

## ЛАБОРАТОРНА РОБОТА № 7

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

**Завдання 1.** Кластеризація даних за допомогою методу k-середніх

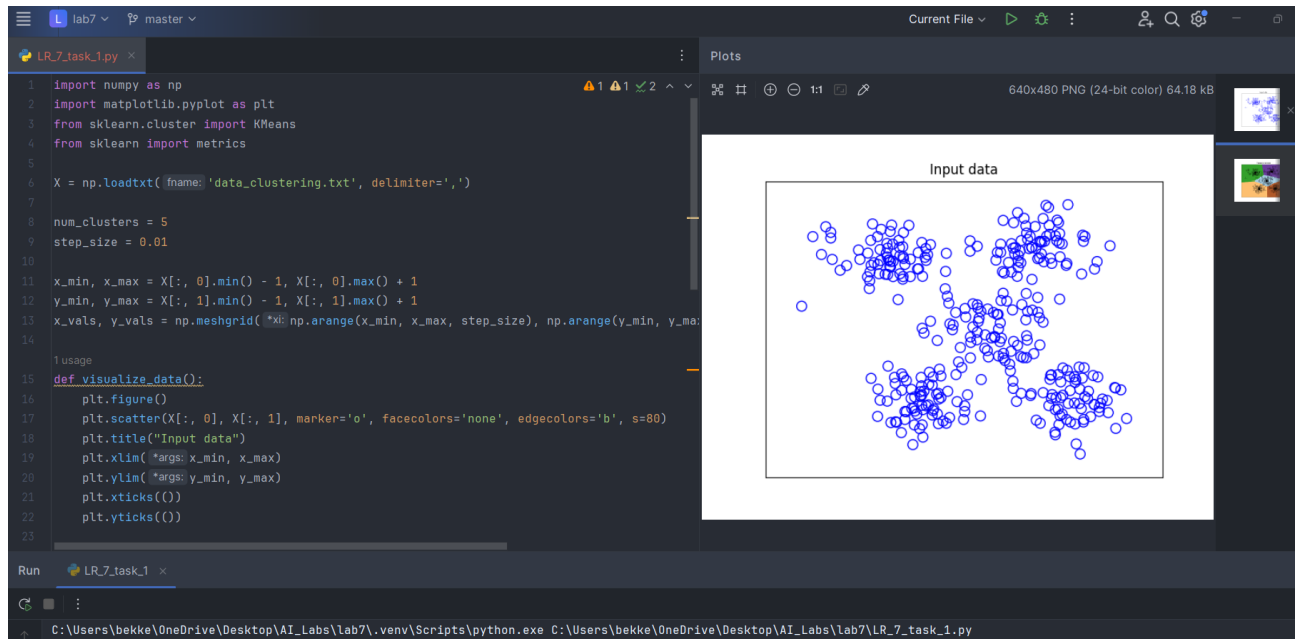


Рис 1. Візуалізація даних

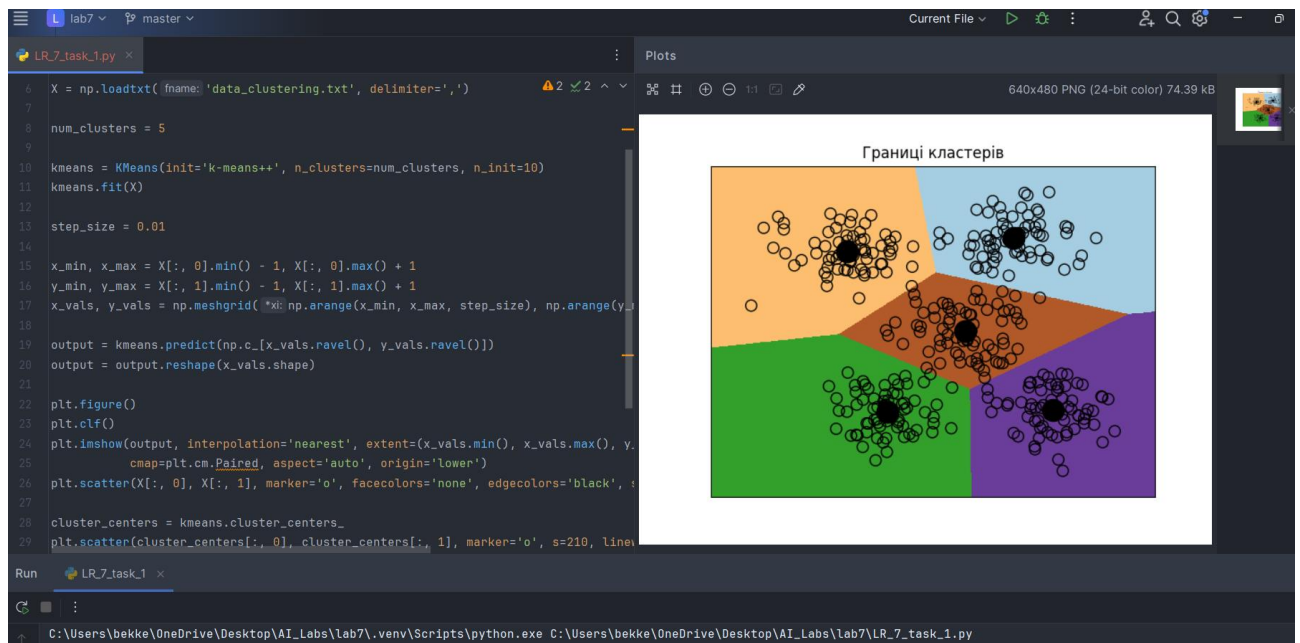


Рис 2. Візуалізація груп даних

**Висновок:** на даному етапі вдалося візуалізувати вхідні дані, а також кластери та їх межі. За допомогою цієї візуалізації можна проаналізувати розподіл даних. Також на графіку можна побачити положення центроїдів (центрів тяжіння кластеру, або головні точки).

