

ЛАБОРАТОРНА РОБОТА №5

ДОСЛІДЖЕННЯ МЕТОДІВ НЕКОНТРОЛЬОВАНОГО НАВЧАННЯ

Мета заняття: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідити методи неконтрольованої класифікації даних у машинному навчанні.

Хід роботи

GitHub репозиторій: https://github.com/AlexanderHorielko/SAI_Horielko_PI-59

Завдання 2.1. Створення класифікаторів на основі випадкових та гранично випадкових лісів

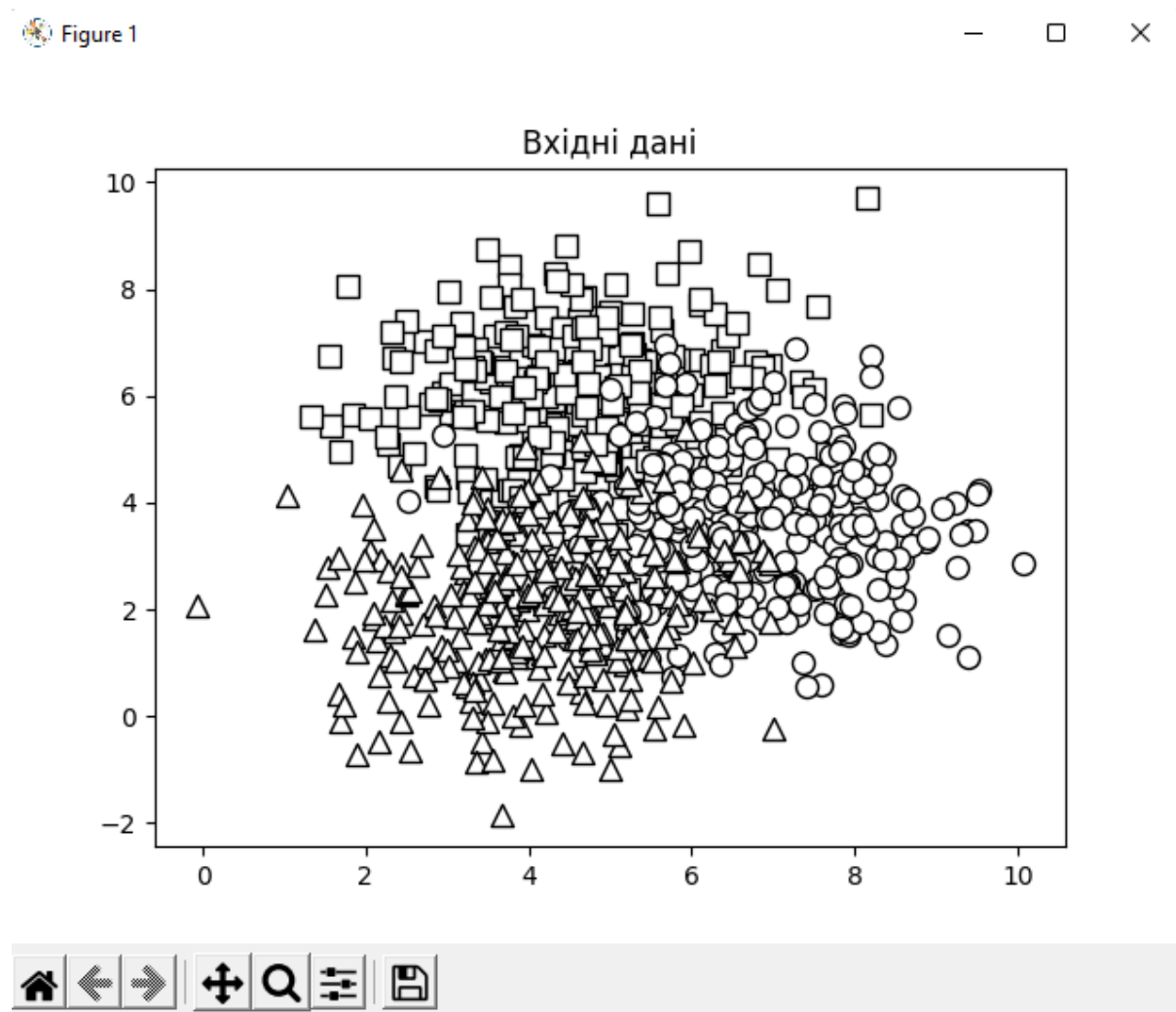


Рисунок 1. Графік розподілу вхідних даних

					ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр5						
Змн.	Арк.	№ докум.	Підпис	Дата							
Розроб.		Горелко О. В.			Звіт з лабораторної роботи			Лім.	Арк.	Аркушів	
Перевір.		Пулеко І. В.								1	3
Керівник								ФІКТ Гр. ПІ-59			
Н. контр.											
Зав. каф.											

