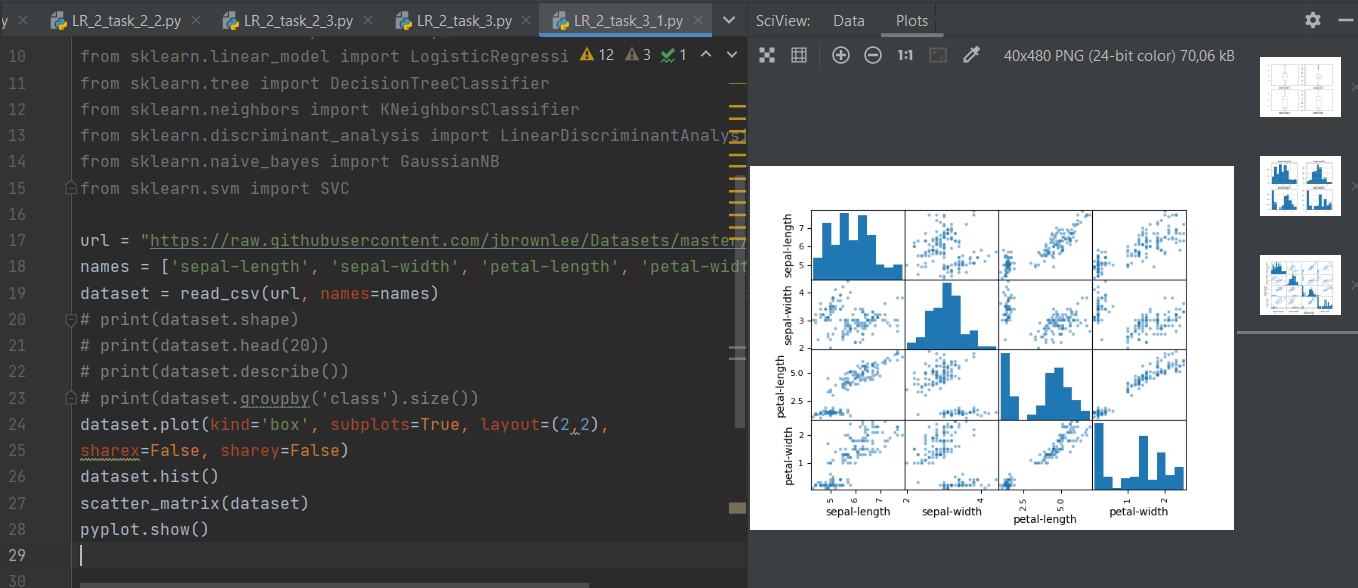
****

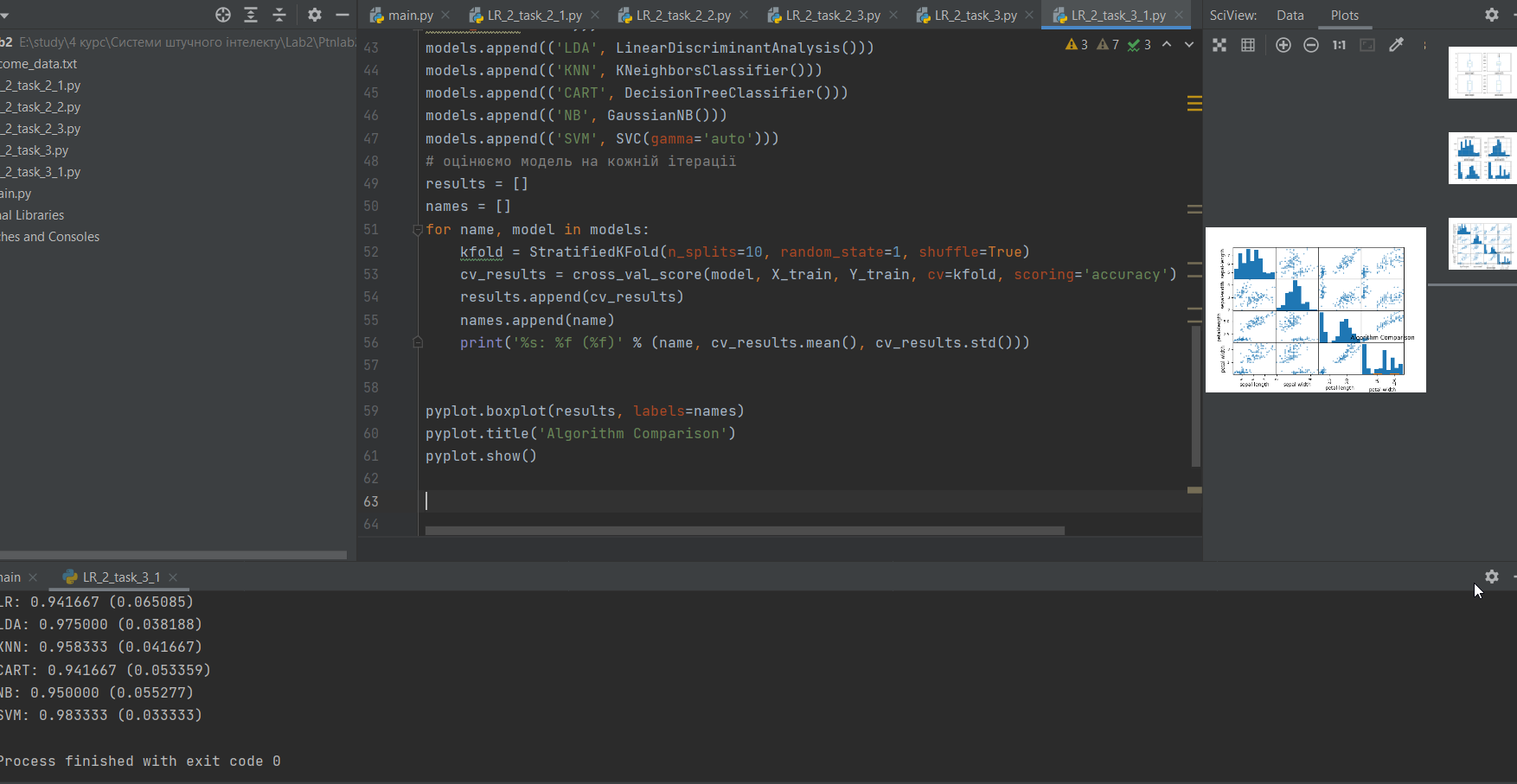
****

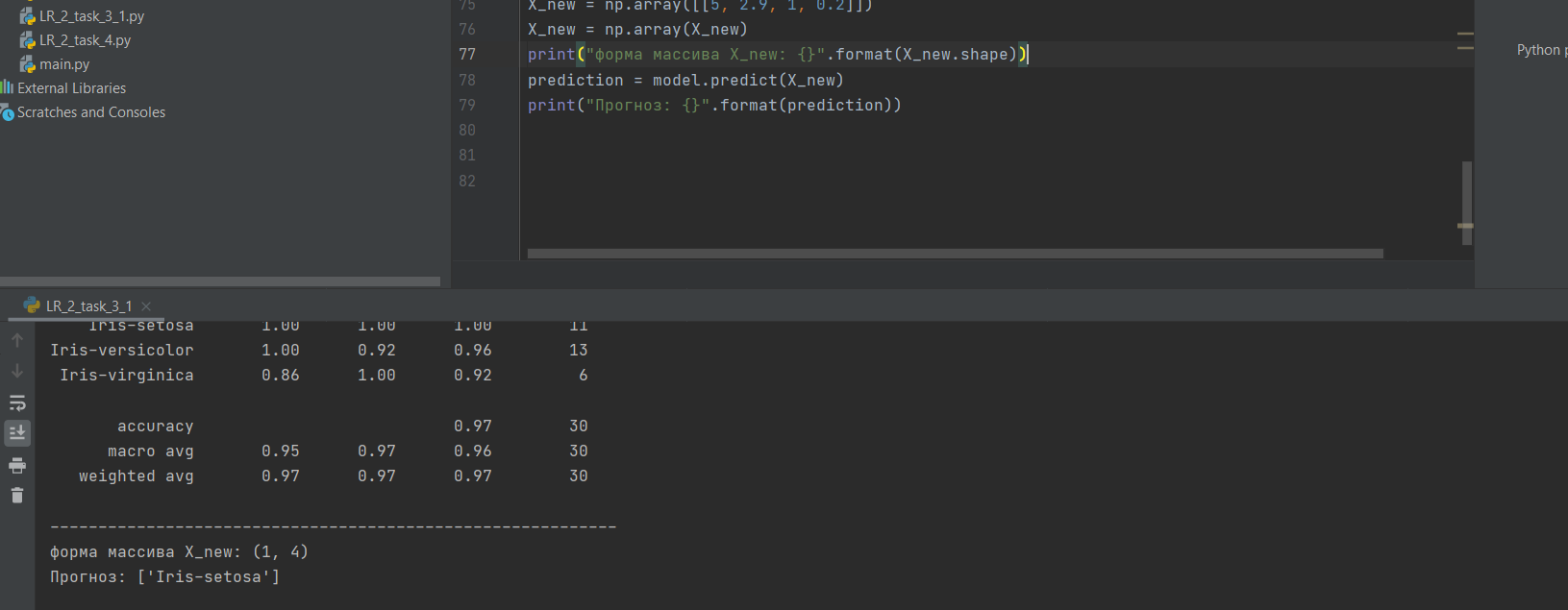
****

****

****

****

