

ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ГЕОДАНИХ

Виконав : студент групи ДА-22
Магас В.В.
Керівник: ст.в., к.т.н.
Булах Б. В.

Мета роботи - застосування засобів інтелектуального аналізу даних для виявлення прихованих закономірностей та зв'язків у просторових баз даних онкохворих людей.

Предмет дослідження - кластерний аналіз просторових баз даних .

Актуальність. У зв'язку бурхливим накопиченням медичних даних, особливо гостро постає проблема застосування аналізу даних для їх опрацювання та подальшого прогнозування, на їх основі.

Основні завдання

- З'ясування специфічних для сфери ІАПД вимог.
- Дослідження існуючих рішень і підходів.
- Вивчення підходів до аналізу просторових даних з використанням алгоритмів кластеризації.
- Постановка прикладної задачі, яка б надала можливість для детального дослідження.
- Безпосереднє використання засобів дослідження до поставленої задачі.
- Аналіз результатів, оцінка використаних та досліджених підходів.

Інтелектуальний аналіз просторових даних(ІАПД) - це процес виявлення раніше невідомих але потенційно корисних та цікавих закономірностей з великих наборів просторових даних.

Методи дослідження в ІАПД

- Кластеризація та метод визначення відхилення.
- Метод асоціації та колокації.
- Класифікація.
- Аналіз тенденцій

Обрані методи та алгоритми дослідження

Кластеризація (або кластерний аналіз) - це задача розбиття множини об'єктів на групи, які називаються кластерами. У середині кожної групи повинні виявитися «схожі» об'єкти, а об'єкти різних групи повинні бути якомога більш відмінні.

- K-means
- BIRCH (Balanced Iterative Reducing and Clustering using Hierarchies)
- DBSCAN (Density-based spatial clustering of applications with noise)

