

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 przeprowadź proces budowy drzewa (algorytmem 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.

2.4. Spróbuj ręcznie uprościć drzewo zastępując kolejne węzły decyzyjne drzewa liśćmi, którym przypisane są klasy decyzyjne. Naskicuj poniżej uproszczone drzewa. Jaka jest trafność klasyfikacji każdego z uproszczonych drzew na zbiorze uczącym?

2.5. Ile węzłów ma najmniejsze, a ile największe drzewo, jakie można skonstruować dla tego problemu? Czy potrafisz stwierdzić jaką trafność klasyfikacji na zbiorze uczącym będą miały te dwa skrajne przypadki drzew decyzyjnych? Co można powiedzieć o skuteczności klasyfikacji na nieznanym zbiorze testującym?