

**Raport z przebiegu laboratorium: Drzewa decyzyjne**

Zespół (imiona i nazwiska):

Data: .....

Grupa: .....

**Zad. 1.** Załóżmy, że zebraliśmy następujące obserwacje dotyczące tego, czy nasz/a przyjaciel/przyjaciółka przyjęła zaproszenie na spotkanie:

Occasion	Day of Week	Weather	TV Tonight	Accepted?
a	<i>weekday</i>	<i>warm</i>	<i>bad</i>	<b>no</b>
b	<i>weekday</i>	<i>warm</i>	<i>good</i>	<b>no</b>
c	<i>weekend</i>	<i>warm</i>	<i>bad</i>	<b>yes</b>
d	<i>weekend</i>	<i>cold</i>	<i>good</i>	<b>no</b>
e	<i>weekend</i>	<i>cold</i>	<i>bad</i>	<b>yes</b>

a) Jaka decyzja jest Twoim zdaniem bardziej prawdopodobna dla nowych przypadków w postaci (*weekday*, \*, \*), gdzie \* oznacza dowolną wartość?      YES      NO      ciężko ocenić

b) Brak pozytywnych przypadków dla *weekday* nie oznacza jednak, że konieczne dla spotkania poza weekendem odpowiedź byłaby negatywna. Zaproponuj hipotezę, która w inny sposób wytłumaczy odmowną decyzję.

**Zad. 2.** Rozrysuj podziały ze względu na wartości na poszczególnych atrybutach. Policz entropie ze względu na decyzje w poszczególnych podzbiorach

a) **Day of Week**

- *weekday*

a (no)	b (no)
--------	--------

Entropia:  $-\frac{0}{2} \cdot \log_2 \frac{0}{2} - \frac{2}{2} \cdot \log_2 \frac{2}{2} = 0$

- *weekend*

c (yes)	d (no)	e (yes)
---------	--------	---------

Entropia:  $-\frac{1}{3} \cdot \log_2 \frac{1}{3} - \frac{2}{3} \cdot \log_2 \frac{2}{3} \approx 0.92$

b) **Weather**

- *warm*

--

Entropia: .....

- *cold*

--

Entropia: .....

c) **TV Tonight**

- *bad*

--

Entropia: .....

- *good*

--

Entropia: .....

**Zad. 3.** Korzystając z materiałów albo Internetu, oraz wyżej wyliczonych wartości, oblicz zysk informacyjny dla poszczególnych atrybutów. Który atrybut powinien znaleźć się w korzeniu drzewa?

**Zad. 4.** Dokończ obliczenia z zad. 3. i narysuj kompletne wynikowe drzewo decyzyjne. Sprawdź następnie, co ono zwróci dla przypadku (*weekend, warm, good*).

**Zad. 5.** Przy użyciu programu WEKA przeprowadź proces budowy drzewa decyzyjnego (*trees/J48*) dla danych z pliku *komputery.arff*. Jako zbiór testowy użyj zbioru uczącego.

a) Naszkicuj poniżej drzewo wygenerowane przez program. Nanieś nazwy atrybutów/klas decyzyjnych reprezentowanych przez węzły i wartości przypisane do poszczególnych gałęzi.

b) Odpowiedz na poniższe pytania:

- Ile węzłów ma wygenerowane drzewo? \_\_\_\_\_

- Ile ma liści? \_\_\_\_\_

- Ile jest ścieżek decyzyjnych wychodzących z korzenia tego drzewa? \_\_\_\_\_

- Jaka decyzja zostanie przydzielona przykładowi (*wysokie, nie, mezczyzna*)? \_\_\_\_\_

- Jaka decyzja zostanie przydzielona przykładowi (*niskie, tak, mezczyzna*)? \_\_\_\_\_

**Zad. 6.** Opisz własnymi słowami ogólną ideę uczenia realizowaną przez drzewa decyzyjne. Pamiętaj, że entropia to tylko pewna miara służąca do konkretnego celu, a nie „kluczowy” komponent algorytmu.