****

****

**LR: 0.941667 (0.065085)**

**LDA: 0.975000 (0.038188)**

**KNN: 0.958333 (0.041667)**

**CART: 0.941667 (0.038188)**

**NB: 0.950000 (0.055277)**

**SVM: 0.983333 (0.033333)**

**0.9666666666666667**

**[[11 0 0]**

**[ 0 12 1]**

**[ 0 0 6]]**

**precision recall f1-score support**

**Iris-setosa 1.00 1.00 1.00 11**

**Iris-versicolor 1.00 0.92 0.96 13**

**Iris-virginica 0.86 1.00 0.92 6**

**accuracy 0.97 30**

**macro avg 0.95 0.97 0.96 30**

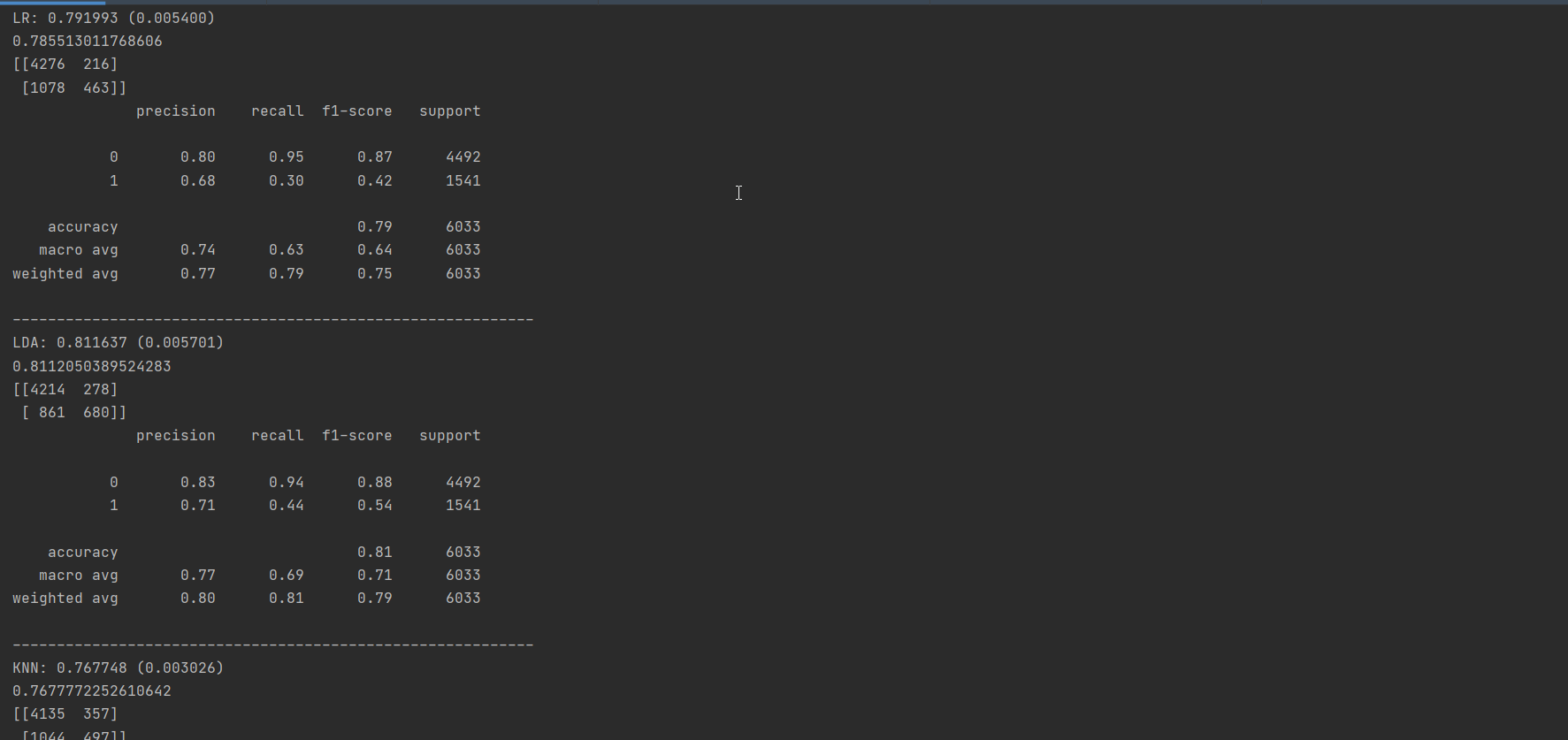
