

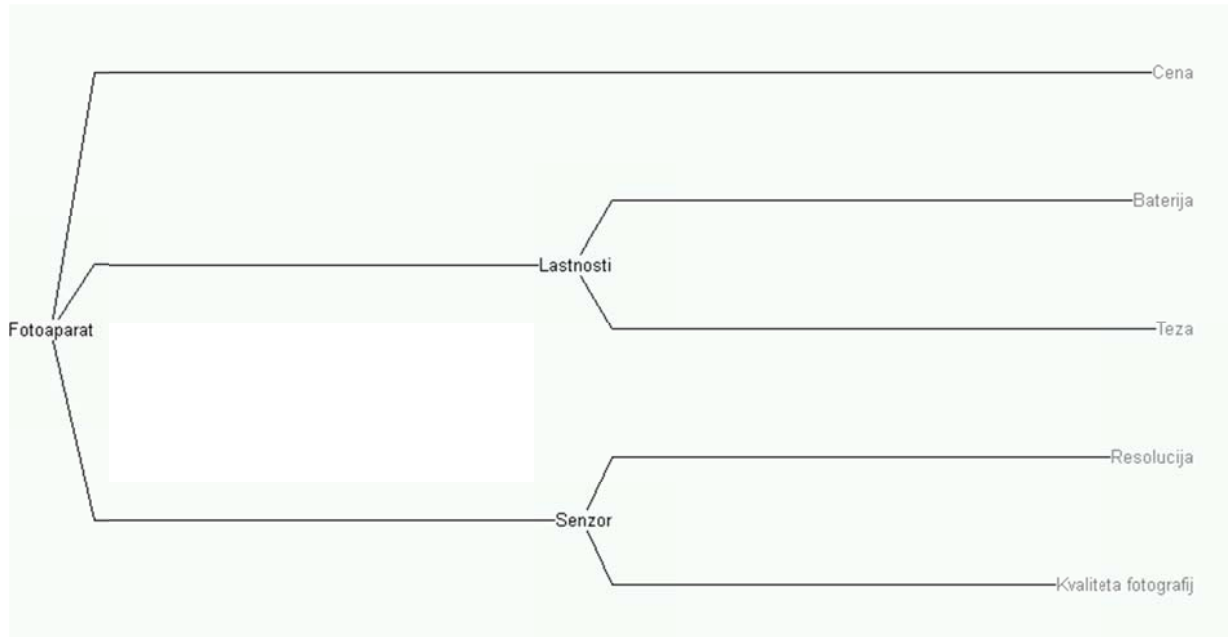
Izpit pri predmetu Odločitveni sistemi (5.2.2014)

Navodila:

Izpit se piše 90 minut. Literatura ni dovoljena. Vrednost naloge je označena pri vsaki nalogi.

1. (30 točk) Izbira fotoaparata

Izbiramo fotoaparat na podlagi 5-ih osnovnih kriterijev, ki smo jih združili v odločitvenem drevesu:



Uteži po posameznih vejah so:

- Fotoaparati: cena(35), Lastnosti(15), Senzor(50)
- Lastnosti: Baterija (80), Teža (20)
- Senzor: Kvaliteta (70), Resolucija (30)

	Teža (g)	Baterija	Cena	Resolucija	Kvaliteta fot.
A (Nikon D5300)	450	600	800	25	80
B (Nikon D5200)	550	500	600	20	80
C (Canon 70D)	750	900	900	25	70
D (Canon 7D)	850	800	1000	20	70

V tabeli so predstavljene vrednosti za izbrane fotoaparate (vrednosti so zaradi lažjega računanja zelo približne – lahko celo napačne). Teža je v gramih (manj je boljše), baterija v številu narejenih posnetkov (več boljše), cena je v evrih, resolucija v megapikslih in kvaliteta je ocena iz laboratorija (več je boljše). Koristnosti kriterijev izračunajte relativno, z uporabo linearne premice.

Naloge:

- a) (5 točk) Izločite manjvredne variante.

Manjvredna varianta je D.

- b) (5 točk) Variante razvrstite od najboljše do najslabše z maxi-min metodo.