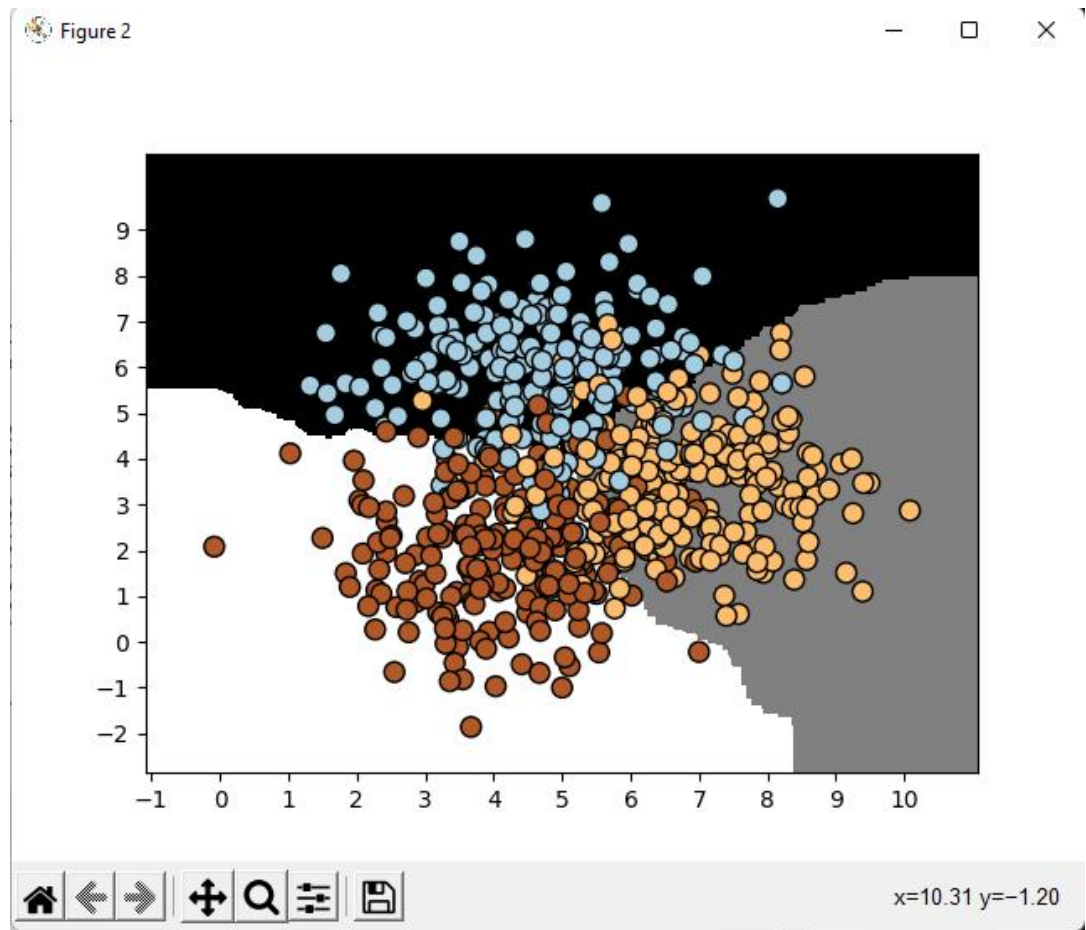


Рисунок 2. Графік результату навчання тренувального набору

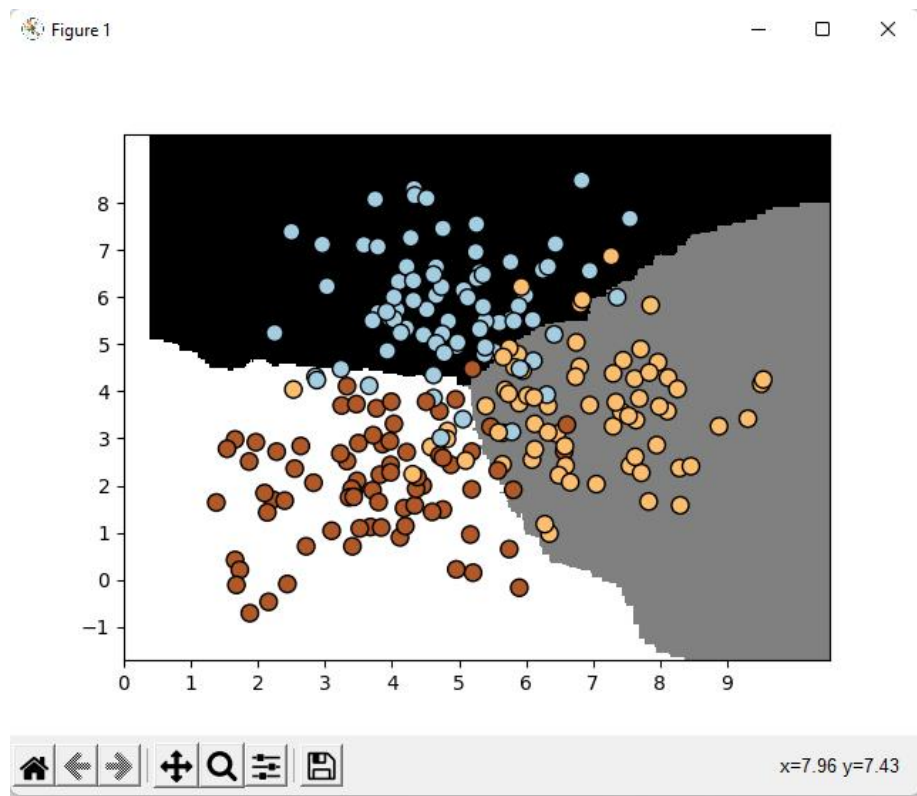


Рисунок 3. Графік результату навчання тестувального набору

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр5	Арк.
		Пулеко І. В.				2
Змн.	Арк.	№ докум.	Підпис	Дата		