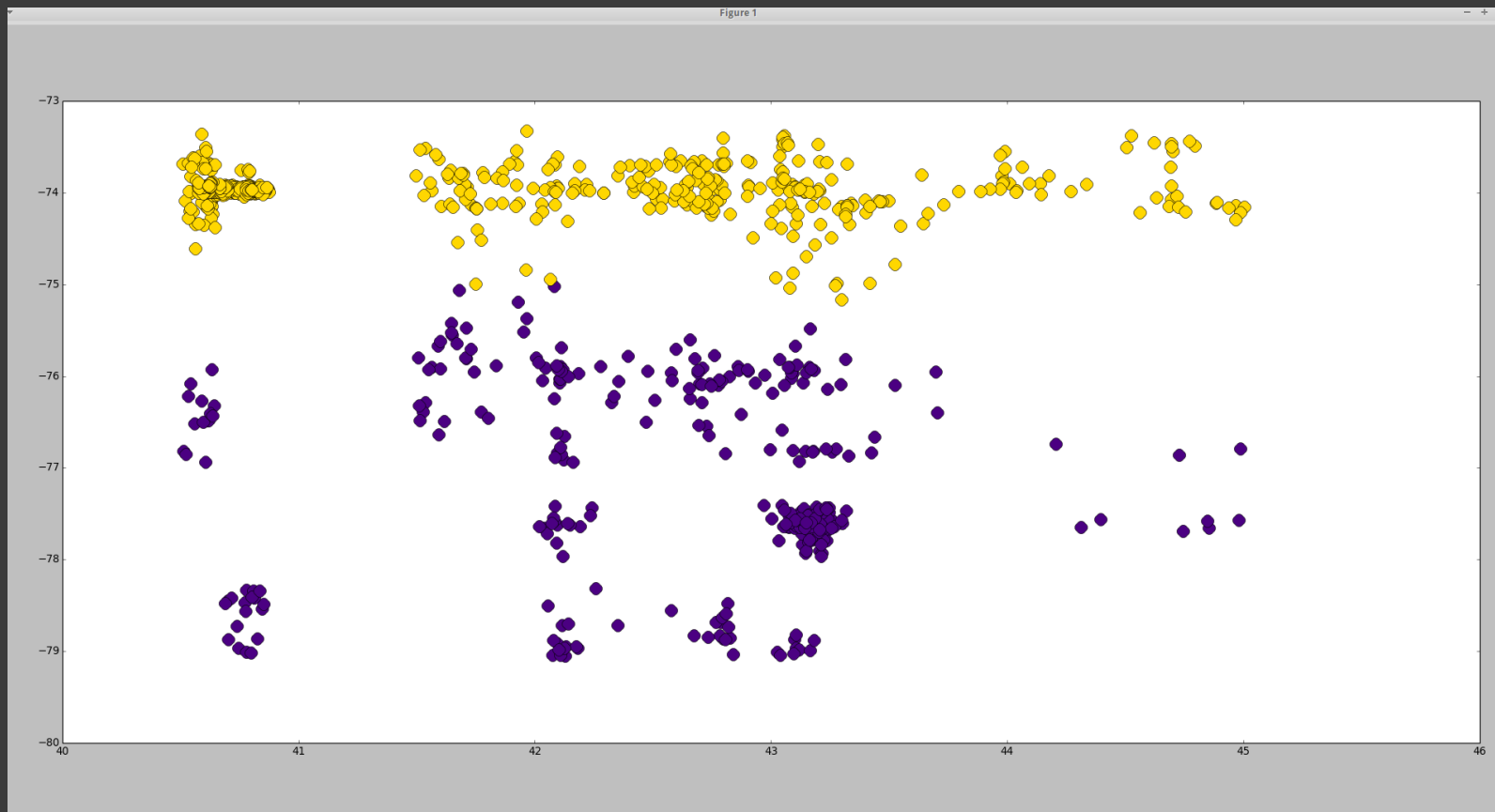
База даних та інструментарій

Просторова База даних : SEER Database

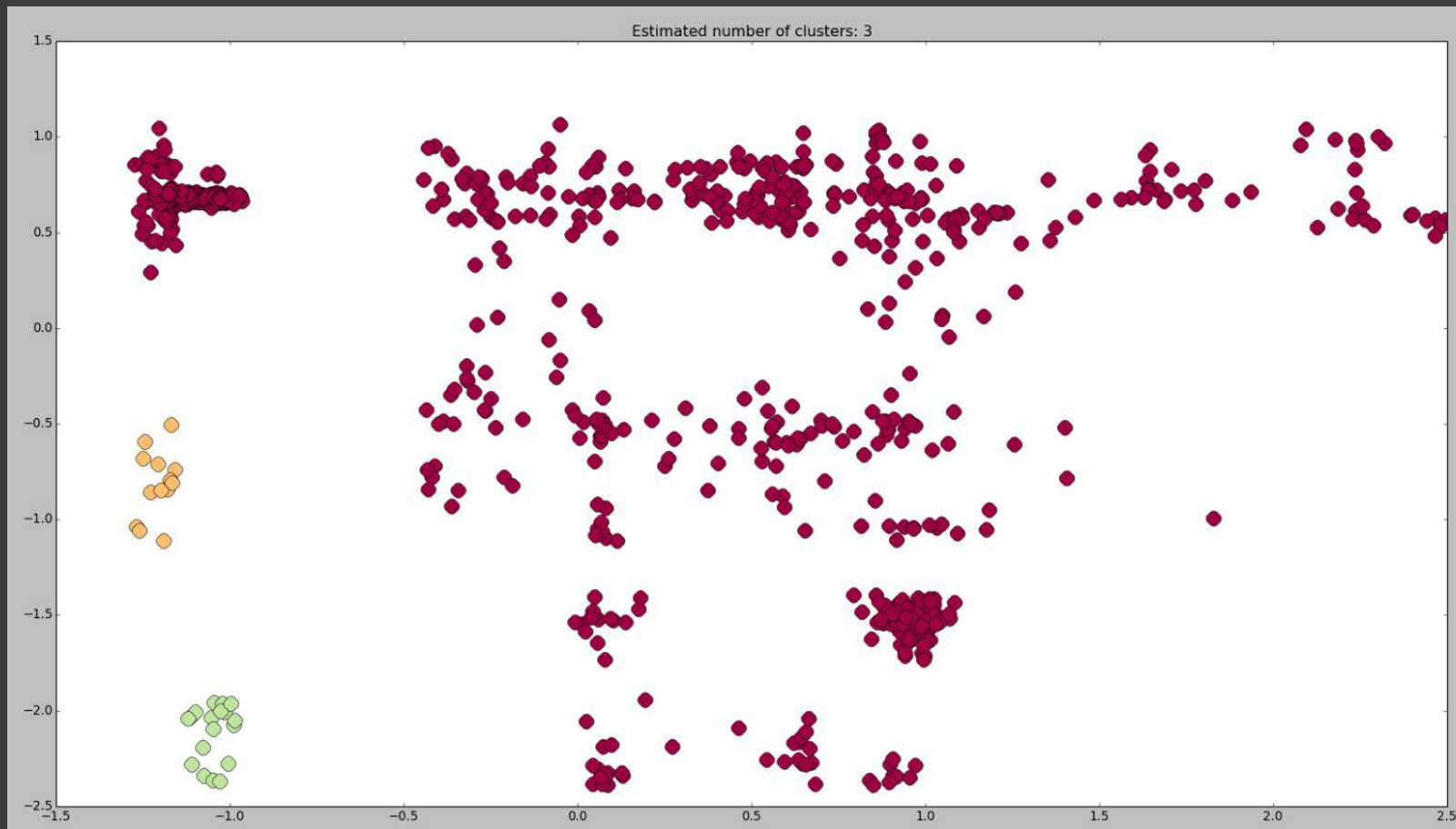
Використаний інструментарій:

- Orange
- Python
- SciPy
- Pandas
- Matplotlib
- Numpy

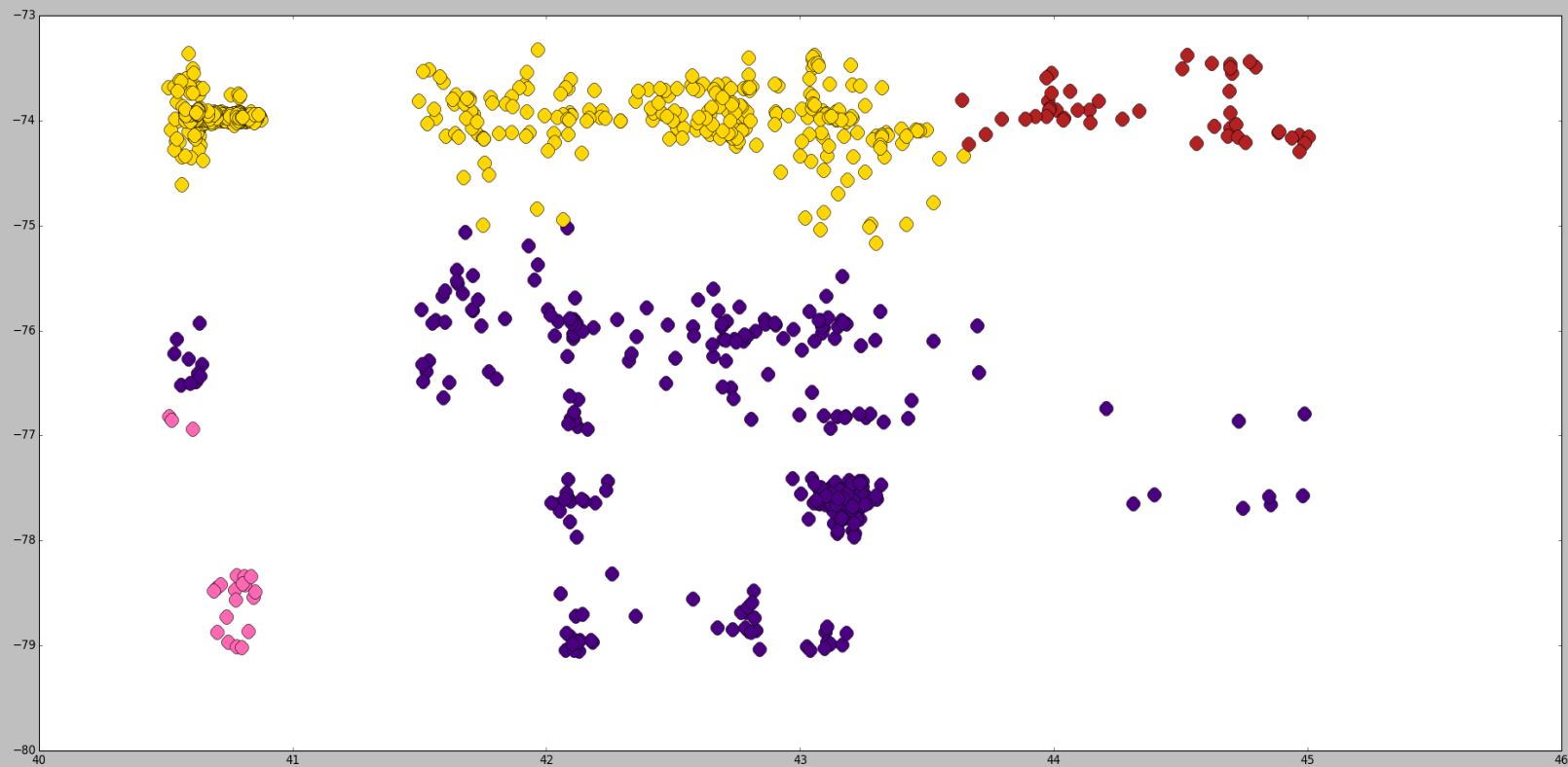
Результати кластеризації алгоритмом k-means