## Завдання 2. Кластеризація К-середніх для набору даних Iris

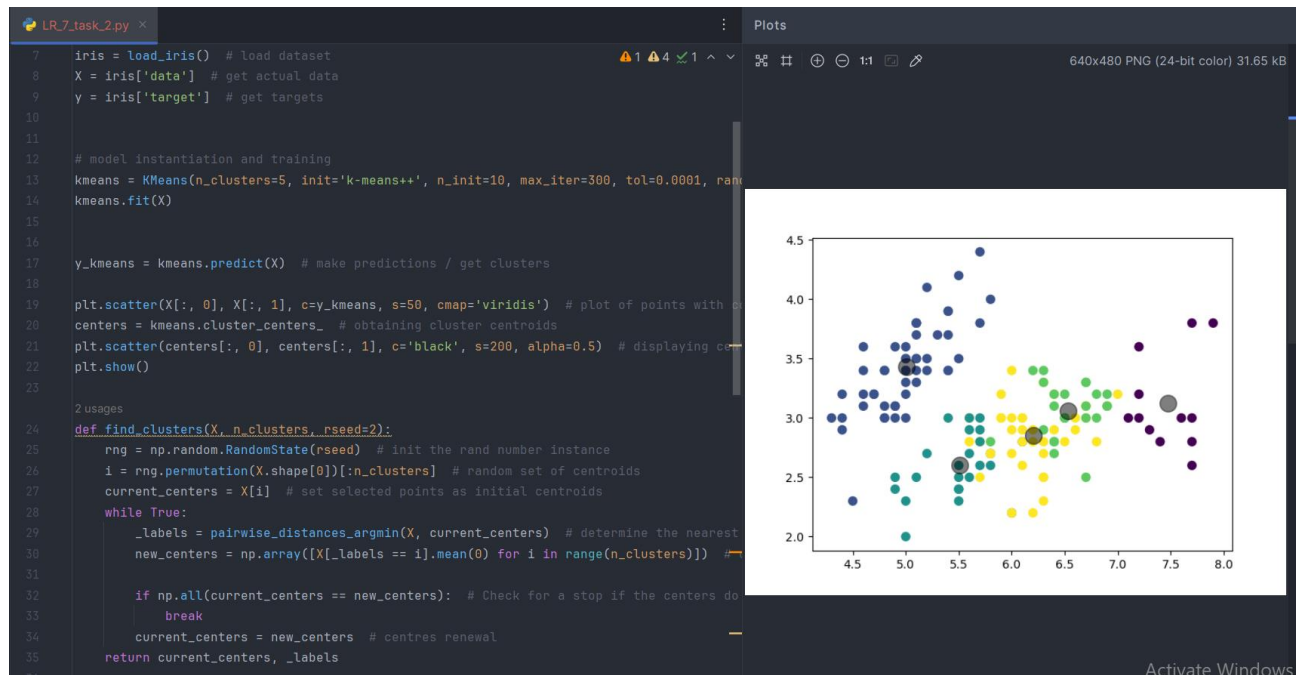


Рис 3. Код програми та візуалізація даних з центроїдами

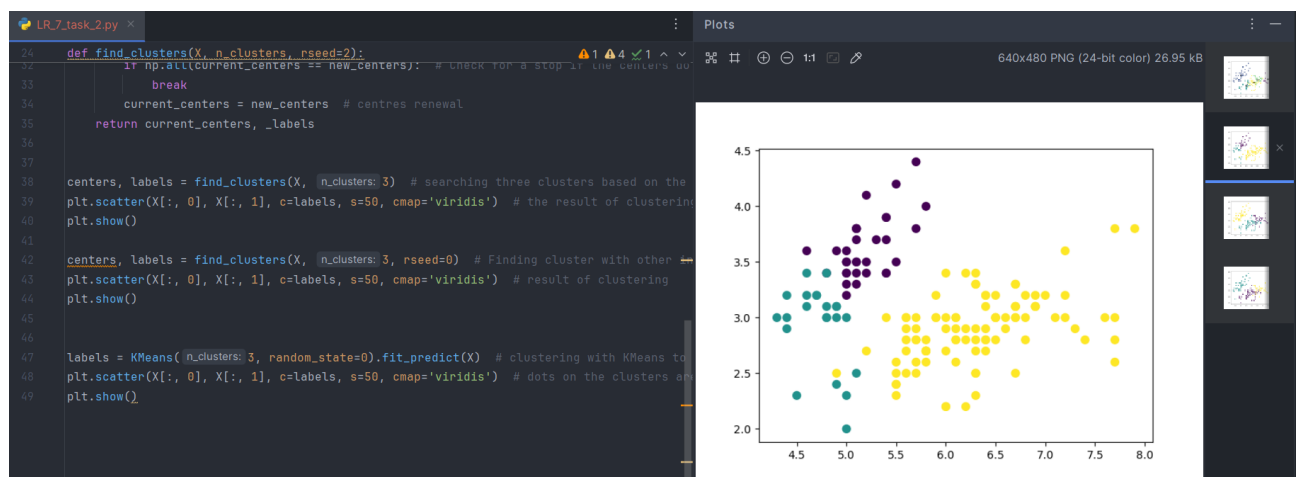


Рис 3. Код програми та візуалізація даних