**weighted avg 0.97 0.97 0.97 30**

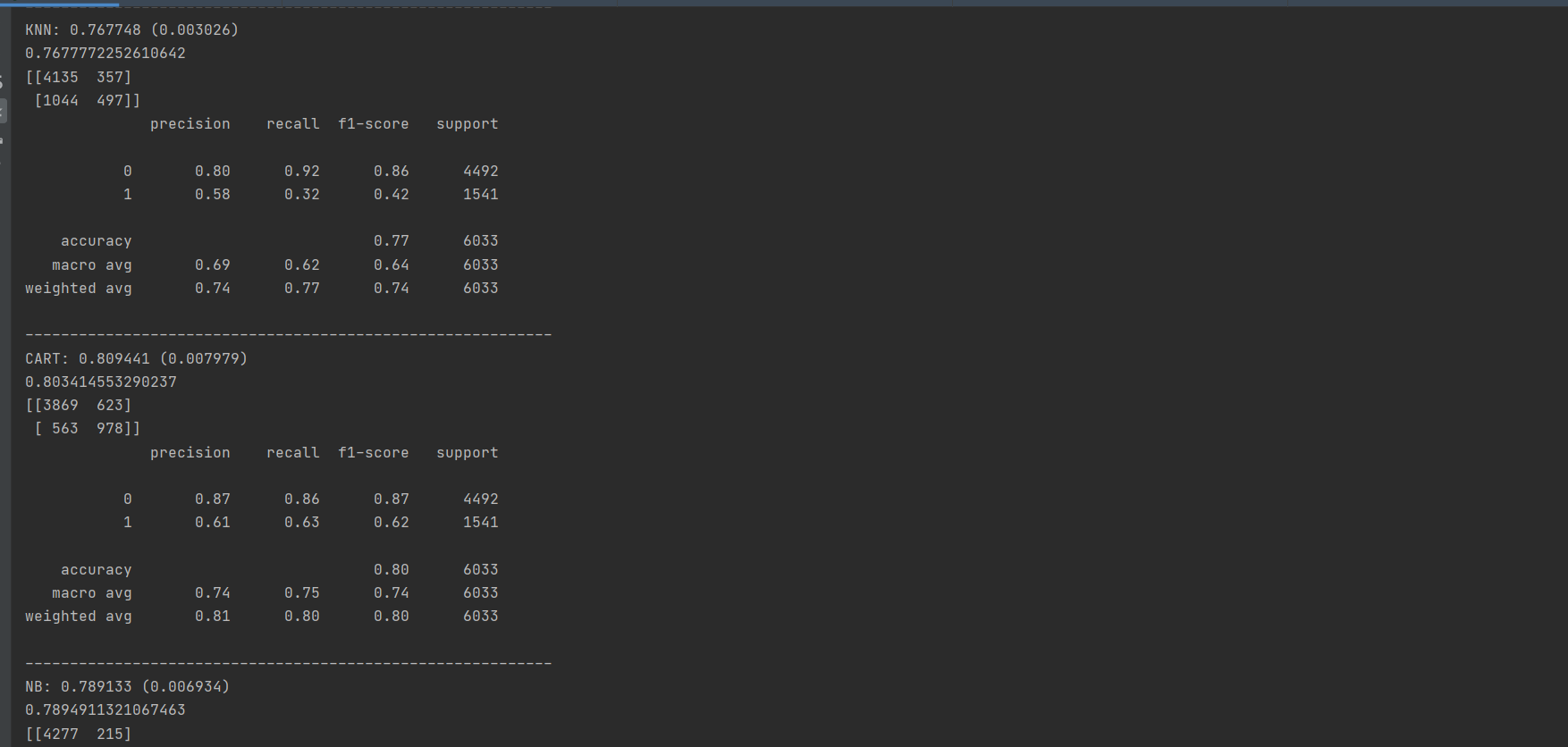
**-----------------------------------------------------------**