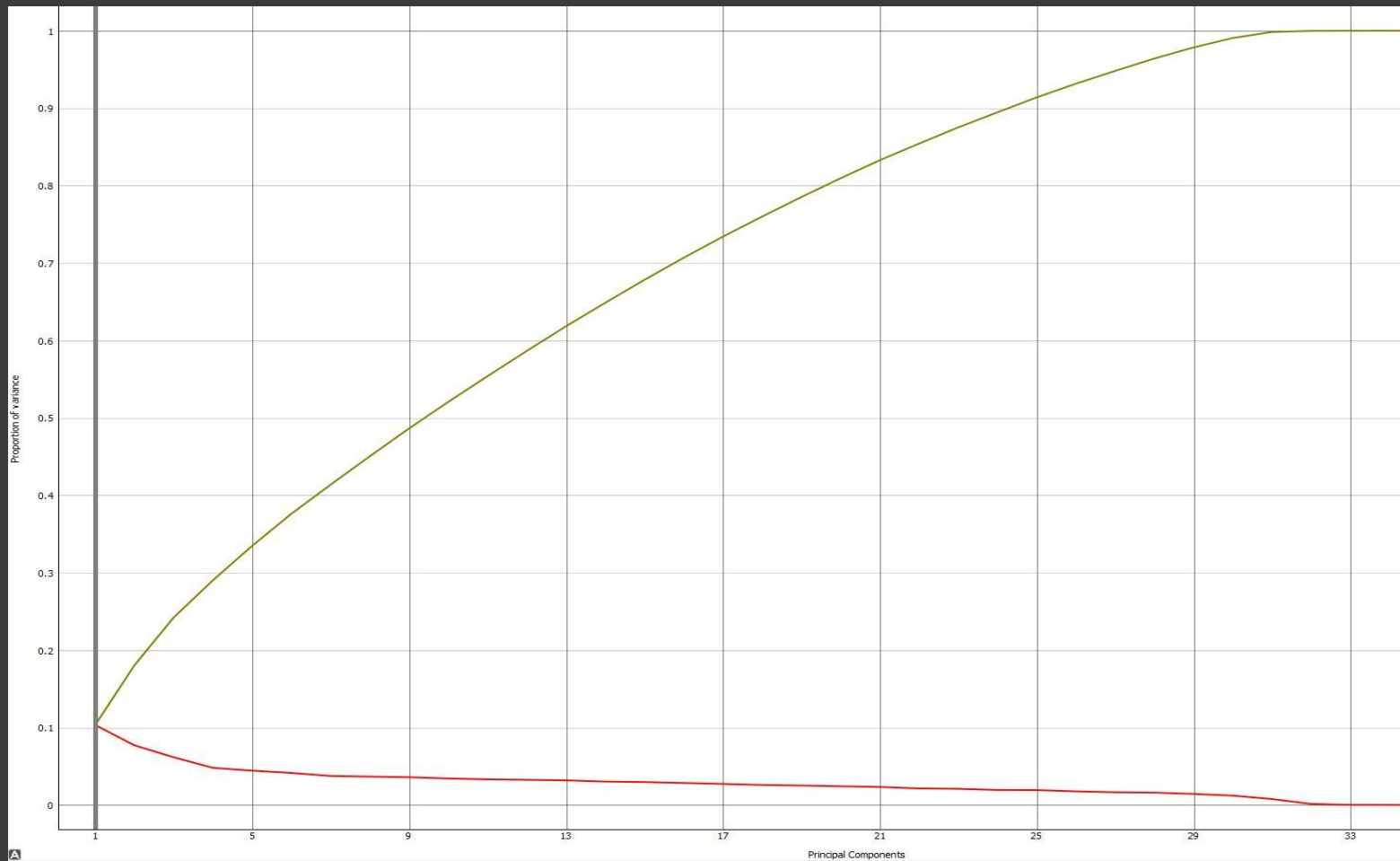
Результати кластеризації алгоритмом DBSCAN



