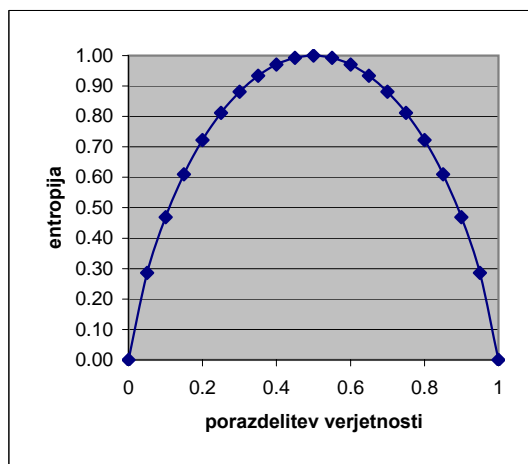


## Entropija in informacijski pridobitek (Information Gain)

verjetnost razreda 1	verjetnost razreda 2	entropija $E(p_1, p_2) = -p_1 \cdot \log_2(p_1) - p_2 \cdot \log_2(p_2)$
$p_1$	$p_2 = 1 - p_1$	
0	1	0.00
0.05	0.95	0.29
0.10	0.90	0.47
0.15	0.85	0.61
0.20	0.80	0.72
0.25	0.75	0.81
0.30	0.70	0.88
0.35	0.65	0.93
0.40	0.60	0.97
0.45	0.55	0.99
0.50	0.50	1.00
0.55	0.45	0.99
0.60	0.40	0.97
0.65	0.35	0.93
0.70	0.30	0.88
0.75	0.25	0.81
0.80	0.20	0.72
0.85	0.15	0.61
0.90	0.10	0.47
0.95	0.05	0.29
1	0	0.00



$$Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot E(S_v)$$

atribut  $\rightarrow$   $A$   
 množica  $\rightarrow$   $S$   
 število primerov v podmnožici  $\rightarrow$   $|S_v|$   
 verjetnost "veje"  $\rightarrow$   $\frac{|S_v|}{|S|}$   
 število primerov v množici  $\rightarrow$   $|S|$

### Učna množica

Oseba	Starost	Dioptrija	Astigmatizem	Solzenje	Leče
O1	mlad	kratko	ne	normalno	DA
O2	mlad	kratko	ne	zmanjšano	NE
O4	mlad	daleko	ne	zmanjšano	NE
O5	mlad	kratko	da	normalno	DA
O6	mlad	kratko	da	zmanjšano	NE
O7	mlad	daleko	da	normalno	DA
O8	mlad	daleko	da	zmanjšano	NE
O10	pr_st_dal	kratko	ne	zmanjšano	NE
O11	pr_st_dal	daleko	ne	normalno	DA
O14	pr_st_dal	kratko	da	zmanjšano	NE
O17	st_daleko	kratko	ne	normalno	NE
O18	st_daleko	kratko	ne	zmanjšano	NE
O19	st_daleko	daleko	ne	normalno	DA
O20	st_daleko	daleko	ne	zmanjšano	NE
O21	st_daleko	kratko	da	normalno	DA
O22	st_daleko	kratko	da	zmanjšano	NE
O24	st_daleko	daleko	da	zmanjšano	NE

### Testna množica

Oseba	Starost	Dioptrija	Astigmatizem	Solzenje	Leče
O3	mlad	daleko	ne	normalno	DA
O9	pr_st_dal	kratko	ne	normalno	DA
O12	pr_st_dal	daleko	ne	zmanjšano	NE
O13	pr_st_dal	kratko	da	normalno	DA
O15	pr_st_dal	daleko	da	normalno	NE
O16	pr_st_dal	daleko	da	zmanjšano	NE
O23	st_daleko	daleko	da	normalno	NE

Petra Kralj

[Petra.Kralj@ijs.si](mailto:Petra.Kralj@ijs.si)

<http://kt.ijs.si/PetraKralj/UNGKnowledgeDiscovery0607.html>