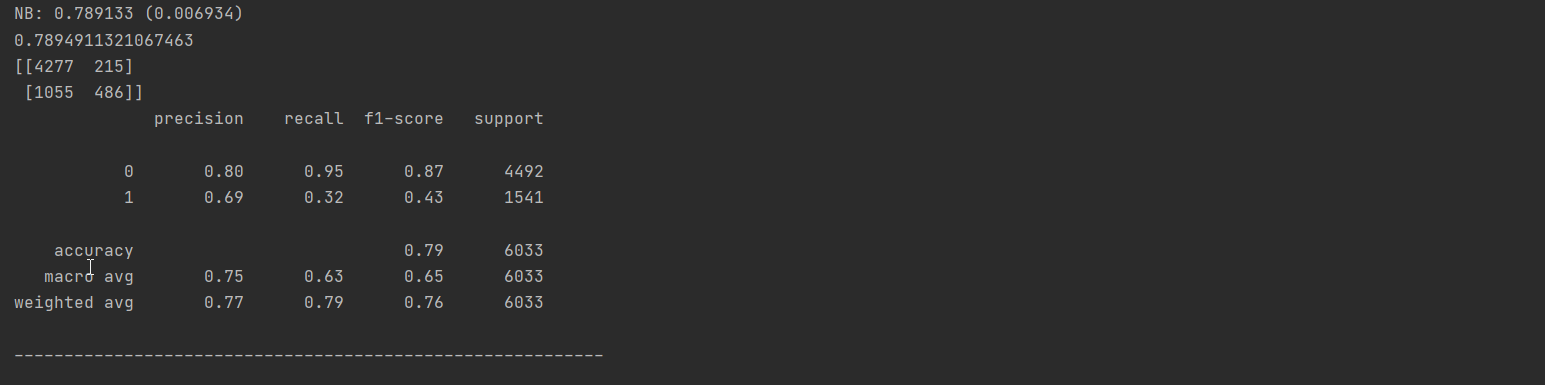
**форма массива X\_new: (1, 4)**

**Прогноз: ['Iris-setosa']**

**Task4**

****

****

****

**LR: 0.791993 (0.005400)**

**0.785513011768606**

**[[4276 216]**

**[1078 463]]**

**precision recall f1-score support**

**0 0.80 0.95 0.87 4492**

**1 0.68 0.30 0.42 1541**

**accuracy 0.79 6033**

**macro avg 0.74 0.63 0.64 6033**

**weighted avg 0.77 0.79 0.75 6033**

**-----------------------------------------------------------**

**LDA: 0.811637 (0.005701)**

**0.8112050389524283**

**[[4214 278]**

**[ 861 680]]**

**precision recall f1-score support**

**0 0.83 0.94 0.88 4492**

**1 0.71 0.44 0.54 1541**

**accuracy 0.81 6033**

**macro avg 0.77 0.69 0.71 6033**

**weighted avg 0.80 0.81 0.79 6033**

**-----------------------------------------------------------**

**KNN: 0.767748 (0.003026)**

**0.7677772252610642**

**[[4135 357]**

**[1044 497]]**

**precision recall f1-score support**

**0 0.80 0.92 0.86 4492**

**1 0.58 0.32 0.42 1541**

**accuracy 0.77 6033**

**macro avg 0.69 0.62 0.64 6033**

**weighted avg 0.74 0.77 0.74 6033**

**-----------------------------------------------------------**

**CART: 0.807244 (0.007977)**

**0.8063981435438422**

**[[3884 608]**

**[ 560 981]]**

**precision recall f1-score support**

**0 0.87 0.86 0.87 4492**

**1 0.62 0.64 0.63 1541**

**accuracy 0.81 6033**

**macro avg 0.75 0.75 0.75 6033**

**weighted avg 0.81 0.81 0.81 