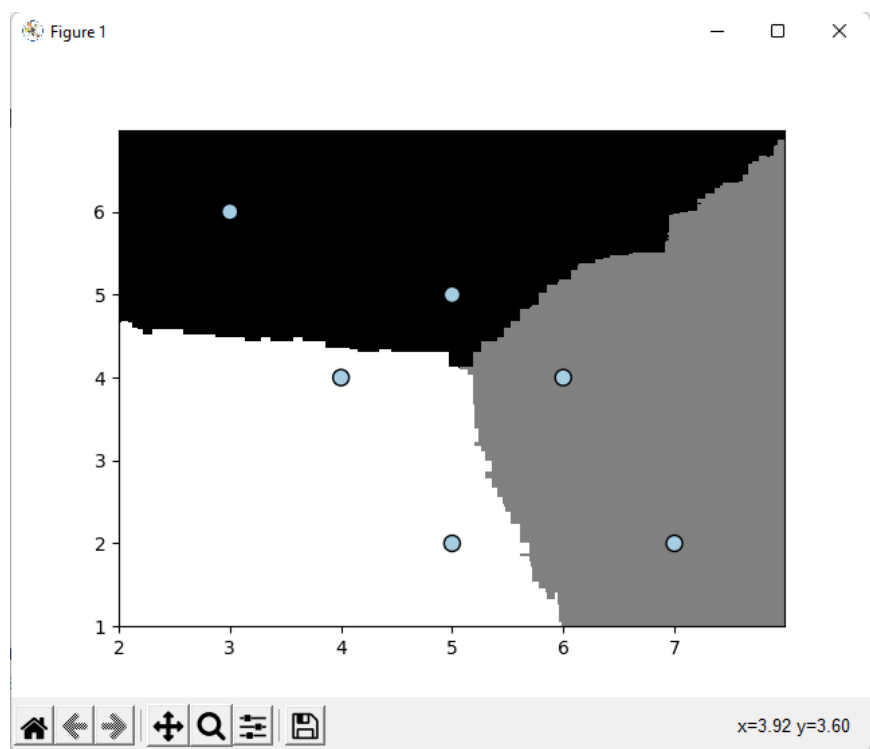


Рисунок 4. Графік результату навчання для тестових точок

```

lab5 task1.py
Project
lab5 D:\University\4(2)\AI\lab5
  Include
  Lib
  Scripts
Terminal: Local
Classifier performance on training dataset

precision    recall  f1-score   support

Class-0      0.89    0.83    0.86    221
Class-1      0.82    0.84    0.83    230
Class-2      0.83    0.86    0.85    224

accuracy          0.85    675
macro avg         0.85    0.85    0.85    675
weighted avg      0.85    0.85    0.85    675

=====

Classifier performance on test dataset

precision    recall  f1-score   support

Class-0      0.92    0.85    0.88     79
Class-1      0.84    0.84    0.84     70
Class-2      0.85    0.92    0.89     76

accuracy          0.87    225
macro avg         0.87    0.87    0.87    225
weighted avg      0.87    0.87    0.87    225

=====

Confidence measure:

Datapoint: [5 5]
Predicted class: Class-0

Datapoint: [3 6]
Predicted class: Class-0

Datapoint: [6 4]
Predicted class: Class-1

Datapoint: [7 2]
Predicted class: Class-1

Datapoint: [4 4]
Predicted class: Class-2

Datapoint: [5 2]
Predicted class: Class-2

(lab5) D:\University\4(2)\AI\lab5:

```

Рисунок 5. Результат класифікатора

В результаті виконання даного завдання було досліджено два види класифікаторів: випадкового лісу та гранично випадкового лісу, наочно представлено їх ефективність у вигляді графіків, результатів класифікації та метрик.