Koristnosti:

	Teža (g)	Baterija	Cena	Resolucija	Kvaliteta fot.
	0.03	0.12	0.35	0.15	0.35
A (Nikon D5300)	100	25	50	100	100
B (Nikon D5200)	75	0	100	0	100
C (Canon 70D)	25	100	25	100	0
D (Canon 7D)	0	75	0	0	0

A, C, B, D

- c) (10 točk) Variante razvrstite z linearno metodo.

$$v(A) = 0.03 * 100 + 0.12 * 25 + 0.35 * 50 + 0.15 * 100 + 0.35 * 100 = 73.5$$

$$v(B) = 0.03 * 75 + 0.12 * 0 + 0.35 * 100 + 0.15 * 0 + 0.35 * 100 = 72.25$$

$$v(C) = 0.03 * 25 + 0.12 * 100 + 0.35 * 25 + 0.15 * 100 + 0.35 * 0 = 36.5$$

- d) (10 točk) Spodaj imamo sliko analize občutljivosti (ni izračunana iz zgornje tabele, ampak iz pravih podatkov; dodan je še en fotoaparat).

- a. Kateri je najboljši fotoaparat?

Če je utež za ceno manjša od 30, potem je najboljši Nikon D7100, drugače pa D5200.

- b. Pri kateri vrednosti uteži je najboljši Nikon 5300?

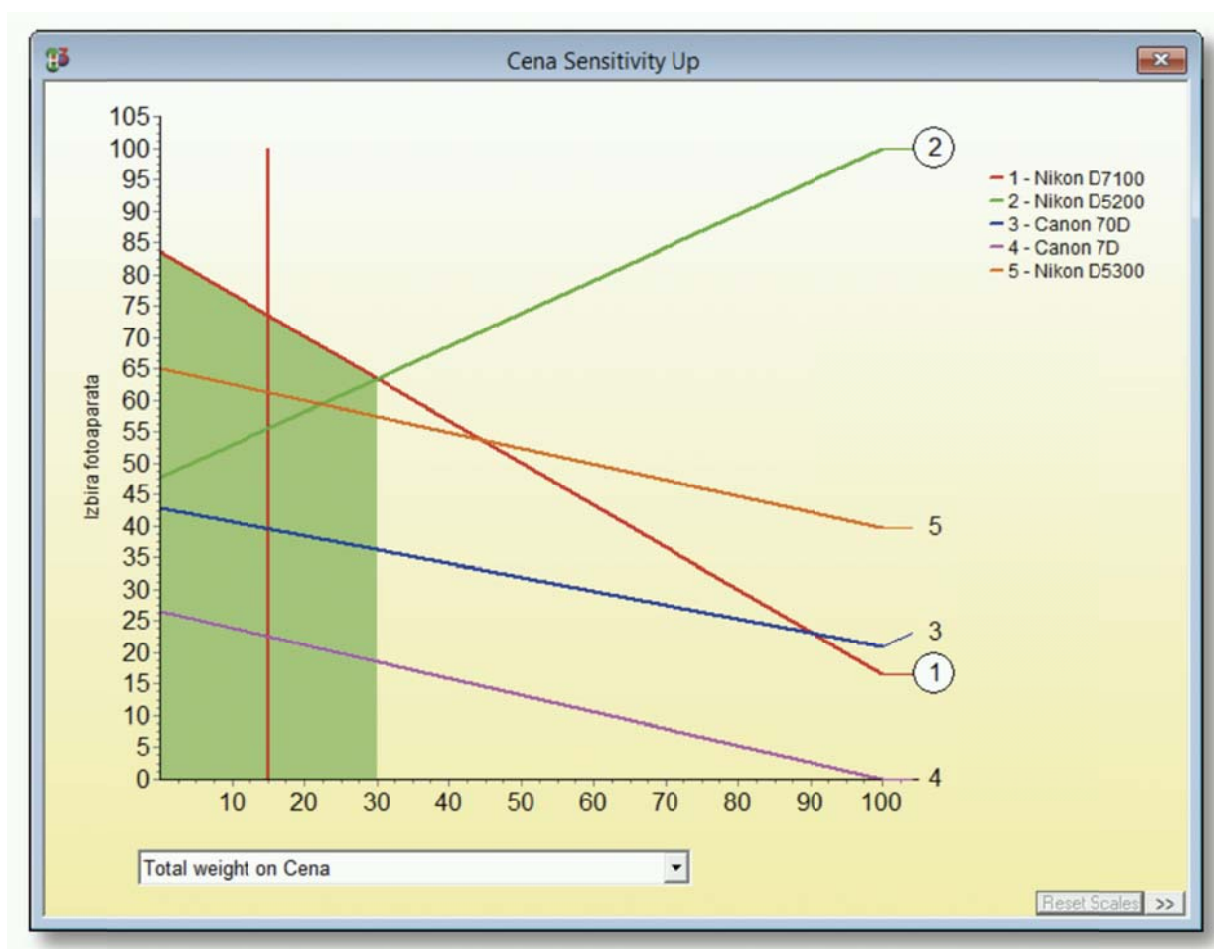
Pri nobeni.

