SPRAWOZDANIE

Zajęcia: Nauka o danych I

Prowadzący: prof. dr hab. Vasyl Martsenyuk

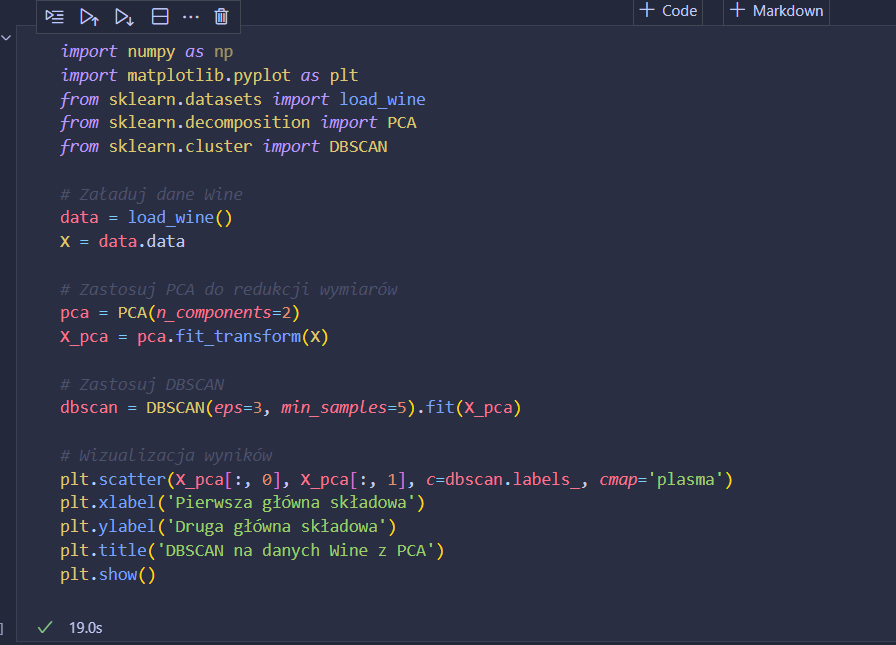
|  |  |
| --- | --- |
| Laboratorium Nr 8  Data 11.01.2025  Temat: „Praktyczne zastosowanie analizy skupień (clustering) do zbiorów danych”  Wariant 10 | Anna Więzik  Informatyka  II stopień, niestacjonarne,  1 semestr, gr.1b |

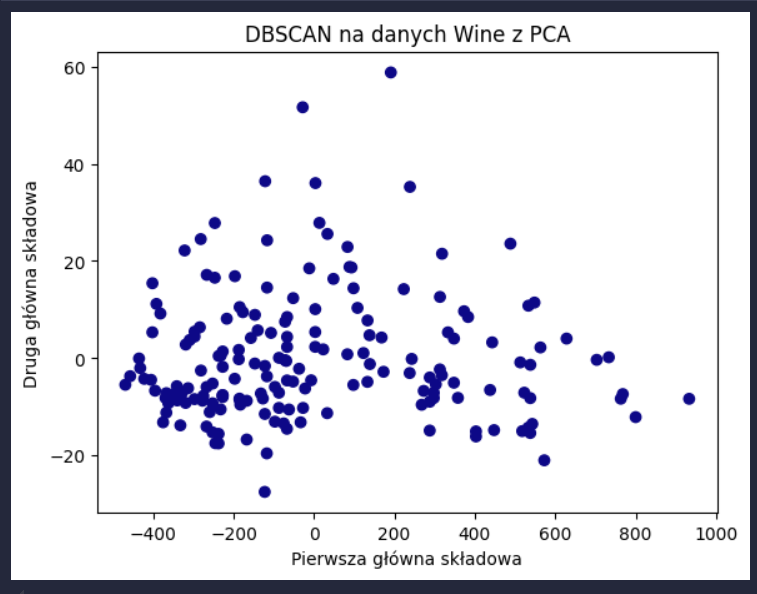
1. Polecenie:

W KNIME u˙zyj hierarchicznej analizy skupie´n na zbiorze Digits. Przedstaw dendrogram i wybierz optymalna liczbe klastr´ow.

Link do repozytorium: <https://github.com/AnaShiro/NoD1_2024>

1. Opis programu opracowanego





1. Wnioski

Analiza skupień to technika uczenia nienadzorowanego, której celem jest podział danych na grupy (klastry) na podstawie ich podobieństwa. Jest często wykorzystywana jako etap wstępny w uczeniu maszynowym, szczególnie do redukcji wymiarowości i uproszczenia złożoności danych, tworzenia cech na podstawie klastrów (clustering-based features), wstępnej analizy struktury danych (eksploracja danych) oraz detekcji anomalii i punktów odstających. K-Means jest jednym z najczęściej stosowanych algorytmów analizy skupień. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) grupuje dane na podstawie gęstości punktów, jednocześnie identyfikując anomalie. Hierarchical Clustering tworzy strukturę drzewa (dendrogram), co pozwala analizować dane na różnych poziomach szczegółowości.