Завдання 2.2. Обробка дисбалансу класів

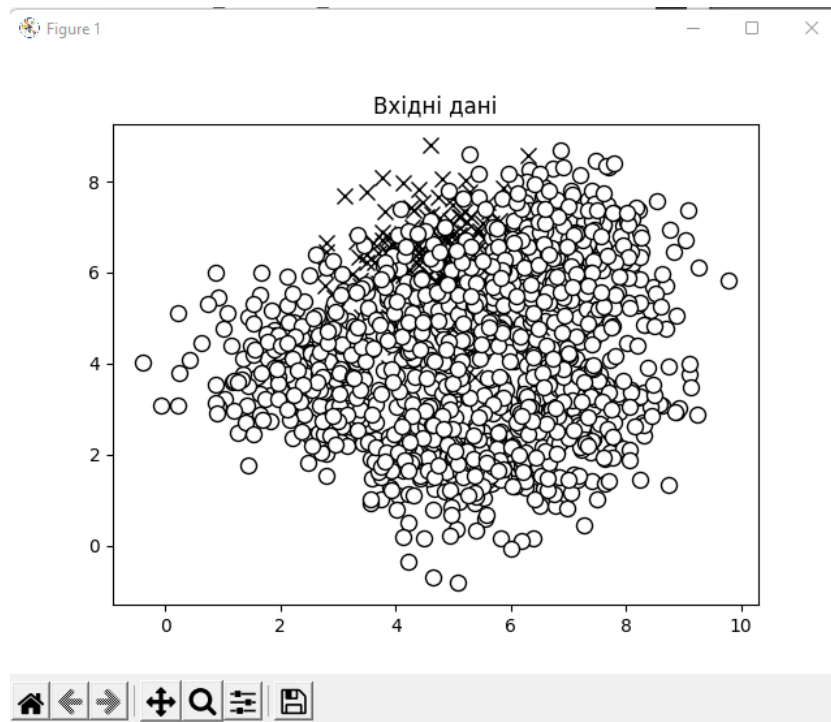


Рисунок 4. Графік розподілу вхідних даних

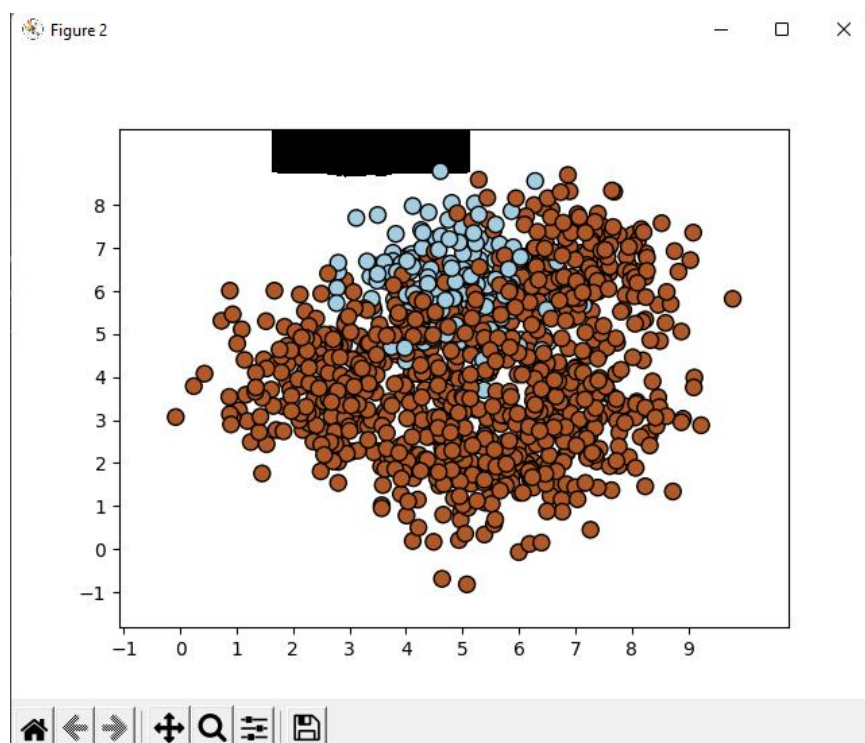


Рисунок 5. Графік даних класифікатора для тестового набору

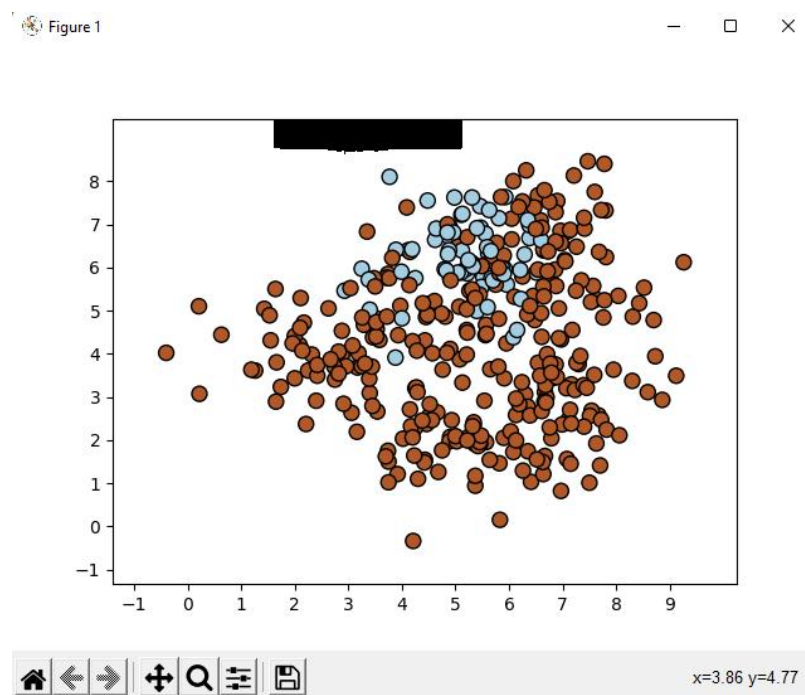


Рисунок 6. Графік даних класифікатора

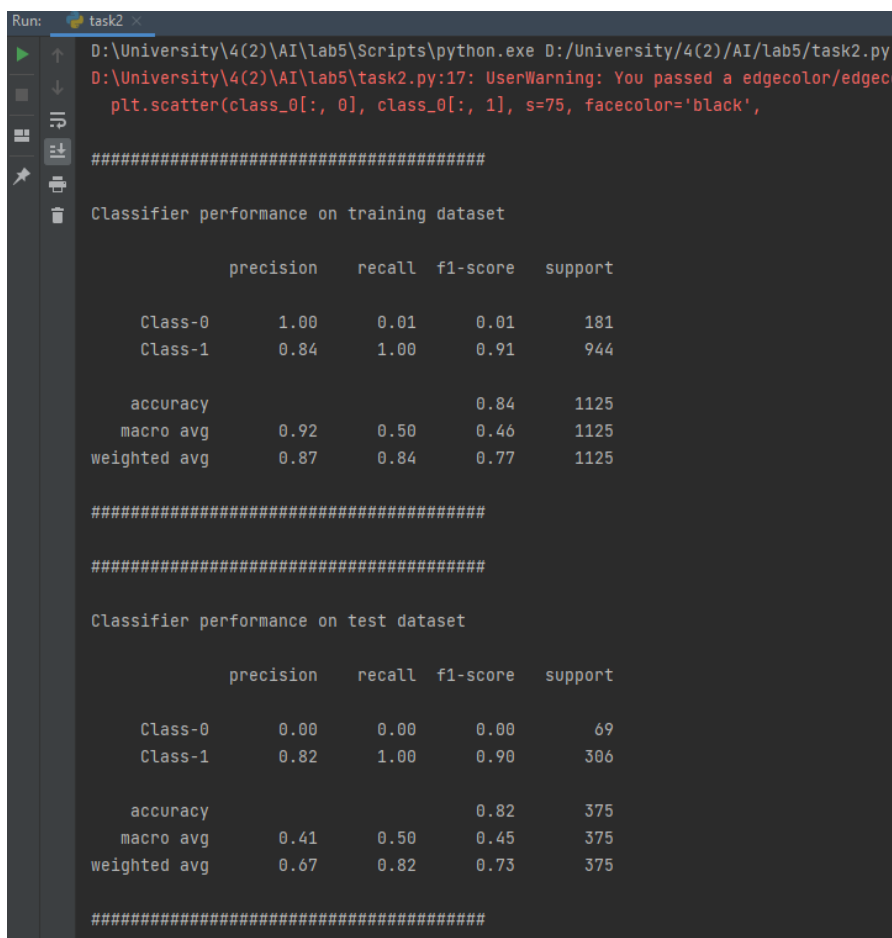


Рисунок 7. Результат даних класифікатора

В результаті виконання даного завдання було досліджено вплив враування дисбалансу в гранично випадковому лісі при використанні несбалансованих даних.

Завдання 2.3. Знаходження оптимальних навчальних параметрів за допомогою сіткового пошуку

```

28
29
30
31
32
33
34
35
36
print('\nGrid scores for the
for i in range(0, len(classi
print(classifier.cv_resu
print('\nBest parameters:',

y_pred = classifier.predict(X_te
print('\nPerformance report:\n')