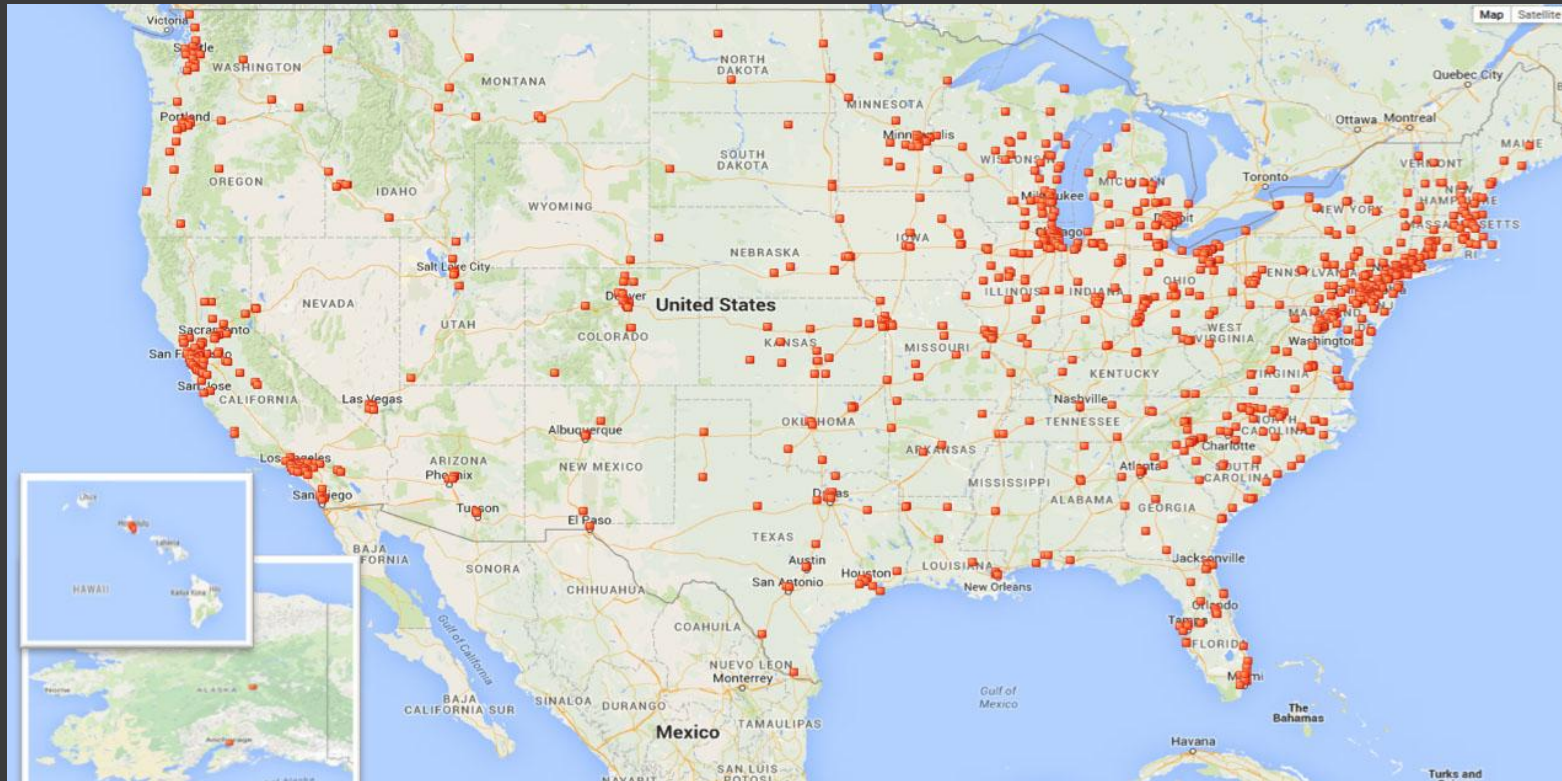
Результати кластеризації алгоритмом BIRCH



Графік лінійного перетворення МГК



Закономірності поширення хворих по території США



Результати порівняння ефективності алгоритмів

Характеристика	Алгоритм		
	K-means	DBSCAN	BIRCH
Однорідність	0.917	0.953	0.960
Повнота	0.766	0.883	0.812
V-міра	0.881	0.917	0.855
Скоригований індекс	0.966	0.952	0.914
Скоригована взаємна інформація	0.855	0.883	0.896
Коефіцієнт силуету	0.722	0.626	0.691

Висновки

- З'ясовано специфічні для ІАПД вимоги.
- Вивчено підходи до аналізу просторових даних з використанням алгоритмів кластеризації.
- Реалізовано алгоритми кластеризації для аналізу просторової БД онкохворих людей.
- Результати роботи представлено у вигляді діаграм, графіків, таблиць.

ДЯКУЮ ЗА УВАГУ!