Metody Analizy Danych – projekt drugi

**Jaki analizę należy wykonać?**

* Klasyfikacja

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

Wymagania zbioru: minimum 10 zmiennych, minimum 800 obserwacji

Grupy maksymalnie 5 osobowe

**Co powinno znaleźć się w projekcie?**

* Tytuł pracy
* Autorzy/Autor
* Streszczenie (maksymalnie 150 słów)
* Słowa kluczowe
* Wprowadzenie
* Przedmiot badania
  + Cel
  + Wstępna analiza danych
    - Przedstawienie dostępnych zmiennych
    - statystyki opisowe (przynajmniej: średnia, mediana, minimum, maksimum, odchylenie standardowe, skośność)
    - podstawowa wizualizacja – np. boxplot, histogramy
    - transformacje danych (skalowanie, logarytmowanie)
    - braki danych – czy występują i jak je obsłużono
    - obserwacje odstające i w jaki sposób je obsłużono
* Opis metod
  + minimum trzy metody i stworzenie modelu hybrydowego (np. średnia ważona score’ów z metod), formalnie więc: minimum cztery mechanizmy decyzyjne
  + cytowanie pracy, w której zaproponowano metodę (ewentualnie pracy, w której użyto metodę) i opis ogólny działania metody
* Rezultaty
  + Mierniki (np. AUC, Accuracy, F1 Score) – minimum jeden
  + Sposób walidacji (np. krzyżowa, leave-one-out, wielokrotne próbkowanie) – minimum jeden sposób
* Przykład użycia modeli na stworzonych sztucznie obserwacjach
* Bibliografia

**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)

**Jak sformatować plik z opisem?**

Sugerowane formatowanie podobne do artykułu naukowego z wybranego czasopisma (bez limitu stron). Przykładowo: <http://qme.sggw.pl/wp-content/uploads/MIBE_Szablon2019.docx>

Pliki z projektem powinny zostać załadowane na MS Teams bez archiwizowania.