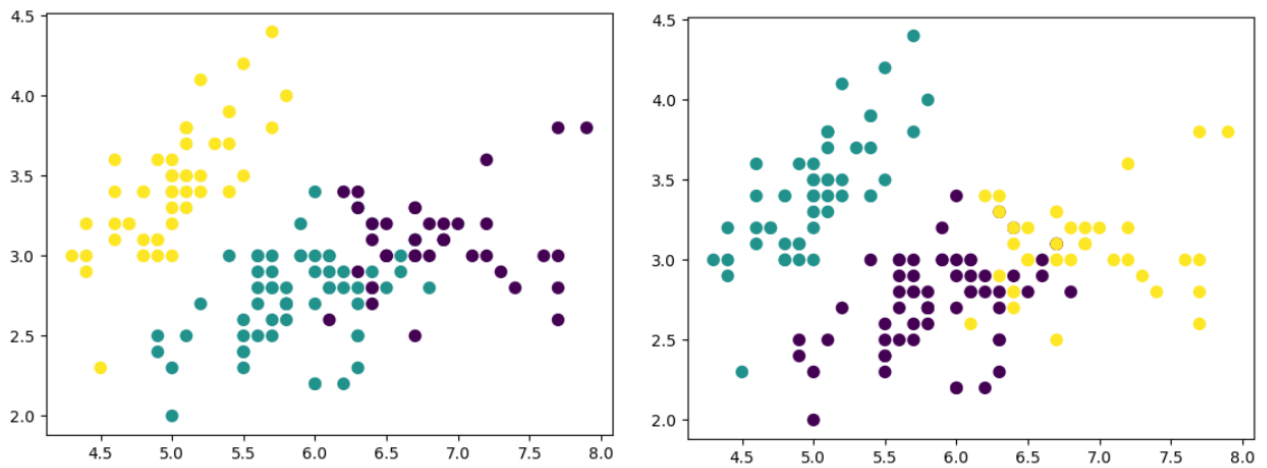


Рис 4. Візуалізація даних з початковими центрами та без

**Висновок:** на даному етапі було візуалізовано стандартну K-means кластерізацію, а також іншу, що використовує метод кластерізації з початковими центрами.