6033**

**-----------------------------------------------------------**

**NB: 0.789133 (0.006934)**

**0.7894911321067463**

**[[4277 215]**

**[1055 486]]**

**precision recall f1-score support**

**0 0.80 0.95 0.87 4492**

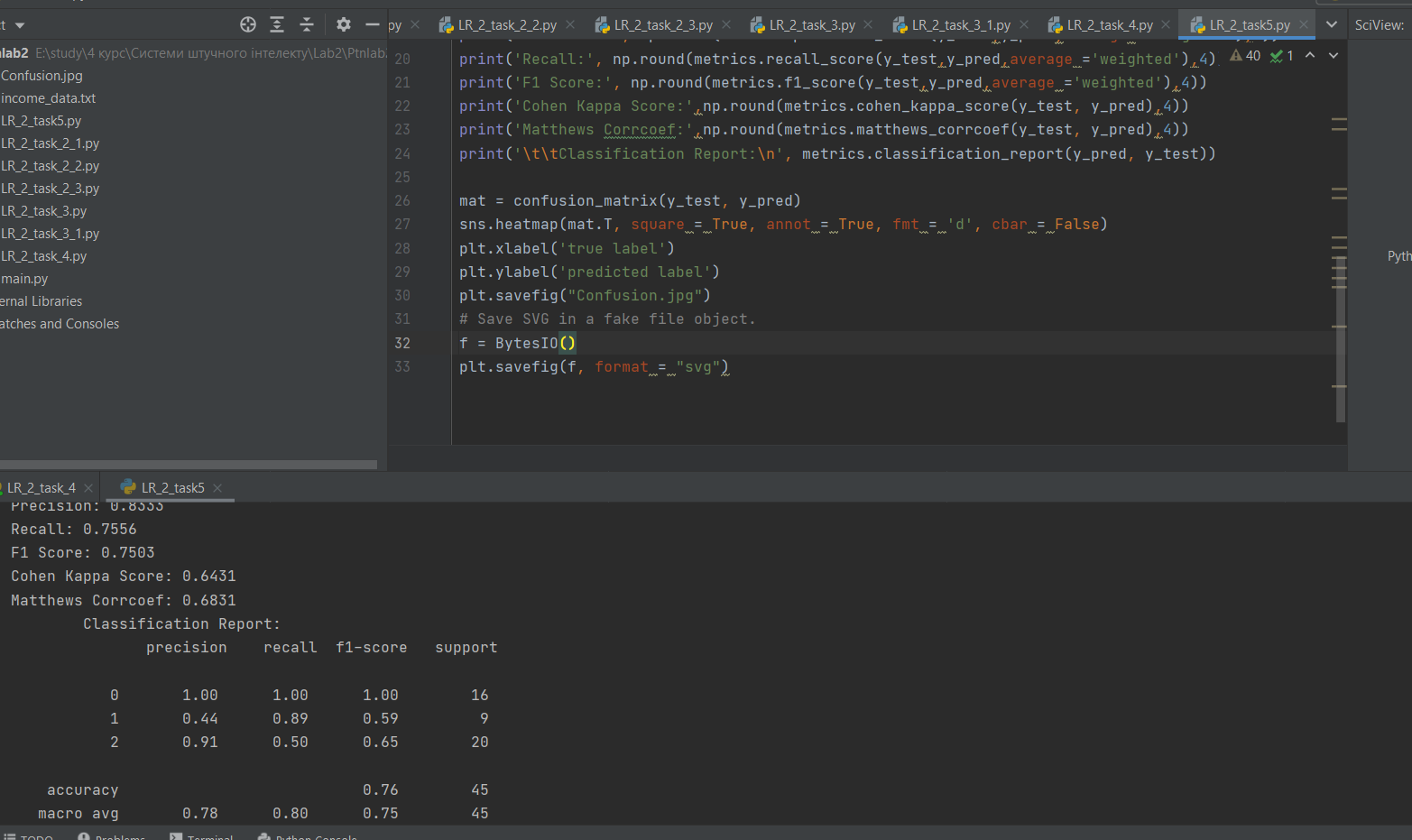
**1 0.69 0.32 0.43 1541**

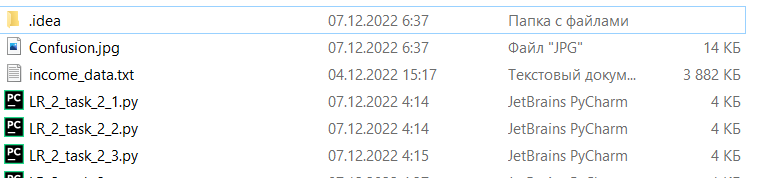
**accuracy 0.79 6033**

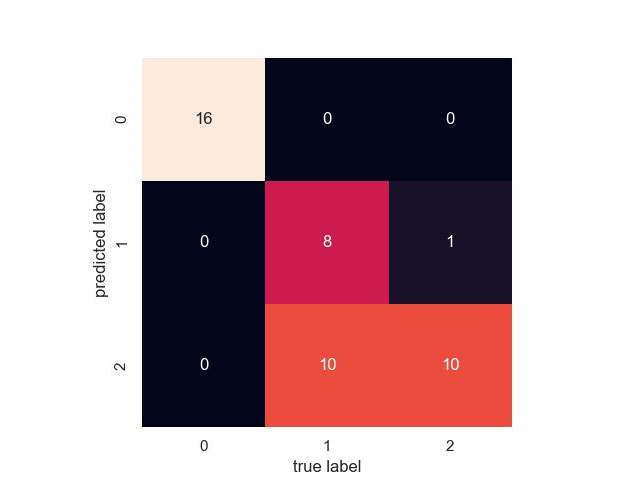
**macro avg 0.75 0.63 0.65 6033**

**weighted avg 0.77 0.79 0.76 6033**

**Task5**

****

****

****

**Confusion matrix (матриця помилок) -** це таблиця особливого компонування, що дає можливість унаочнювати продуктивність алгоритму, зазвичай [керованого навчання](https://uk.wikipedia.org/wiki/%D0%9A%D0%B5%D1%80%D0%BE%D0%B2%D0%B0%D0%BD%D0%B5_%D0%BD%D0%B0%D0%B2%D1%87%D0%B0%D0%BD%D0%BD%D1%8F).

В [прогнозній аналітиці](https://uk.wikipedia.org/w/index.php?title=%D0%9F%D1%80%D0%BE%D0%B3%D0%BD%D0%BE%D0%B7%D0%BD%D0%B0_%D0%B0%D0%BD%D0%B0%D0%BB%D1%96%D1%82%D0%B8%D0%BA%D0%B0&action=edit&redlink=1), табли́ця невідповідностей — це таблиця з двома рядками та двома стовпцями, що повідомляє число хибно позитивних ([англ.](https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%B0_%D0%BC%D0%BE%D0%B2%D0%B0) false positives), хибно негативних ([англ.](https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%B0_%D0%BC%D0%BE%D0%B2%D0%B0) false negatives), істинно позитивних ([англ.](https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%B0_%D0%BC%D0%BE%D0%B2%D0%B0) true