**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ**

**«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ**

**ІМ. ІГОРЯ СІКОРСЬКОГО»**

**Кафедра АСОІУ**

Звіт з лабораторної роботи

за темою: «**НЕІЄРАРХІЧНІ МЕТОДИ КЛАСТЕРНОГО АНАЛІЗУ**»

Виконав студент

групи ІП-91

Кочев Геннадій

Мета роботи: Практичне засвоєння неієрархічного кластерного аналізу багатовимірних даних на прикладі методу k-середніх.

Теоретичні відомості: Неієрархічні методи мають за основу вже задану кількість кластерів (k-means, PAM кластеризація) або використовують складні алгоритми знаходження їх кількості (CLOPE, карти Кохонена).

Серед неієрархічних методів кластеризації особливої уваги заслуговують ітеративні методи. Вони працюють за наступним алгоритмом:

1)вихідні дані розбиваються на певну кількість кластерів та обчислюються центри тяжіння цих кластерів;

2)кожна точка даних поміщується в кластер з найближчим центром

тяжіння;

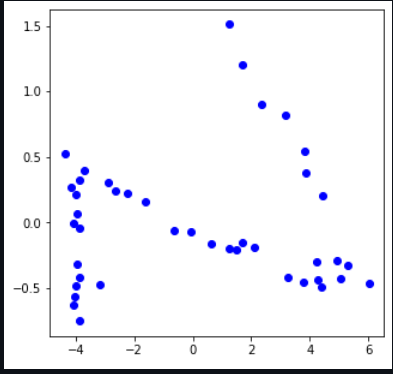
3)обчислюються нові центри тяжіння кластерів; кластери не замінюються на нові доти, поки не будуть повністю переглянуті всі дані;

4)кроки 2 і 3 повторюються доти, поки не перестануть змінюватись кластери.

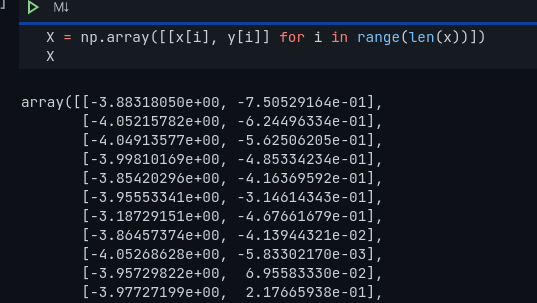
Набір даних: представляє собою док-файл, що містить таблицю з даними та назвами колонок.

Виконання: робота була виконана за допомогою мови програмування Python3 у середовищі Jupyter Notebook.

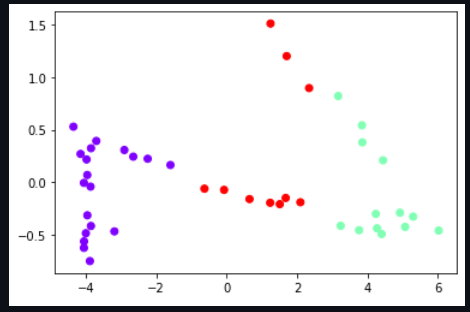
Результати роботи: маємо такий розподіл даних за графіком:



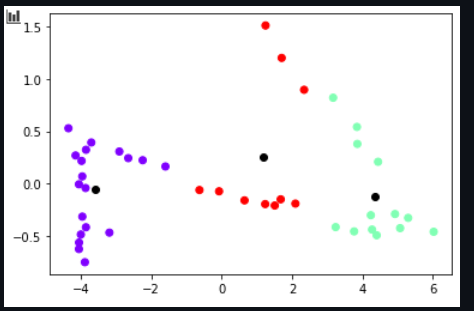
Переформатуємо координати точок до зручному формату:



Розфарбуємо точки за розподілом на 3 кластери:



А тепер з центрами:



Висновок: під час виконання лабораторної роботи мені потрібно було знайти набір даних, переформатувати його до зручного для роботи виду та провести неієрархічну кластеризацію за методом k-середніх. Для переформування даних docx, pandas, numpy, розбиття даних на координати - sklearn (decomposition, preprocessing), графіки - matplotlib.

Програма складається з таких етапів:

* імпортування модулів
* перетворення даних до матриці
* встановлення параметрів для кластеризації
* знахождення центрів кластерів
* виведення точок та центрів різними кольорами на графіках

Таким чином, мені вдалося вивчити методи підходу до неієрархічної кластеризації для обраного набору даних. Робота виконана.