

ЛАБОРАТОРНА РОБОТА № 7

Наївний Байєс в Python

Мета роботи: набути навичок працювати з даними і опонувати роботу у Python з використанням теореми Байєса.

Завдання 2.1. Кластеризація даних за допомогою методу k-середні. Провести кластеризацію даних методом k-середніх. Використовувати файл вхідних даних: data_clustering.txt.

Код:

```
import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans

# Завантаження даних
data = np.loadtxt('data_clustering.txt', delimiter=',')

# Візуалізація початкових даних
plt.scatter(data[:, 0], data[:, 1], s=30, color='blue')
plt.title("Initial Data Points")
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.show()

# Кількість кластерів
num_clusters = 5
kmeans = KMeans(n_clusters=num_clusters, init='k-means++', random_state=42)

# Навчання моделі
kmeans.fit(data)
labels = kmeans.labels_
centroids = kmeans.cluster_centers_

# Візуалізація результатів кластеризації
for i in range(num_clusters):
    cluster_points = data[labels == i]
    plt.scatter(cluster_points[:, 0], cluster_points[:, 1], label=f'Cluster {i+1}')
plt.scatter(centroids[:, 0], centroids[:, 1], color='red', marker='x', s=200, label='Centroids')
plt.title("K-Means Clustering")
```

					ДУ «Житомирська політехніка».24.121.12.000 – Лр7			
Змн.	Арк.	№ докум.	Підпис	Дата				
Виконав	Левкович О.О.				Звіт з лабораторної роботи		Лім.	Арк.
Перевір.	Іванов Д.А.							1
Керівник								9
Н. контр.							ФІКТ Гр. ІПЗк-23-1	
Зав. каф.								

```
plt.xlabel("X-axis")
plt.ylabel("Y-axis")
plt.legend()
plt.show()
```

Результат:

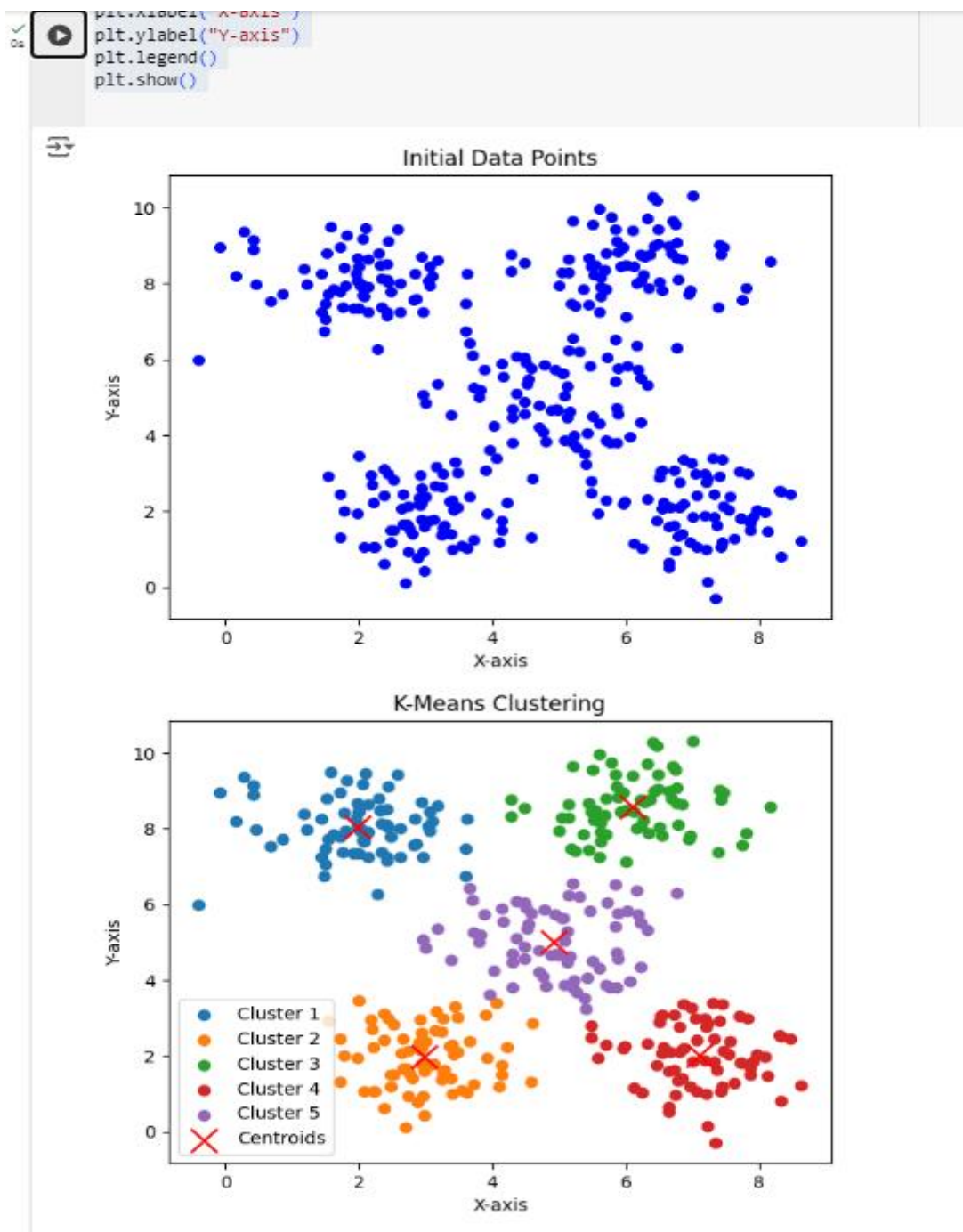


Рис. 1. Результат кластеризації даних методом k-середніх.

		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				2
Змн.	Арк.	№ докум.	Підпис	Дата		

Висновок: У роботі застосовано метод К-середніх для кластеризації даних з файлу data_clustering.txt. Було визначено 5 кластерів, що наочно продемонстровано на графіку з центроїдами. Результат показує групування схожих точок, що допомагає виявити структуру даних.

Завдання 2.2. Кластеризація К-середніх для набору даних Iris.

Виправлений код:

```
from sklearn.cluster import KMeans # Додано импорт KMeans
from sklearn.metrics import pairwise_distances_argmin

import numpy as np
from sklearn.datasets import load_iris
import matplotlib.pyplot as plt

# Завантаження даних Iris
iris = load_iris()
X = iris['data']
y = iris['target']

# Використання KMeans з n_clusters=5
kmeans = KMeans(n_clusters=5, random_state=42) # Додано random_state для
відтворюваності
kmeans.fit(X)
y_kmeans = kmeans.predict(X)

# Побудова графіку кластерів
plt.scatter(X[:, 0], X[:, 1], c=y_kmeans, s=50, cmap='viridis')
centers = kmeans.cluster_centers_
plt.scatter(centers[:, 0], centers[:, 1], c='black', s=200, alpha=0.5)
plt.title("KMeans Clustering with 5 Clusters")
plt.show()

# Визначення функції find_clusters
def find_clusters(X, n_clusters, rseed=2):
    rng = np.random.RandomState(rseed)
    i = rng.permutation(X.shape[0])[:n_clusters]
    centers = X[i]
    while True:
        # Призначення міток за відстанями до центрів
        labels = pairwise_distances_argmin(X, centers)
        # Перерахунок центрів
```

		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				3
Змн.	Арк.	№ докум.	Підпис	Дата		

```

        new_centers = np.array([X[labels == i].mean(0) for i in range(n_clusters)])
        # Перевірка на зупинку
        if np.all(centers == new_centers):
            break
        centers = new_centers
    return centers, labels

# Застосування find_clusters
centers, labels = find_clusters(X, 3)
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis')
plt.title("Custom Clustering (rseed=2)")
plt.show()

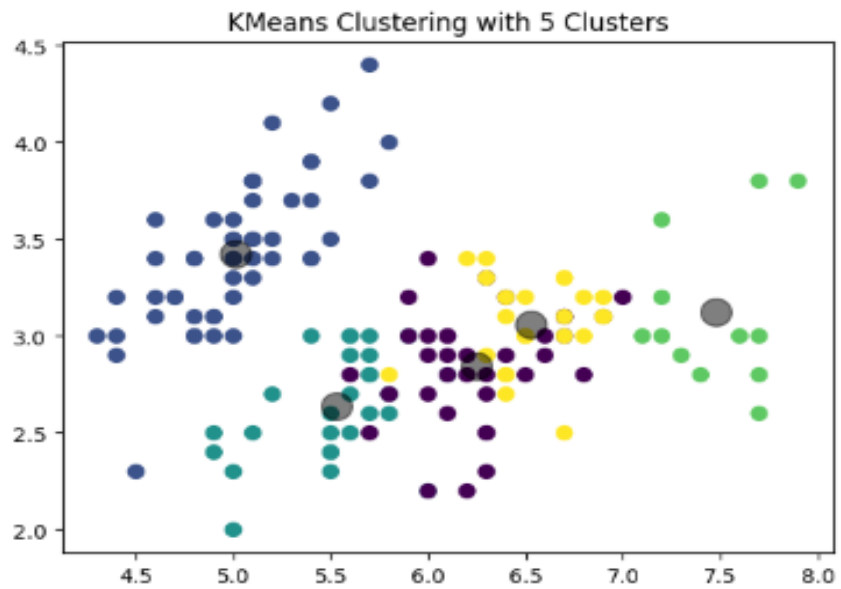
centers, labels = find_clusters(X, 3, rseed=0)
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis')
plt.title("Custom Clustering (rseed=0)")
plt.show()

# Порівняння з KMeans
labels = KMeans(3, random_state=0).fit_predict(X)
plt.scatter(X[:, 0], X[:, 1], c=labels, s=50, cmap='viridis')
plt.title("KMeans Clustering with 3 Clusters")
plt.show()

