



# SZTUCZNA INTELIGENCJA

## ZADANIA ZALICZENIOWE



Do zaliczenia ćwiczeń są do zrealizowania 3 modele obliczeniowe w RapidMinerze oraz jeden w postaci implementacji własnej:

1. Zbudowanie w RapidMinerze możliwie efektywnego modelu obliczeniowego (wykorzystanie grupy metod z optymalizacją parametrów modeli) dla nietrywialnego zbioru etykietowanych danych uczących (kilka set lub tysięcy wzorców) wykorzystującego uczenie nadzorowane (supervised learning) dla wybranego problemu klasyfikacji.
2. Zbudowanie w RapidMinerze możliwie efektywnego modelu obliczeniowego (wykorzystanie grupy metod z optymalizacją parametrów modeli) dla nietrywialnego zbioru etykietowanych danych uczących (kilka set lub tysięcy wzorców) wykorzystującego uczenie nadzorowane (supervised learning) dla wybranego problemu regresji (aproksymacji funkcji nieliniowej).
3. Zbudowanie w RapidMinerze możliwie efektywnego modelu obliczeniowego (wykorzystanie grupy metod z optymalizacją parametrów modeli) dla nietrywialnego zbioru nieetykietowanych danych uczących (kilka set lub tysięcy wzorców) wykorzystującego uczenie nienadzorowane (unsupervised learning) dla wybranego problemu klasteryzacji.
4. Zaimplementowanie w wybranym języku programowania (np. C++, C#, Python, Java) jednego z poniższych modeli obliczeniowych z zakresu inteligencji obliczeniowej i wytrenowanie go:
  - Metodę k Najbliższych Sąsiadów z automatycznym wyznaczaniem optymalnej wartości „k” dla wybranych 3 zbiorów uczących
  - Drzewa lub diagramy decyzyjne dla wybranych 3 zbiorów uczących
  - Samoorganizujące się mapy - sieci Kohonena (SOM) – z graficzną wizualizacją powstałych klastrów dla 3 zbiorów uczących
  - Algorytm genetyczny lub algorytm ewolucyjny zastosowany do wybranych zbiorów danych uczących w celu znalezienia globalnego lub lokalnego minimum funkcji czy modelu.
  - Sieć neuronową typu MLP (Multi-Layer Perceptron) możliwie z walidacją krzyżową (cross-validation) uczoną metodą propagacji wstecznej błędów (backpropagation).
  - Sieć neuronową typu RBF (Radial Basis Function Network) możliwie z walidacją krzyżową (cross-validation) uczoną metodą propagacji wstecznej błędów (backpropagation).
  - Sieć skojarzeniową AGDS z wyszukiwaniem wzorców najbardziej podobnych do wskazanego lub kombinacji parametrów wejściowych.
  - Inną metodę z zakresu inteligencji obliczeniowej uzgodnioną z prowadzącym ćwiczenia.

Dane można szukać na stronie [UCI ML Repository](#) lub wykorzystać jakieś własne zbiory.

Ćwiczenia należy oddać prowadzącemu na ostatnich zajęciach (w postaci kodu źródłowego i w wersji skompilowanej), wytłumaczyć sposób działania programu, kodu oraz wybranej metody inteligencji obliczeniowej.



# ZADANIA ZALICZENIOWE



Sposób realizacji zadań w Rapid Minerze:

1. Poszukać kilka fajnych zbiorów danych (zawierających kilkaset lub kilka tysięcy wzorców oraz od kilku do kilkudziesięciu atrybutów) na stronie [UCI ML Repository – All Data Sets](#) z różnych kategorii do różnych zadań na zaliczenie:
  - Classification
  - Regression
  - Clustering
2. Przeczytać opisy tych zbiorów, sposób formatowania. Wciągnąć sobie dane bezpośrednio do Rapid Minera, jak się da, a jak nie, to wczytać sobie dane do arkusza kalkulacyjnego i odpowiednio przerobić, np. niektóre zbiory nie posiadają nagłówków kolumn itp. W razie potrzeby można część danych symbolicznych zastąpić danymi numerycznymi.
3. Następnie spróbować zbudować prosty projekt klasyfikacji, regresji, klasteryzacji ... tak jak to było, opisane w przerobionych tutorialach na ćwiczeniach.
4. W następnej kolejności podnosimy sobie poprzeczkę i próbujemy dodać walidację krzyżową (X-Validation, CrossValidation), bloczki optymalizujące ilość atrybutów danych (Optimize Selection...), bloczki optymalizujące parametry metody (Optimize Parameters...), oraz zbudować komitet kilku różnych klasyfikatorów czy też dobierać rozwiązanie z wykorzystaniem bloczków ewolucyjnych.
5. Próbujemy znaleźć możliwie najlepsze rozwiązanie, które dobrze klasyfikuje dane uczące, jak i te walidacyjne czy testowe.
6. Na ostatnie zajęcia w semestrze przygotowujemy prezentację działania swoich rozwiązań (ok. 15 min/osobę), próbując pokazać wszystkim, co i jak udało się osiągnąć, wraz z opisem własnego podejścia do zbioru danych.
7. Końcowe projekty w Rapid Minerze i te programistyczne w jednej paczce, nazwa pliku: skrót przedmiotu + Imię i Nazwisko, spakowane w formacie RAR lub ZIP przesyłamy prowadzącemu po prezentacji do oceny.