**Завдання 3.** Оцінка кількості кластерів з використанням методу зсуву середнього

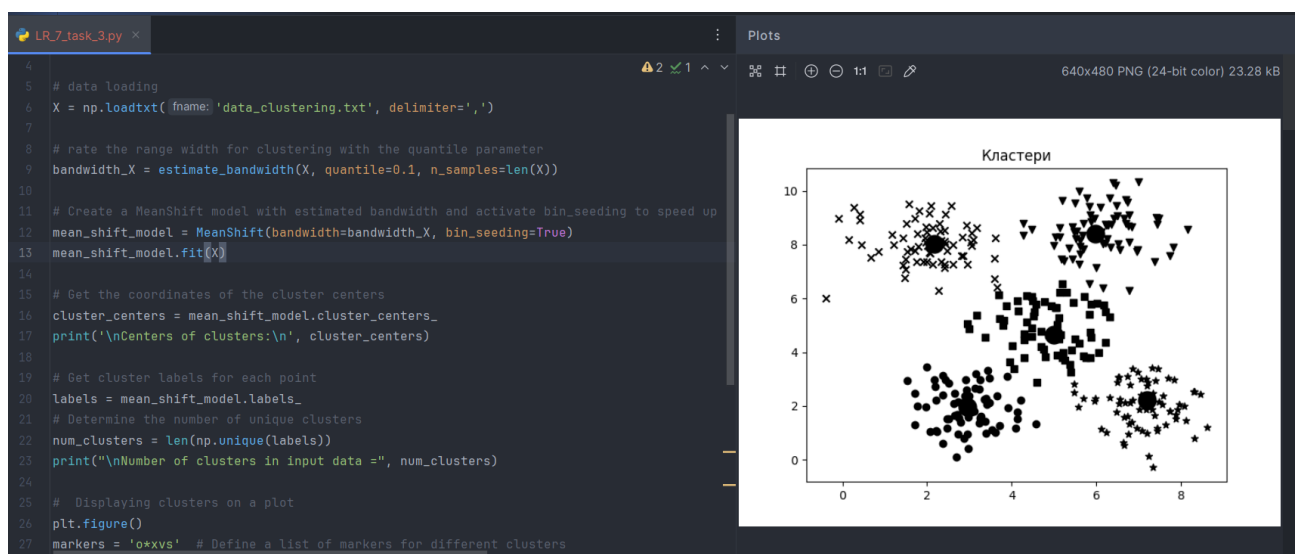
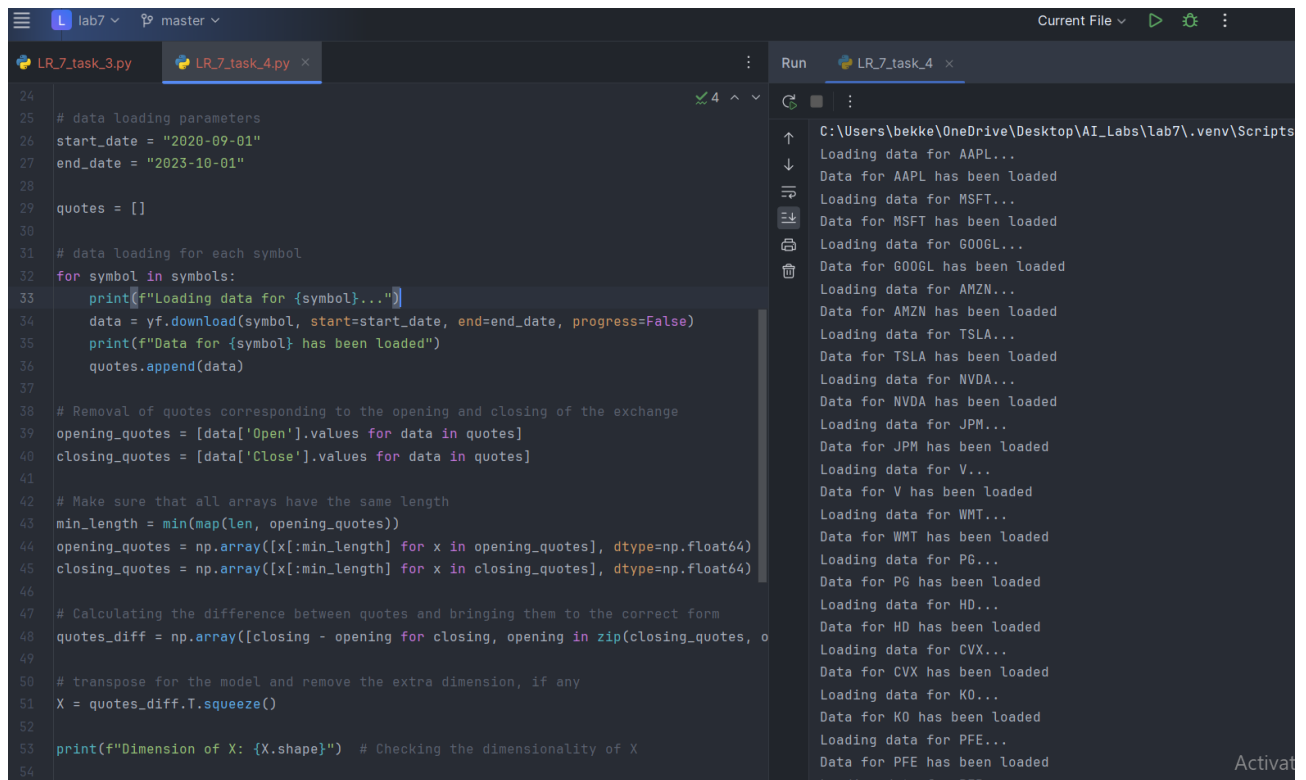


Рис 5. Візуалізація кластеризації

**Висновок:** на даному етапі вдалося виконати кластеризацію даних із використанням алгоритму MeanShift, який автоматично визначає кількість кластерів у даних та знаходить їх центри, що можна побачити на графіку

