

# Drzewa decyzyjne

Zwierzę	Ma skrzydła	Ma futro	Żyje w wodzie	Klasa (cel)
Kot	Nie	Tak	Nie	Ssak
Pies	Nie	Tak	Nie	Ssak
Delfin	Nie	Nie	Tak	Ssak
Orzeł	Tak	Nie	Nie	Ptak
Kura	Tak	Nie	Nie	Ptak
Rekin	Nie	Nie	Tak	Ryba
Łosoś	Nie	Nie	Tak	Ryba

## Entropia

$$H(S) = -p_1 \cdot \log_2(p_1) - p_2 \cdot \log_2(p_2) - p_3 \cdot \log_2(p_3) - \dots - p_k \cdot \log_2(p_k)$$

$p_i$  – udział klasy  $i$  w zbiorze:  $p_i = \frac{|S_i|}{|S|}$ ,

$k$  – liczba klas w zbiorze,

jeżeli  $p_i = 0$ , przyjmujemy  $0 \cdot \log_2(0) = 0$ .

- ◆ Entropia mówi, **jak bardzo dane są pomieszane**.
- ◆ Entropia = 0 → wszystko tej samej klasy
- ◆ Entropia = 1 → pełny chaos (po równo różnych klas)

## Entropia podziału

$$H_{\text{podziału}} = \frac{|S_1|}{|S|} \cdot H(S_1) + \frac{|S_2|}{|S|} \cdot H(S_2) + \dots + \frac{|S_n|}{|S|} \cdot H(S_n)$$

$H(S_j)$  obliczamy wg wzoru powyżej dla każdej grupy  $S_j$ ,

$|S_j|$  – liczba elementów w grupie  $j$ ,

$|S|$  – liczba elementów w całym zbiorze.

$$IG(\text{cecha}) = H(S) - H_{\text{podziału}}$$

<b>p</b>	<b><math>\log_2(p)</math></b>	<b><math>-p \cdot \log_2(p)</math></b>
0	-	0
$1/7 \approx 0.143$	$\log_2(0.143) \approx -2.807$	$0.143 * 2.807 \approx 0.401$
$2/7 \approx 0.286$	$\log_2(0.286) \approx -1.807$	$0.286 * 1.807 \approx 0.517$
$3/7 \approx 0.429$	$\log_2(0.429) \approx -1.222$	$0.429 * 1.222 \approx 0.524$
1	$\log_2(1) = 0$	0