**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

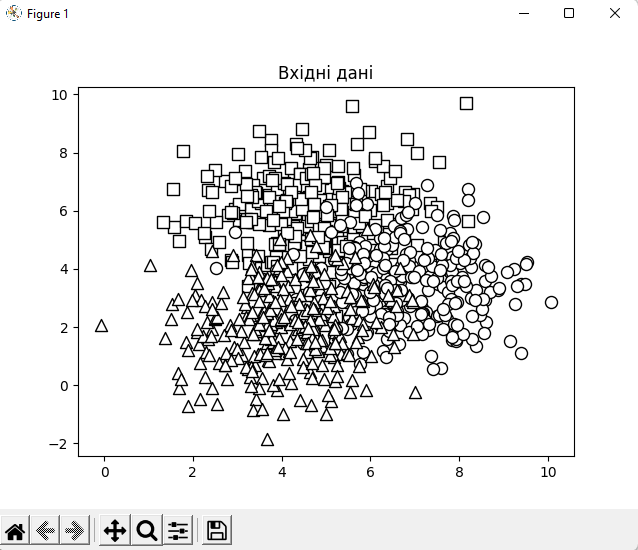


Рисунок 1. Графік розподілу вхідних даних

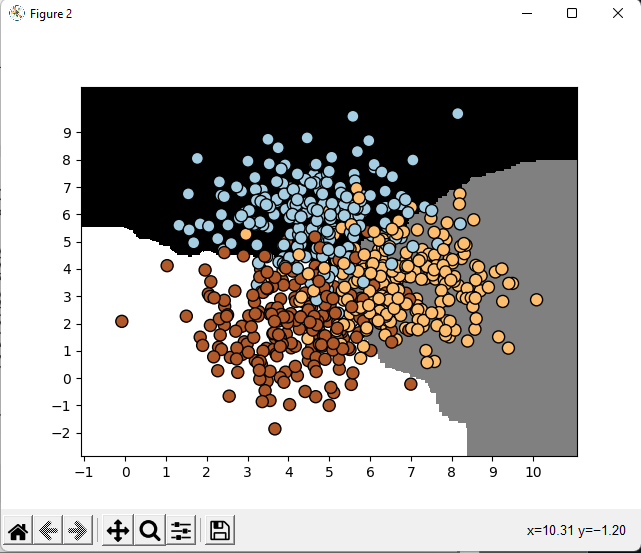


Рисунок 2. Графік результату навчання тренувального набору

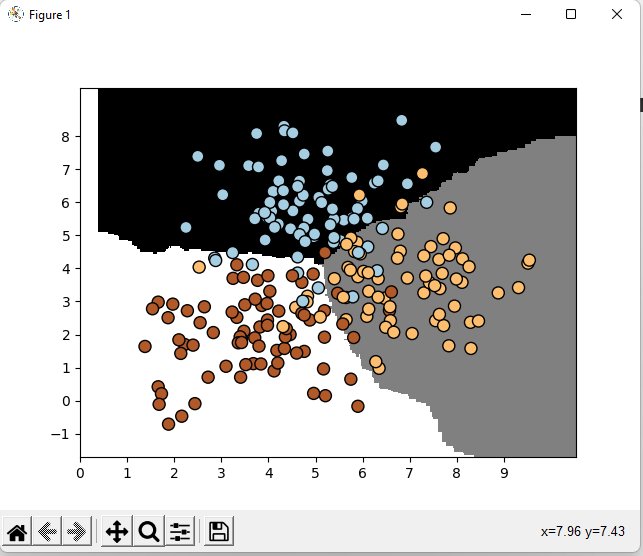


Рисунок 3. Графік результату навчання тестувального набору

