

1 V Soči so ribiči po prvi svetovni vojni naselili potočno postrv, ki je skoraj izpodrinila soško. Da bi soško postrv rešili pred izumrtjem, jo razmnožujejo v ribogojnicah in vračajo v reko, potočne postrvi pa lovijo in izločajo. Na pogled je med vrstama postrvi pogosto težko razlikovati, vrsto lahko zanesljivo določijo le na podlagi genskih testov. Denimo, da pri tem opazujejo gene A, B, C in D. Vrsto tedaj določijo s pomočjo tabele na desni.

A	B	C	D	vrsta
0	0	0	0	potočna
0	0	0	1	potočna
0	1	1	0	potočna
0	1	1	1	potočna
0	2	1	0	potočna
0	2	0	1	soška
1	0	0	0	potočna
1	0	0	1	potočna
1	1	1	0	soška
1	2	0	0	potočna
1	2	1	1	soška

- a) Zgradite klasifikacijsko drevo za te podatke, tako da kriterije izbirate glede na njihov informacijski prispevek!
- b) Porežite drevo s pomočjo m-ocene napake ($m=2$)!

2 Da bi preverili, v kolikšni meri drži teza, da mladenk take krvi res ni, pa bomo storili tole. Zbrali smo podatke o 93 mladenkah, med njimi 52 slovenske in 41 madžarske krvi. V vzorcih krvi smo testirali prisotnost genov A, B in C. Rezultate kaže tabela na levi. Vsi trije geni, denimo, so bili prisotni pri dvanajstih mladenkah, ki so bile vse Slovenke. Gen C je manjkal pri devet Slovenkah in šest Madžarkah...

5. *Bog živi vas, Slovenke,
prelepe, žlahtne rožice!
Ni take je mladenke,
ko naše je krvi dekle;
naj sinov zarod nov
iz vas bo strah
sovražnikov!*

- a) Iz danih podatkov sestavite klasifikacijsko drevo, globoko le dva nivoja (torej tako, da ima koren in potem kvečjemu še po eno vozlišče na vsaki veji).
- b) Ker vozlišča niso čista – v vsakem vozlišču bo nekaj slovenskih in nekaj madžarskih mladenk –, lahko za vsako vozlišče izračunamo preostalo entropijo. To ste pravzaprav že izračunali med gradnjo modela. Pripisite jo k listom. Zdaj pa izračunajte pričakovano preostalo entropijo: to dobimo tako, da entropijo v vsakem listu pomnožimo z verjetnostjo, da bo neka naključno izbrana mladenka klasificirana v ta list. (Verjetnost je torej enaka številu mladenk v listu, deljenemu s številom vseh mladenk.)
- c) Kaj nam pove pričakovana preostala entropija in kako je povezana s tezo, ki jo preverjamo?

A	B	C	razmerje
✓	✓	✓	12:0
✓	✓	×	9:6
✓	×	✓	12:1
✓	×	×	5:5
×	✓	✓	7:6
×	✓	×	1:9
×	×	✓	4:6
×	×	×	2:8

3 Natakariji v nekem lokalnu se zabavajo tako, da stavijo, kakšno pizzo bo naročil kateri gost. Eden od njih si bo izboljšal možnosti zadetka s pomočjo klasifikacijskega drevesa. Podatke, ki jih kani uporabiti, je zbral v tabeli.

spol	starost	sam(a)?	tip pizze
Ž	mlada	da	bučna
Ž	stara	ne	klasika
Ž	mlada	ne	bučna
Ž	stara	da	klasika
M	mlad	ne	klasika
M	star	ne	klasika
M	mlad	da	mesna
M	star	da	mesna

Sestavite klasifikacijsko drevo tako, da v vsakem koraku izberete atribut z največjim informacijskim prispevkom. V vsakem vozlišču pripisite tudi, za koliko zmanjša pričakovano entropijo.

4 Za uspešno pridelavo solate je ključnega pomena preventivno škropljenje pred njeno najpogostejšo boleznijo, listno plesnijo. Večletna opazovanja smo strnili v tabeli na desni (N pomeni nizko in V visoko vrednost). Tako smo, denimo, štirinajstkrat zapazili nizko sklenjenost glav, visoko zračno vlago in nizko temperaturo; pri tem so glave dvanajstkrat ostale zdrave, dvakrat pa so začele plesniti.

sklenjenost glav	zračna vlaga	temperatura	plesen (NE:DA)
N	N	N	10:0
N	N	V	8:2
N	V	N	12:2
N	V	V	3:7
V	N	N	11:3
V	N	V	4:8
V	V	N	8:4
V	V	V	2:7

Sestavite klasifikacijsko drevo za napovedovanje (verjetnost pojavitve) plesni!

5 Pa pogledjmo. Čisto res tole ni. Urško so do sedme ure videli plesati štirikrat, sedem prosilcev pa je zavrnila. Mestne čenče so vestno zapisovale značilnosti prosilcev, ki so se jim zdele pomembne: ali so imeli v gumbnici rožo, ali jim je moško glavo krasil nov klobuk in ali so ta dan že kdaj plesali.

5. *Jih dokaj jo prosil, al vsakmu odreče,
prešerna se brani in ples odlašuje,
si vedno izgovore nove zmišljuje,
že sonce zahaja, se mrak približuje,
že sedem odbila je ura in čez,
ko jela ravnat se je Urška na ples*

a) Iz podatkov sestavite klasifikacijsko drevo.

roža	nov klobuk	že plesal	gre plesat
✓	✓	✓	✓
✓	✓		✓
✓	✓	✓	✓
✓		✓	
✓		✓	
✓		✓	
	✓	✓	✓
	✓	✓	
	✓	✓	
		✓	

Opomba: gornje naloge so izpitne naloge s predmeta Odločitveni sistemi (VSŠ Informatika, 3. letnik).

Naloge so iztrgane iz konteksta (izpiti pri tem predmetu imajo navadno rdečo nit) in tudi nekoliko spremenjene (določena podvprašanja so odstranjena, v prvi nalogi je eno vprašanje dodatno).