

# Zaawansowane metody i techniki analizy danych

## OPIS PROJEKTU

*Celem projektu jest zastosowanie nabytej wiedzy do samodzielnie przeprowadzonej analizy różnorodnych danych oraz wyciągania uogólnionych wniosków na ich podstawie.*

**Termin oddania projektu podaje prowadzący i jest on nieprzekraczalny!**

**Dane użyte w projekcie nie mogą być danymi używanymi podczas laboratoriów!**

### Część 1

1. Znajdź odpowiednie dane do analizy. Dane muszą zawierać kilkaset przykładów oraz atrybuty ciągłe i dyskretne. Dokonaj wstępnego przygotowania danych np. wypełnij brakujące wartości atrybutów, dokonaj dyskretyzacji atrybutów ciągłych.
2. Przypisz odpowiedni typ danych każdej z kolumn a następnie opisz kolumny za pomocą np. minimalna/maksymalna wartość, najczęściej występująca itp.
3. Przeanalizuj dane np. pokaż strukturę wybranego zbioru danych, znajdź zależności pomiędzy poszczególnymi kolumnami, zaproponuj dodatkowe kolumny itp.
4. Przeprowadź badania dla dwóch różnych algorytmów klasyfikacji. Jeśli to możliwe uwzględnij różne ustawienia parametrów algorytmów, spróbuj określić ich wpływ na wyniki. Oceń wyniki uczenia się za pomocą walidacji krzyżowej.
5. Przeprowadź badania dla dwóch różnych algorytmów grupowania. Jeśli to możliwe uwzględnij różne ustawienia parametrów algorytmów, spróbuj określić ich wpływ na wyniki. Oceń wyniki grupowania ze względu na liczbę grup. Znajdź/wygeneruj (mały) zbiór danych, w których znany jest podział na grupy. Sprawdź, który z wybranych algorytmów generuje podział na klastry zbliżony do prawdziwego.
6. Uzyskane wyniki przedstaw w czytelnej formie tabel i/lub wykresów. Przeanalizuj uzyskane wyniki, podejmij próbę ich wyjaśnienia, przedyskutuj wynikające z nich wnioski. Napisany kod opisz stosownymi komentarzami.