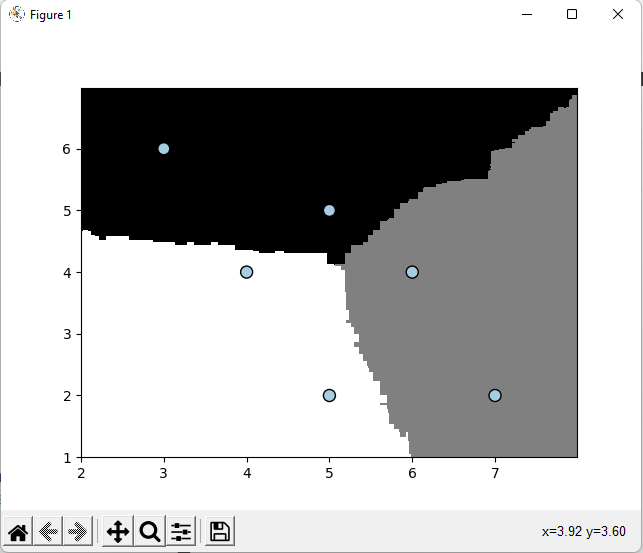


Рисунок 4. Графік результату навчання для тестових точок

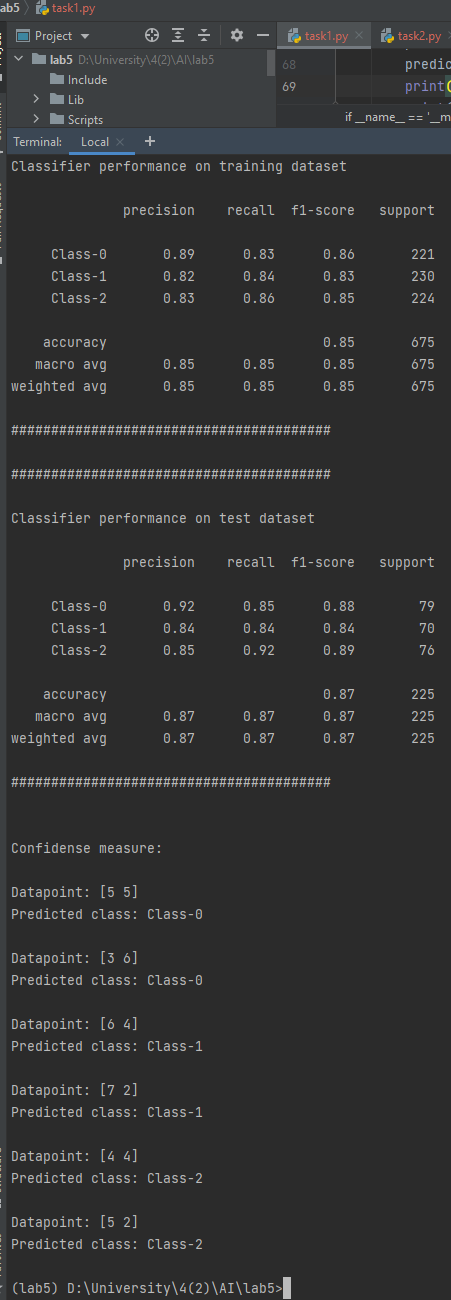


Рисунок 5. Результат класифікатора

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

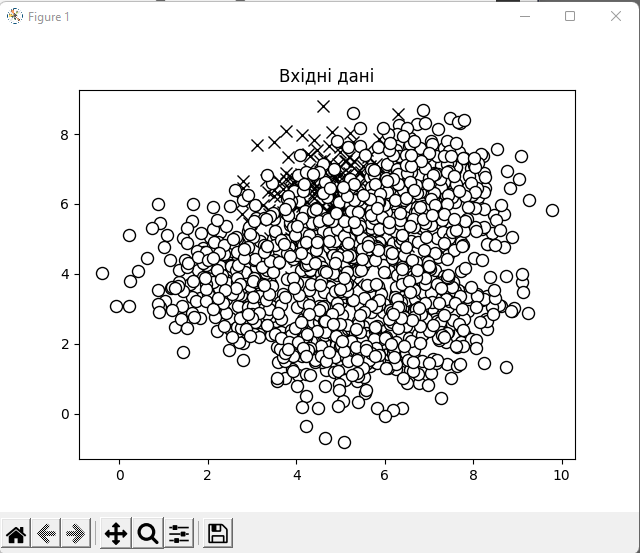


Рисунок 4. Графік розподілу вхідних даних

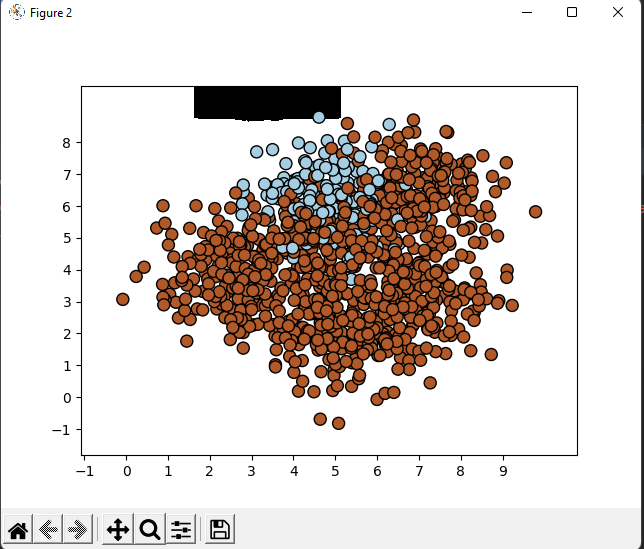


