Metody Analizy Danych – zadanie projektowe 1a

**Jaki analizę można wybrać?**

* porządkowanie obiektów
* analiza skupień

Dane: własne dane (np. wybrane z GUS), repozytoria Machine Learningu, dane z pakietów w R

Grupy maks. 5 os.

**Co powinno znaleźć się w projekcie?**

* Tytuł pracy
* Autorzy/Autor
* Streszczenie (maksymalnie 150 słów)
* Słowa kluczowe
* Wprowadzenie
* Przedmiot badania
  + Cel i zakres badania
  + Przegląd literatury (min. 1 cytowanie powiązane tematycznie i krótki opis co było badane)
  + Zmienne wybrane do analizy (opis i uzasadnienie zmiennych oraz podział na stymulanty/destymulanty) – minimum sześć zmiennych
  + Wstępna analiza danych
    - statystyki opisowe (przynajmniej: średnia, mediana, minimum, maksimum, odchylenie standardowe, skośność)
    - podstawowa wizualizacja – np. boxplot, histogramy
    - braki danych – czy występują i jak je obsłużono
    - obserwacje odstające i w jaki sposób je obsłużono
* Opis metod
  + wzory wraz z opisami oznaczeń
  + cytowanie pracy w której zaproponowano metodę/ewentualnie pracy, w której użyto metodę)
* Rezultaty (w postaci tabelarycznej i/lub graficznej oraz omówienie wyników)
* Podsumowanie (ocena realizacji celu, odniesienie do pozycji z przeglądu literatury)
* Bibliografia

W przypadku porządkowania obiektów stworzenie minimum trzech rankingów (np. z różnych metod normalizacji, różnych metod porządkowania), w przypadku analizy skupień podział na skupienia minimum dwoma metodami (np. k-średnich, warda). Należy porównać ze sobą uzyskane rankingi/skupienia (np. podobieństwo rankingów, ocena jakości skupień).

**Co należy oddać?**

* Zbiór danych + Plik z opisem (pdf) + plik z obliczeniami (w zależności od tego gdzie wykonane to .R, .py, .xlsx)

lub

* Zbiór danych + Jupyter Notebook (.ipynb i wygenerowany .html)

lub

* Zbiór danych + R Notebook (.Rmd i wygenerowany .html)