```

Результат:

		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				
Змн.	Арк.	№ докум.	Підпис	Дата		4



		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				5
Змн.	Арк.	№ докум.	Підпис	Дата		

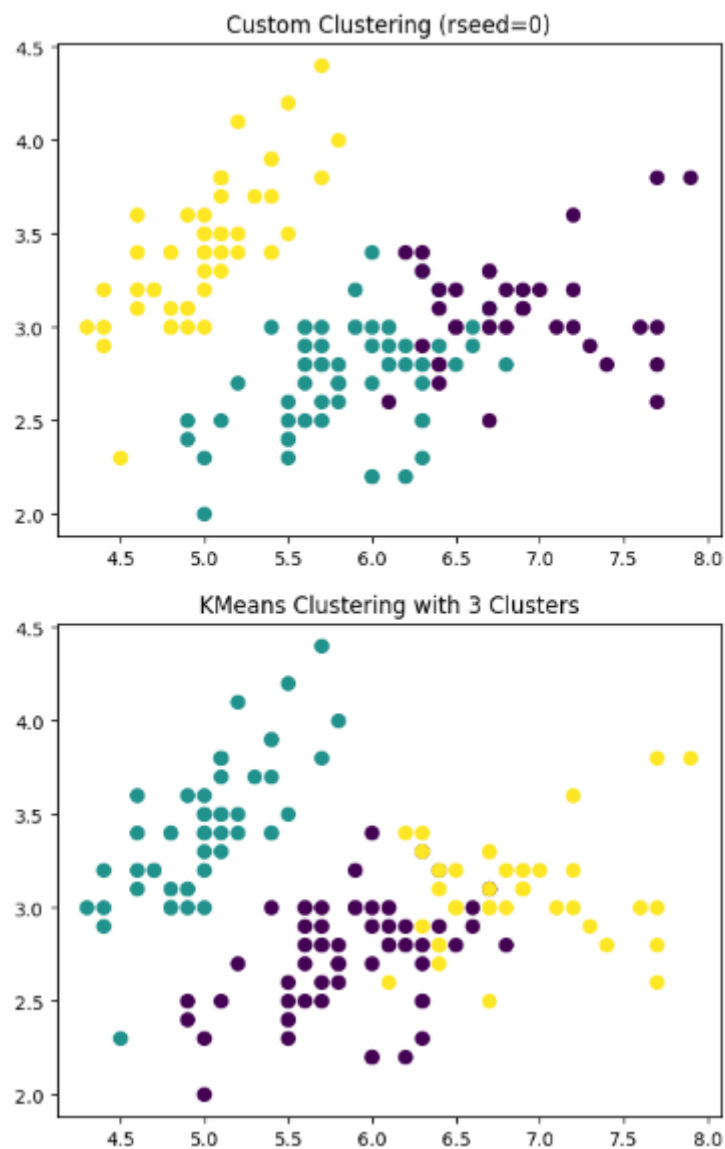


Рис. 2. Демонструє результати кластеризації для одного й того ж набору даних різними методами.

Висновок: Код порівнює кластеризацію даних Iris за допомогою стандартного KMeans та кастомного алгоритму. KMeans забезпечує стабільні результати, а кастомний метод залежить від початкових умов (rseed).

Завдання 2.3. Оцінка кількості кластерів з використанням методу зсуву середнього.

Код:

		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				6
Змн.	Арк.	№ докум.	Підпис	Дата		

```

import numpy as np
import matplotlib.pyplot as plt
from sklearn.cluster import MeanShift, estimate_bandwidth
from itertools import cycle

X = np.loadtxt('data_clustering.txt', delimiter=',')
# Оцінка ширини вікна для X
bandwidth_X = estimate_bandwidth(X, quantile=0.1, n_samples=len(X))

meanshift_model = MeanShift(bandwidth=bandwidth_X, bin_seeding=True)
meanshift_model.fit(X)

# Витягуємо центри всіх кластерів.
cluster_centers = meanshift_model.cluster_centers_
print('\nCenters of clusters:\n', cluster_centers)

# Витягуємо кількість кластерів.
labels = meanshift_model.labels_
num_clusters = len(np.unique(labels))
print("\nNumber of clusters in input data =", num_clusters)
plt.figure()
markers = 'o*xvs'
for i, marker in zip(range(num_clusters), markers):
    # Відображення на графіку точок, приналежачих
    # поточному кластеру
    plt.scatter(X[labels==i, 0], X[labels==i, 1], marker=marker,
color='black')

# Відображення на графіку центру кластера
cluster_center = cluster_centers[i]
plt.plot(cluster_center[0], cluster_center[1], marker='o',
markerfacecolor='black', markeredgecolor='black',
markersize=15)

plt.title('Кластери')
plt.show()