Рисунок 5. Графік даних класифікатора для тестового набору

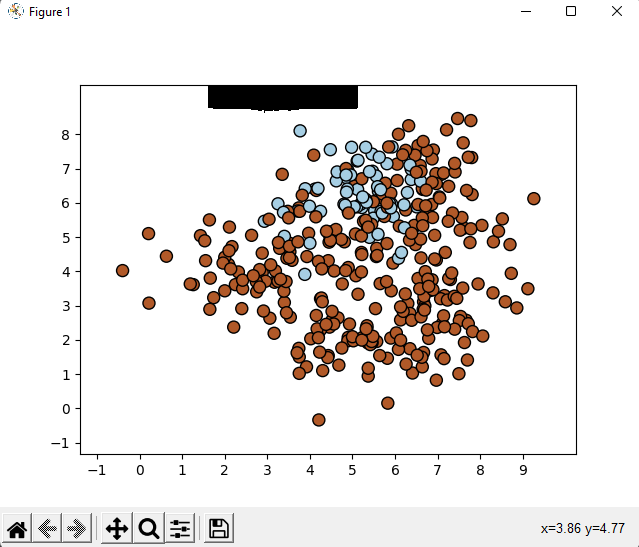


Рисунок 6. Графік даних класифікатора

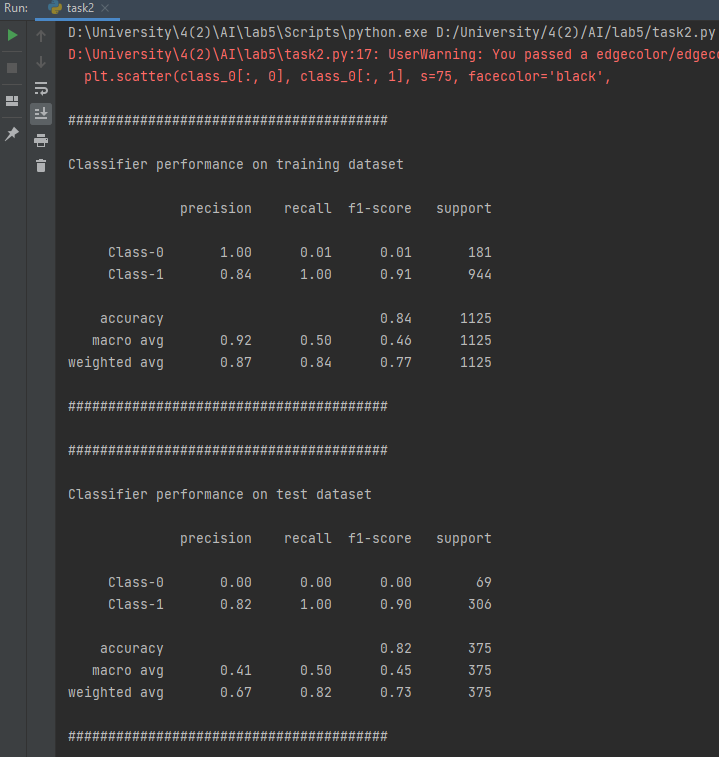


Рисунок 7. Результат даних класифікатора

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

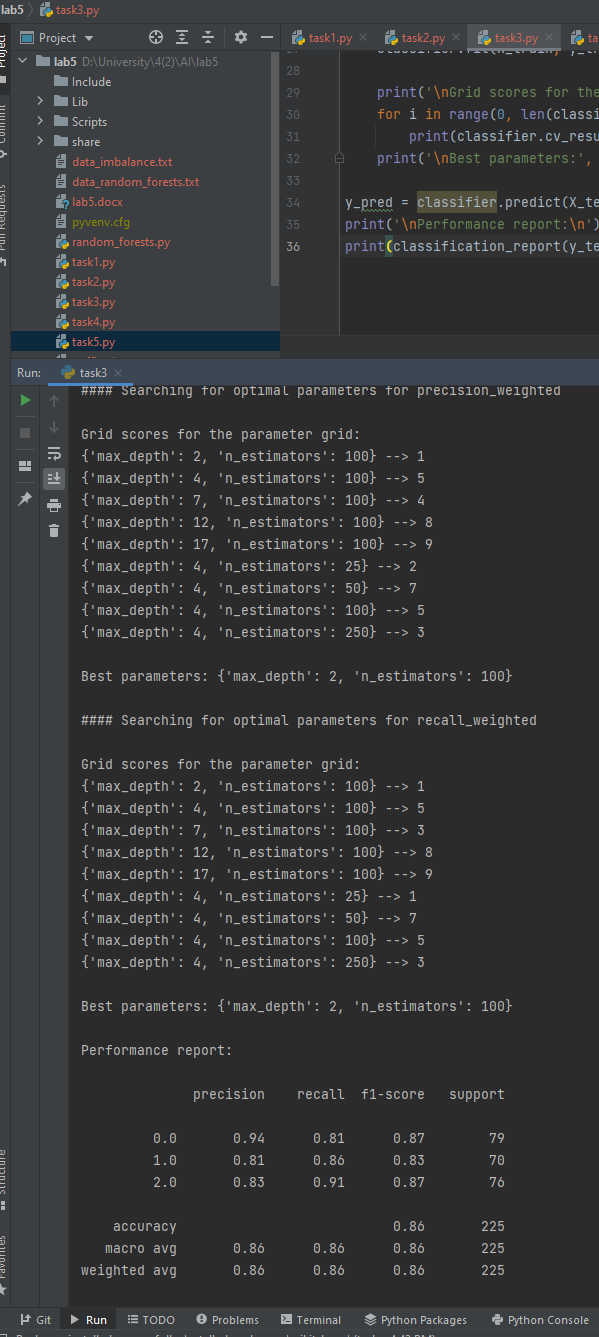


Рисунок 7. Результат пошуку оптимальних параметрів

