

EXPERTNE SYSTEMY

STROJOVE UČENIE

CVIČENIE: GENEROV. ROZHOD. STROMOV - ID3

TRENOVACIA MNOŽINA

	BLUETOOTH	ZNAČKA	FOTOGRAFUJE	TRIEDA	
1.	nema'	nokia	N	-	T_1
2.	ma'	siemens	N	-	T_1
3.	ma'	nokia	A	+	T_2
4.	ma'	siemens	A	-	T_8
5.	nema'	siemens	A	-	
6.	ma'	samsung	A	+	
7.	ma'	nokia	N	-	
8.	nema'	nokia	A	+	

Odvodte rozhodovací strom pre danú trenovaciu množinu. Práňte algoritmus ID3.

BLUETOOTH

ma'	nema'	spolu
2+3-	1+2-	3+5-
$H(ma') = 0.971$	$H(nema') = 0.918$	$H(S) = 0.954$
$P(ma') = 5/8$	$P(nema') = 3/8$	
$H(S, BLUETH) = 0.951$		$I(BLUETH) = 0.003$

$$H(S) = -\sum_{j=1}^2 \frac{n_j}{n_1+n_2} \log_2 \frac{n_j}{n_1+n_2}$$

$$H(S) = -\frac{3}{3+5} \log_2 \frac{3}{3+5} - \frac{5}{3+5} \log_2 \frac{5}{3+5} = -\frac{3}{8} \log_2 \frac{3}{8} - \frac{5}{8} \log_2 \frac{5}{8} = 0.954$$

$$H(ma') = -\frac{2}{5} \log_2 \frac{2}{5} - \frac{3}{5} \log_2 \frac{3}{5} = 0.971$$

$$H(nema') = -\frac{1}{3} \log_2 \frac{1}{3} - \frac{2}{3} \log_2 \frac{2}{3} = 0.918$$

$$H(S, A) = \sum_{i=1}^2 \frac{m_i}{m_1+m_2} \cdot H(S_i)$$

$$H(S, BLUETH) = \frac{5}{8} H(ma') + \frac{3}{8} H(nema') = \frac{5}{8} \cdot 0.971 + \frac{3}{8} \cdot 0.918 = 0.951$$

$$I(BLUETH) = H(S) - H(S, A)$$

$$I(BLUETH) = H(S) - H(S, BLUETH) = 0.954 - 0.951 = 0.003$$

ZNAČKA

nokia	siemens	samsung	spolu
2+ 2-	0+ 3-	1+ 0-	3+ 5-
$H(\text{nokia}) = 1$	$H(\text{siemens}) = 0$	$H(\text{samsung}) = 0$	$H(S) = 0.954$
$P(\text{nokia}) = 1/2$	$P(\text{siemens}) = 3/8$	$P(\text{samsung}) = 1/8$	
$H(S, \text{ZNAČKA}) = 0.5$			$I(\text{ZNAČKA}) = 0.454$

$$H(\text{nokia}) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = -2 \left(\frac{1}{2} \log_2 \frac{1}{2} \right) = -1(-1) = 1$$

$$H(\text{siemens}) = -\frac{0}{3} \log_2 \frac{0}{3} - \frac{3}{3} \log_2 \frac{3}{3} = -0 - 1 \log_2 1 = 0$$

$$H(\text{samsung}) = -\frac{1}{1} \log_2 \frac{1}{1} - \frac{0}{1} \log_2 \frac{0}{1} = 0$$

$$H(S, \text{ZNAČKA}) = \frac{1}{2} \cdot 1 + \frac{3}{8} \cdot 0 + \frac{1}{8} \cdot 0 = 0.5$$

$$I(\text{ZNAČKA}) = 0.954 - 0.5 = \underline{\underline{0.454}}$$

FOTOGRAFUJE

A	N	spolu
3+ 2-	0+ 3-	3+ 5-
$H(A) = 0.971$	$H(N) = 0$	$H(S) = 0.954$
$P(A) = 5/8$	$P(N) = 3/8$	
$H(S, \text{FOTOGRAFUJE}) = 0.607$		$I(\text{FOT}) = 0.347$

$$H(A) = -\frac{3}{5} \log_2 \frac{3}{5} - \frac{2}{5} \log_2 \frac{2}{5} = 0.971$$