```

Результат:

		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				7
Змн.	Арк.	№ докум.	Підпис	Дата		

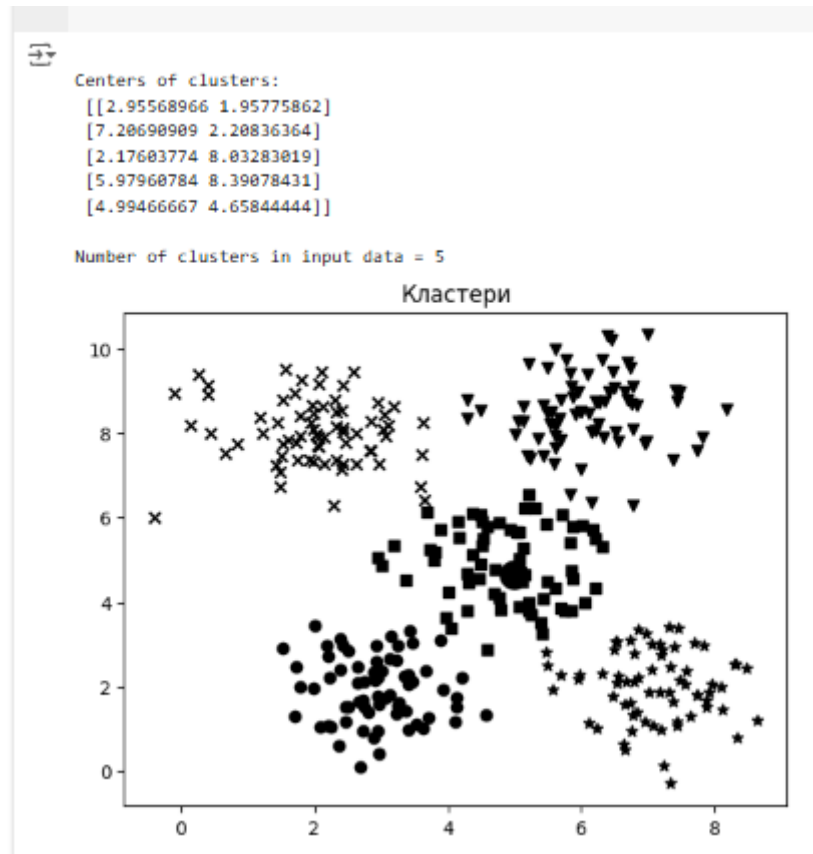


Рис. 3. Результат щодо оцінка кількості кластерів з використанням методу зсуву середнього.

Висновок: Результат виконує кластеризацію набору даних із файлу data_clustering.txt за допомогою алгоритму MeanShift. Автоматично оцінюється ширина вікна (bandwidth), визначаються центри кластерів та їх кількість. Результати відображаються на графіку, де кожен кластер має свій маркер.

Ці результати показують ефективність методу кластеризації, який успішно виявив та розділив дані на п'ять окремих груп. Чітко виражені кластери свідчать про якісну сегментацію даних.

Завдання 2.4. Знаходження підгруп на фондовому ринку з використанням моделі поширення подібності.

		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				8
Змн.	Арк.	№ докум.	Підпис	Дата		

Відсутність файлу **company_symbol_mapping.json** та функція **quotes_historical_yahoo_ochl** яка була частиною бібліотеки **matplotlib.finance**, але починаючи з версії **Matplotlib 2.0** модуль **matplotlib.finance** був видалений.

***Висновок:** Під час виконання лабораторної роботи, використовуючи мову програмування **Python** та сервіс **google collab** в якому йшли спеціалізовані бібліотеки, було проведено дослідження методів неконтрольованої класифікації даних у сфері машинного навчання.*

Github - <https://github.com/TAMOTO24/-Intelligen-Systems>

		Левкович О.О.			ДУ «Житомирська політехніка». 24.121.12.000 – Лр7	Арк.
		Іванов Д.А.				9
Змн.	Арк.	№ докум.	Підпис	Дата		