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

**Завдання 2.4.** Обчислення відносної важливості ознак

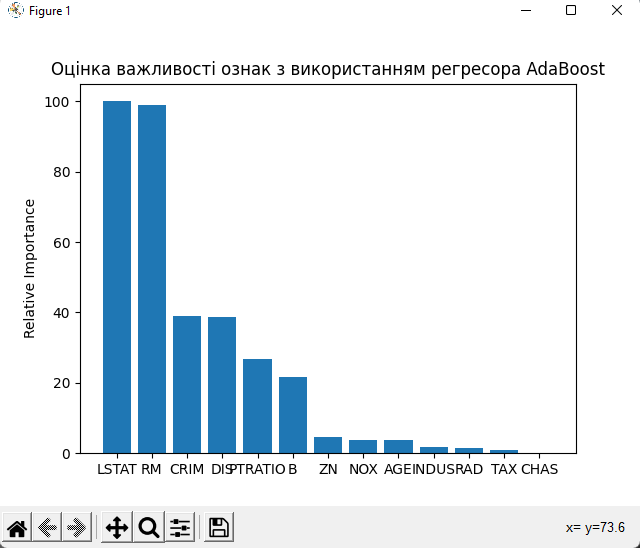


Рисунок 8. Діаграма оцінки важливості ознак

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

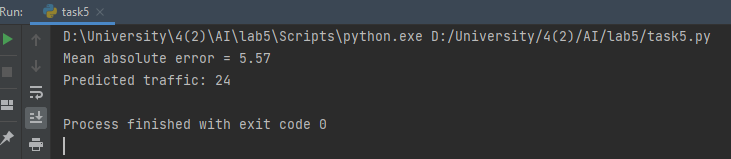


Рисунок 9. Результат

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