print(classification_report(y_te

```

```

Run: task3
#### Searching for optimal parameters for precision_weighted

Grid scores for the parameter grid:
{'max_depth': 2, 'n_estimators': 100} --> 1
{'max_depth': 4, 'n_estimators': 100} --> 5
{'max_depth': 7, 'n_estimators': 100} --> 4
{'max_depth': 12, 'n_estimators': 100} --> 8
{'max_depth': 17, 'n_estimators': 100} --> 9
{'max_depth': 4, 'n_estimators': 25} --> 2
{'max_depth': 4, 'n_estimators': 50} --> 7
{'max_depth': 4, 'n_estimators': 100} --> 5
{'max_depth': 4, 'n_estimators': 250} --> 3

Best parameters: {'max_depth': 2, 'n_estimators': 100}

#### Searching for optimal parameters for recall_weighted

Grid scores for the parameter grid:
{'max_depth': 2, 'n_estimators': 100} --> 1
{'max_depth': 4, 'n_estimators': 100} --> 5
{'max_depth': 7, 'n_estimators': 100} --> 3
{'max_depth': 12, 'n_estimators': 100} --> 8
{'max_depth': 17, 'n_estimators': 100} --> 9
{'max_depth': 4, 'n_estimators': 25} --> 1
{'max_depth': 4, 'n_estimators': 50} --> 7
{'max_depth': 4, 'n_estimators': 100} --> 5
{'max_depth': 4, 'n_estimators': 250} --> 3

Best parameters: {'max_depth': 2, 'n_estimators': 100}

Performance report:

      precision    recall  f1-score   support

      0.0         0.94      0.81      0.87         79
      1.0         0.81      0.86      0.83         70
      2.0         0.83      0.91      0.87         76

   accuracy          0.86
  macro avg          0.86
 weighted avg          0.86

```

Рисунок 7. Результат пошуку оптимальних параметрів

Під час виконання даного завдання ми дослідили процес оптимізації параметрів класифікатора у відповідності до певної метрики.

