

## SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

Laboratorium Nr 8 Data 11.01.2025 Temat: „Praktyczne zastosowanie analizy skupień (clustering) do zbiorów danych” Wariant 11	Szymon Nycz Informatyka II stopień, niestacjonarne, 1 semestr, gr.1b
---	---

### 1. Polecenie:

W Pythonie zastosuj DBSCAN na danych Wine z `sklearn.datasets`. Zwizualizuj wyniki na dwóch głównych składowych PCA.

Link do repozytorium:

[https://github.com/Maciek332/Semestr\\_1\\_Nycz/tree/master/NoD](https://github.com/Maciek332/Semestr_1_Nycz/tree/master/NoD)

## 2. Opis programu opracowanego

```
from sklearn.datasets import load_digits
from sklearn.decomposition import PCA
from scipy.cluster.hierarchy import dendrogram, linkage
import matplotlib.pyplot as plt

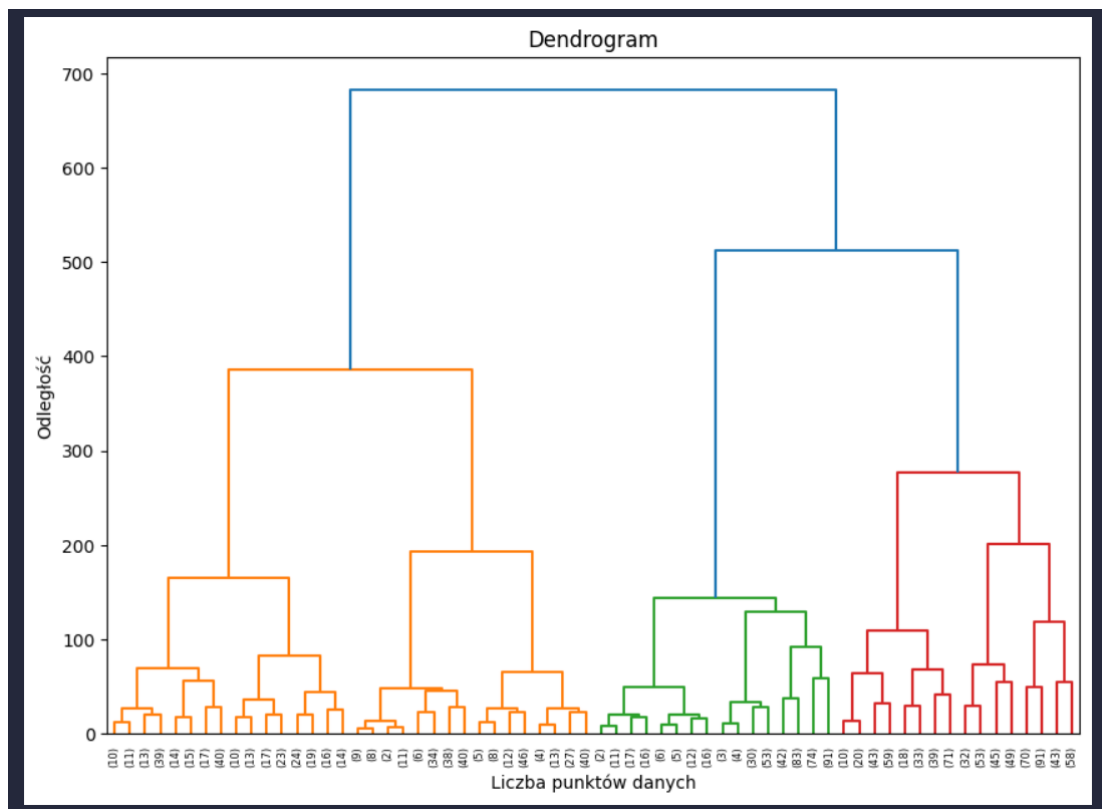
# Załaduj dane Digits
digits = load_digits()
X = digits.data

# Zastosuj PCA do redukcji wymiarów
pca = PCA(n_components=2)
X_pca = pca.fit_transform(X)

# Wykonaj hierarchiczną analizę skupień
linked = linkage(X_pca, 'ward')

# Wygeneruj dendrogram
plt.figure(figsize=(10, 7))
dendrogram(linked, truncate_mode='level', p=5)
plt.title('Dendrogram')
plt.xlabel('Liczba punktów danych')
plt.ylabel('Odległość')
plt.show()
```

✓ 1.7s



### 3. Wnioski

Analiza skupień to technika uczenia nienadzorowanego, której celem jest grupowanie danych na podstawie ich podobieństwa. Jest często stosowana jako wstępny etap w uczeniu maszynowym do redukcji wymiarowości, tworzenia cech na podstawie klastrów, wstępnej analizy struktury danych i detekcji anomalii. K-Means to jeden z najczęściej używanych algorytmów do analizy skupień. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) grupuje dane według gęstości punktów, identyfikując przy tym anomalie. Hierarchical Clustering tworzy dendrogram, co umożliwia analizę danych na różnych poziomach szczegółowości.