

Raport z przebiegu laboratorium: Upraszczenie drzew decyzyjnych

Zespół: (imiona i nazwiska):

Data:

Dzień i godzina zajęć:

Grupa:

1. Używając programu WEKA zbuduj drzewo decyzyjne (algorytmem J48) dla problemu diagnozowania raka piersi, dla którego dane uczące umieszczono w pliku *breast-cancer.arff*.

1.1. Porównaj drzewa wygenerowane z użyciem mechanizmu upraszczania i bez niego. Jakie są różnice w rozmiarach drzewa a jakie w skuteczności klasyfikacji? Weź pod uwagę skuteczność na zbiorze uczącym i zbiorze testowym (uzyskanym za pomocą podziału pierwotnego zbioru danych metodami *percentage-split* i *k-fold cross validation*).

1.2. Przeanalizuj działanie poszczególnych mechanizmów upraszczania drzewa – jak je oceniasz? Naszkicuj wykres zależności trafności klasyfikowania na zbiorach uczącym i testującym (z użyciem walidacji krzyżowej) od rozmiaru drzewa (przy różnych ustawieniach parametrów *confidence factor* oraz *min-num-objects*).

2. Zastosuj poznane algorytmy klasyfikacji dla problemu oceny stanów gry w kółko i krzyżyk (plik *tic-tac-toe.arff*). Który algorytm i przy jakich parametrach spisuje się najlepiej? Czy potrafisz uzasadnić dlaczego dany klasyfikator dobrze radzi sobie z tym konkretnym problemem?