**Завдання 4.** Знаходження підгруп на фондовому ринку з використанням моделі поширення подібності



```
24
25 # data loading parameters
26 start_date = "2020-09-01"
27 end_date = "2023-10-01"
28
29 quotes = []
30
31 # data loading for each symbol
32 for symbol in symbols:
33     print(f"Loading data for {symbol}...")
34     data = yf.download(symbol, start=start_date, end=end_date, progress=False)
35     print(f"Data for {symbol} has been loaded")
36     quotes.append(data)
37
38 # Removal of quotes corresponding to the opening and closing of the exchange
39 opening_quotes = [data['Open'].values for data in quotes]
40 closing_quotes = [data['Close'].values for data in quotes]
41
42 # Make sure that all arrays have the same length
43 min_length = min(map(len, opening_quotes))
44 opening_quotes = np.array([x[:min_length] for x in opening_quotes], dtype=np.float64)
45 closing_quotes = np.array([x[:min_length] for x in closing_quotes], dtype=np.float64)
46
47 # Calculating the difference between quotes and bringing them to the correct form
48 quotes_diff = np.array([closing - opening for closing, opening in zip(closing_quotes, opening_quotes)])
49
50 # transpose for the model and remove the extra dimension, if any
51 X = quotes_diff.T.squeeze()
52
53 print(f"Dimension of X: {X.shape}") # Checking the dimensionality of X
54
```

Run LR\_7\_task\_4

C:\Users\bekke\OneDrive\Desktop\AI\_Labs\lab7\.venv\Scripts

Loading data for AAPL...  
Data for AAPL has been loaded  
Loading data for MSFT...  
Data for MSFT has been loaded  
Loading data for GOOGL...  
Data for GOOGL has been loaded  
Loading data for AMZN...  
Data for AMZN has been loaded  
Loading data for TSLA...  
Data for TSLA has been loaded  
Loading data for NVDA...  
Data for NVDA has been loaded  
Loading data for JPM...  
Data for JPM has been loaded  
Loading data for V...  
Data for V has been loaded  
Loading data for WMT...  
Data for WMT has been loaded  
Loading data for PG...  
Data for PG has been loaded  
Loading data for HD...  
Data for HD has been loaded  
Loading data for CVX...  
Data for CVX has been loaded  
Loading data for KO...  
Data for KO has been loaded  
Loading data for PFE...  
Data for PFE has been loaded

Рис 6. Завантаження даних

```
Cluster 1 => Apple Inc., Microsoft Corporation, Alphabet Inc., Amazon.com Inc., Tesla Inc., NVIDIA Corporation, Visa Inc., Home Depot
Cluster 2 => JPMorgan Chase & Co., Chevron Corporation
Cluster 3 => Walmart Inc., Procter & Gamble, Coca-Cola Company, PepsiCo Inc.
Cluster 4 => Pfizer Inc.
```

Рис 7. Результати

На даному етапі за допомогою API yfinance вдалося завантажити фінансові дані для зазначених в коді компаній, обчислити зміни між відкриттям і закриттям цін акцій, стандартизувати ці зміни, та виконати кластеризацію компаній. За допомогою метода GraphicalLassoCV вдалося побудувати коваріансе-матрицю, яка допомагає зрозуміти взаємозв'язки між компаніями. Також за допомогою алгоритму кластеризації AffinityPropagation, вдалося погрупувати компанії в кластери на основі подібності їх поведінки на ринку. В результаті було отримано кластеризації дані, за допомогою яких можна побачити які компанії належать до кожного кластеру.

**Висновок:** на даній лабораторній роботі я навчився використовувати кластеризацію даних за допомогою методу k-середніх, застосував цей метод на практиці. Також навчився оцінювати кількість кластерів з використанням методу зсуву середнього, застосував ці знання на практиці. Github: [link](#)