

Tutorial 6: Analiza skupień

1. Przygotować zbiór danych do pracy (przykładowo: może być to treningowy zbiór danych do prognozowania zapotrzebowań na pożyczki, taki jak tutaj: <https://datahack.analyticsvidhya.com/contest/practice-problem-loan-prediction-iii/>)
2. Przygotować krótką prezentację o trzech popularnych algorytmach w analizie skupień: metoda k-means, metoda grupowania wokół centroidów (partitioning around medoids) oraz metoda aglomeracyjnego klastrowania hierarchicznego.
3. Przedyskutować typowe wyzwania poszczególnych metod, w szczególności różne wielkości poszczególnych klasterów, różne gęstości punktów, złożoność obliczeniową grupowania wokół centroidów itp. Zilustrować wspomniane problemy prostymi przykładami.

ZADANIE:

Zadanie programistyczne ma polegać na napisaniu kodu w oparciu o wybrane metody analizy skupień i przetestowania go na wybranym wcześniej zbiorze danych testowych.