

Raport z przebiegu laboratorium: Drzewa decyzyjne

Zespół: (imiona i nazwiska):

Data:

Dzień i godzina zajęć:

Grupa:

1. Dokonaj analizy zbioru danych uczących znajdujących się w pliku *komputery.arff*.

1.1. Jakie są liczebności przykładów należących do każdej z klas decyzyjnych?

1.2. Oblicz miarę entropii dla zbioru wszystkich przykładów.

1.3. Jakie podziały obiektów generuje każdy z atrybutów warunkowych? Jaki jest rozkład klas dla przykładów o każdej z możliwych wartości wybranego atrybutu?

1.4. Używając entropii warunkowej i miary przyrostu informacji określ który atrybut zostanie wybrany jako najkorzystniejszy do zbudowania testu w korzeniu drzewa.

2. Przy użyciu programu WEKA zbuduj drzewo decyzyjne (algorytm ID3) dla danych z pliku *komputery.arff*.

2.1. Naskicuj poniżej drzewo wygenerowane przez program.

2.2. Ile węzłów ma wygenerowane drzewo? Ile ma liści? Ile jest możliwych ścieżek decyzyjnych wychodzących z korzenia drzewa? Jak wygląda zestaw warunków z najdłuższej ścieżki?

2.3. Przy użyciu wygenerowanego drzewa dokonaj klasyfikacji 3 samodzielnie stworzonych przykładów, które nie pojawiają się w danych uczących. Podaj opis przykładów oraz wynik klasyfikacji.

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

3.1. Zbuduj drzewo bez mechanizmu upraszczania (opcja *unpruned* ustawiona na wartość *true*). Jaki jest rozmiar i skuteczności klasyfikacji tego drzewa? Weź pod uwagę skuteczność na zbiorze uczącym i nieznanym zbiorze testowym (uzyskanym za pomocą podziału pierwotnego zbioru danych metodami *percentage-split* i *k-fold cross validation*, z *k* ustawionym na wartość 10).

3.2. Zbuduj drzewo z włączonym mechanizmem upraszczania (opcja *unpruned* ustawiona na wartość *false*). Jaki jest rozmiar i skuteczności klasyfikacji tego drzewa? Jakie wnioski można wyciągnąć porównując rozmiar i skuteczność klasyfikacji drzew uproszczonych z drzewami wygenerowanymi wcześniej?