positives) та істинно негативних ([англ.](https://uk.wikipedia.org/wiki/%D0%90%D0%BD%D0%B3%D0%BB%D1%96%D0%B9%D1%81%D1%8C%D0%BA%D0%B0_%D0%BC%D0%BE%D0%B2%D0%B0) true negatives) результатів. Це уможливлює аналіз, докладніший за просту пропорцію правильних класифікацій (точність). Точність видаватиме оманливі результати, якщо набір даних є незбалансованим, тобто коли число спостережень в різних класах сильно різниться.

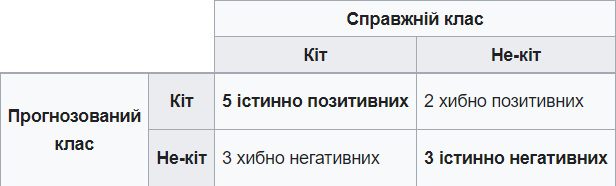
припустімо, що ми перевіряємо класифікатор, який розрізняє котів та псів. Для цього ці 13 зображень подамо у класифікатор, і, нехай, класифікатор зробив 8 точних прогнозів, та 5 помилок: для 3 котів було помилково зроблено прогноз, що це пси (перші три прогнози), й для 2 псів було зроблено помилковий прогноз, що це коти (крайні 2 прогнози).

прогнозований = [0,0,0,1,1,1,1,1,0,0,0,1,1]

Маючи ці два мічені набори (справжній та прогнозований), ми можемо створити матрицю невідповідностей, що узагальнюватиме ці результати перевірки класифікатора:



В цій матриці невідповідностей система порахувала, що із 8 зображень котів 3 були псами, а для 2 з 5 зображень псів було зроблено прогноз, що це коти.

****

**https://gitlab.com/2019-2023/ipz19-3/lysovyi-maksym/AI**

**Висновок:** використовуючи спеціалізовані бібліотеки та мову програмування Python дослідив різні методи класифікації даних та навчився їх порівнювати.