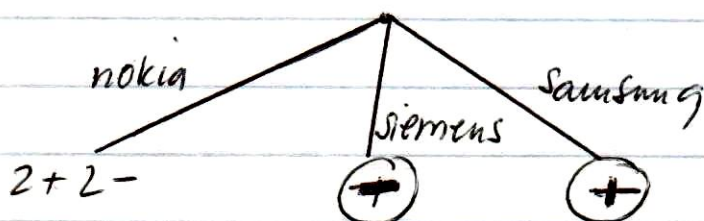
$$H(N) = -\frac{0}{3} \log_2 \frac{0}{3} - \frac{3}{3} \log_2 \frac{3}{3} =$$

$$= -0 - 1 \cdot \log_2 1 = 0$$

$$H(S, \text{FOTOGRAFUJE}) = \frac{5}{8} \cdot H(A) + \frac{3}{8} \cdot H(N) = \frac{5}{8} \cdot 0.971 + \frac{3}{8} \cdot 0 = 0.607$$

$$I(\text{FOTOGRAFUJE}) = H(S) - H(S, \text{FOTOGRAFUJE}) = 0.954 - 0.607 = \underline{\underline{0.347}}$$

ZNAČKA



NOVÁ TRÉNOVACIA MNOŽINA

	BLUTNIS	ZNAČKA	FOTOGRAFUJE	TRIEDA
1	nema'	nokia	N	-
2	ma'	nokia	A	+
3	ma'	nokia	N	-
4	nema'	nokia	A	+

2

^EBLUTNIS

ma'	nema'	spolu
1+ 1-	1+ 1-	2+ 2-
$H(ma') = 1$	$H(nema') = 1$	$H(S) = 1$
$P(ma') = 1/2$	$P(nema') = 1/2$	
$H(S, BLUTNIS) = 1$		$I(BLUTNIS) = 0$

$$H(S) = -\frac{2}{4} \log_2 \frac{2}{4} - \frac{2}{4} \log_2 \frac{2}{4} = -2 \left(\frac{1}{2} \log_2 \frac{1}{2} \right) = -(-1) = 1$$

$$H(ma') = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$H(nema') = -\frac{1}{2} \log_2 \frac{1}{2} - \frac{1}{2} \log_2 \frac{1}{2} = 1$$

$$H(S, BLUTNIS) = 0,5 \cdot 1 + 0,5 \cdot 1 = 1 \Rightarrow I(BLUTNIS) = 1 - 1 = 0$$

FOTOGRAFUJE

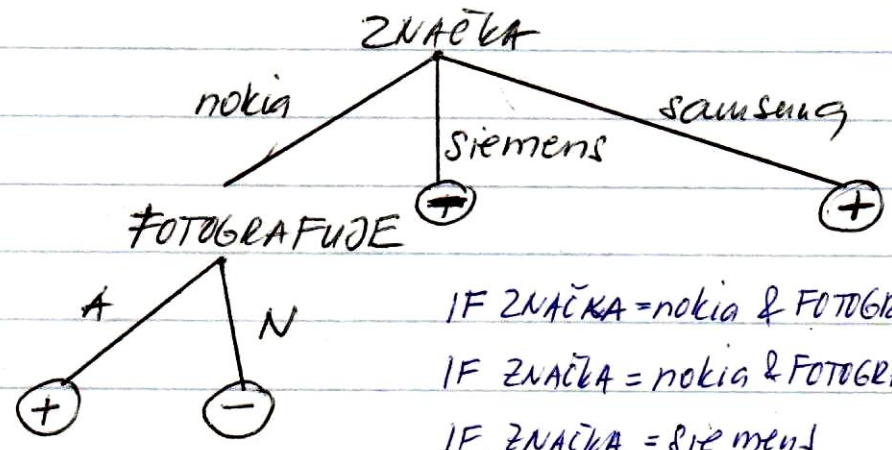
A	N	spolu
2+ 0-	0+ 2-	2+ 2-
$H(A) = 0$	$H(N) = 0$	$H(S) = 1$
$P(A) = 0,5$	$P(N) = 0,5$	
$H(S, FOTOGRAFUJE) = 0$		$I(FOT.) = 1$

$$H(A) = -\frac{2}{2} \log_2 \frac{2}{2} - \frac{0}{2} \log_2 \frac{0}{2} = 0$$

$$H(N) = -\frac{0}{2} \log_2 \frac{0}{2} - \frac{2}{2} \log_2 \frac{2}{2} = 0$$

$$H(S, FOTOGRAFUJE) = 0,5 \cdot 0 + 0,5 \cdot 0 = 0$$

$$I(FOTOGRAFUJE) = 1 - 0 = 1$$



- IF ZNAČKA = nokia & FOTOGRAFUJE = A THEN (+)
- IF ZNAČKA = nokia & FOTOGRAFUJE = N THEN (-)
- IF ZNAČKA = siemens THEN (-)
- IF ZNAČKA = samsung THEN (+)

3