Завдання 2.4. Обчислення відносної важливості ознак

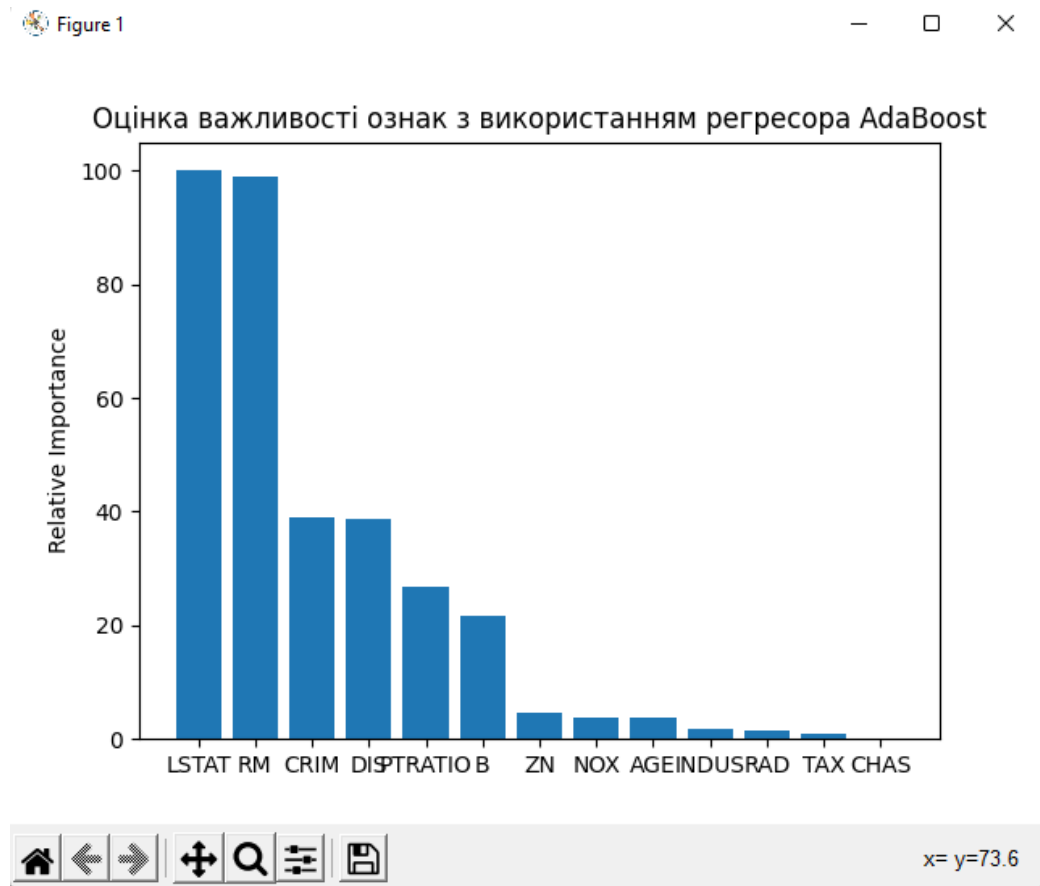


Рисунок 8. Діаграма оцінки важливості ознак

Відповідно до отриманої діаграми, найбільш важливими ознаками є LSTAT (відсоток малозабезпеченого населення) та RM (середня кількість кімнат), а знехтувати можна CHAS (чи межує з річкою).

В результаті виконання даного завдання ми навчилися аналізувати важливість характеристик датасету за допомогою регресора AdaBoost.

Завдання 2.5. Прогнозування інтенсивності дорожнього руху за допомогою класифікатора на основі гранично випадкових лісів

```
Run: task5
D:\University\4(2)\AI\lab5\Scripts\python.exe D:/University/4(2)/AI/lab5/task5.py
Mean absolute error = 5.57
Predicted traffic: 24
Process finished with exit code 0
```

Рисунок 9. Результат

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр5	Арк.
		Пулеко І. В.				7
Змн.	Арк.	№ докум.	Підпис	Дата		

Висновок: використовуючи спеціалізовані бібліотеки та мову програмування Python дослідив методи неконтрольованої класифікації даних у машинному навчанні.

		Горелко О. В.			ЖИТОМИРСЬКА ПОЛІТЕХНІКА.21.121.05.000 – Лр5	Арк.
		Пулеко І. В.				8
Змн.	Арк.	№ докум.	Підпис	Дата		