- c. Od katerih fotoaparátov je Canon 70D boljši?

Od Canon 7D je vedno boljši, od Nikona D7100 pa le, če je utež za ceno zelo visoka > 90 .

- d. Če bi se Nikon D7100 pocenil, kako bi se spremenila krivulja?

Desna stran premice bi se dvignila, leva stran bi ostala na istem mestu (cca 84).

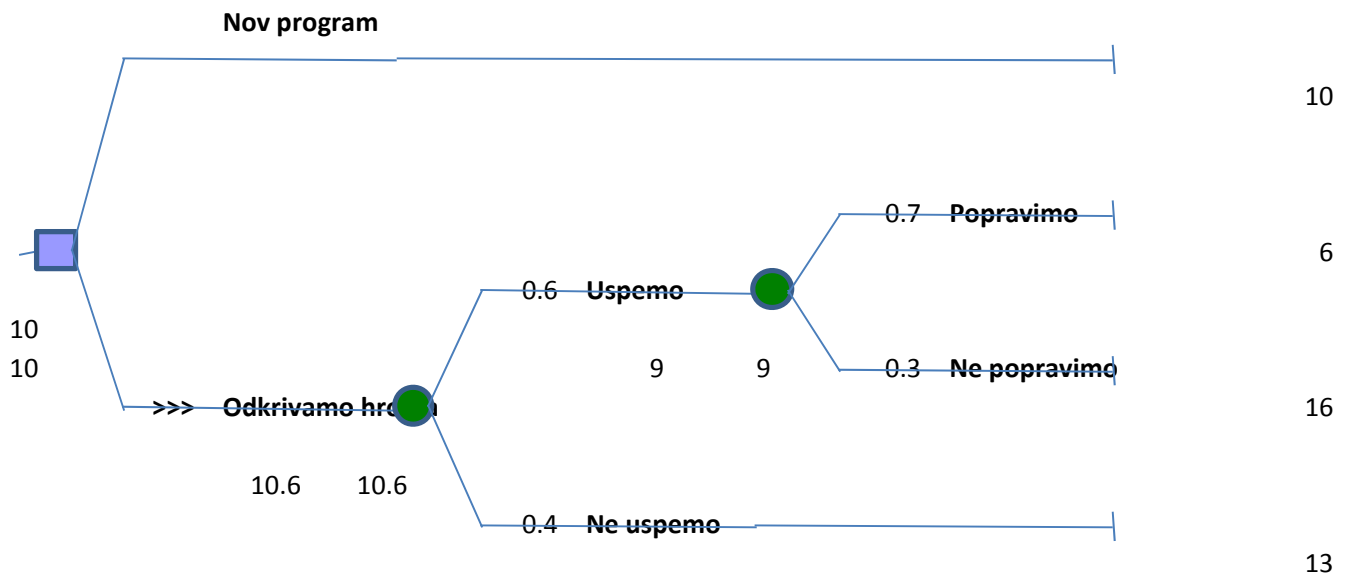


2. (25 točk) Težave s hrošči

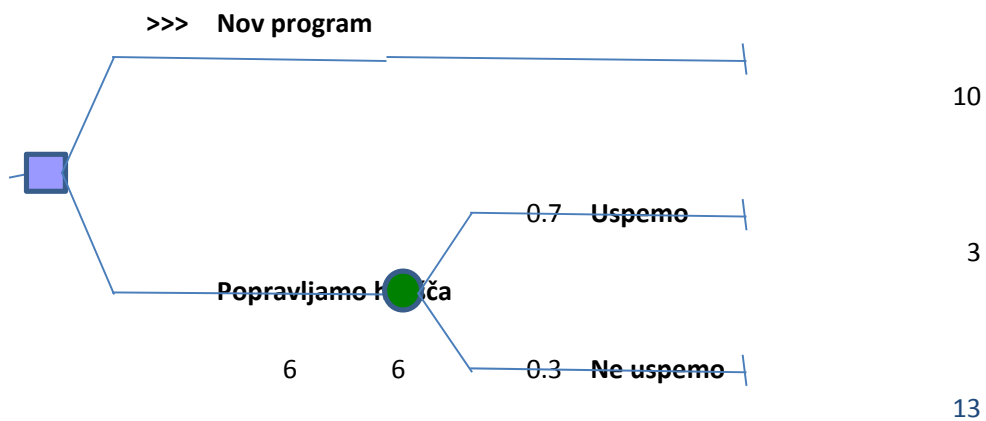
V novi službi imamo precejšno nesrečo; dobili smo program od našega predhodnika in ta program ima hrošča. Ta hrošč je precej izmuzljiv, ni ga mogoče enostavno odkriti oz. popraviti, verjetno bomo potrebovali kar nekaj časa. Imamo dve možnosti: a) ali se odločimo, da napišemo celotni program na novo, kar bi nam vzelo 10 dni dela, ali b) poskusimo najprej odkriti hrošča (3dni dela, verjetnost, da odkrijemo je 60%) in ga potem popravimo (3 dni dela z verjetnostjo uspeha 70%) oz. če ne uspe, napišemo celotni program na novo.

- a) (10 točk) Katera možnost je boljša? Narišite ustrezni model.

Pričakovan čas dela, če odkrivamo hrošča, je 10.6 dni. Zato je bolje takoj začeti z razvojem novega programa.



- b) (5 točk) Koliko smo pripravljeni plačati našemu predhodniku, da odkrije hrošča? Ker je njegov program, predpostavimo, da najde hrošča v trenutku. Cena našega dela je 100EUR/dan.



Pričakovan čas je tokrat 6 dni. Glede na prejšno možnost (10 dni), smo pridobili 4 dni, torej smo pripravljeni plačati največ 400 EUR.

- c) (10 točk) Predpostavka v točki a) je precej nerealna. Pogosto niti sam avtor ne uspe najti hrošča. Poznamo:

- a. $P(\text{predhodnik odkrije} \mid \text{hrošč rešljiv}) = 0.9$
- b. $P(\text{predhodnik odkrije} \mid \text{hrošč ni rešljiv}) = 0.3$
- c. $P(\text{hrošč rešljiv}) = 0.7$

Kakšna je verjetnost, da bo predhodnik odkril hrošča? Kakšna je verjetnost, da bo hrošč rešljiv, če predhodnik odkrije hrošča?

Ena možna rešitev je, da obrnemo drevo. Lahko pa kar izračunamo z verjetnostnim računom:

$$P(\text{predhodnik odkrije}) = 0.9 * 0.7 + 0.3 * 0.3 = 0.72$$

$$P(\text{hrošč rešljiv} \mid \text{predhodnik odkrije}) = 0.7 * 0.9 / 0.72 = 0.875$$

3. (35 točk) Izbira nadaljevanke

	Igra prestolov	Sherlock	Prijatelji	Seks v mestu	Prava kri
uporabnik A	5	4	3	?	4
uporabnik B	3	?	2	1	2
uporabnik C	1	1	4	3	1
uporabnik D	4	4	3	5	5

- (a) (15 točk) Na podlagi »user-based collaborative filtering« tehnike napovejte kakšno oceno bi uporabnik A dal nadaljevanke Seks v mestu. Za mero podobnosti uporabite Pearsonov količnik. Upoštevajte vse uporabnike z nenegativnim količnikom.

Pearsonov količnik za podobnost med dvema uporabnika (a in b):

$$\text{similarity}(a, b) = \frac{\sum_p (r_{a,p} - \bar{r}_a) \cdot (r_{b,p} - \bar{r}_b)}{\sqrt{\sum_p (r_{a,p} - \bar{r}_a)^2} \cdot \sqrt{\sum_p (r_{b,p} - \bar{r}_b)^2}}$$

$$\text{sim}(A, B) = 1 / \sqrt{2} = 0.71$$

$$\text{sim}(A, C) < 0$$

$$\text{sim}(A, D) = (1 * (-0.2) + (-1) * (-1.2)) / \sqrt{2} / \sqrt{0.04 + 0.04 + 1.44 + 0.64} = 0.48$$

$$\text{napoved}(A, \text{Seks v mestu}) = 4 + (0.71 * (-1) + 0.48 * 0.8) / (0.71 + 0.48) = 3.72$$

	Igra prestolov	Sherlock	Prijatelji	Seks v mestu	Prava kri
uporabnik A	1	1	0	0	1
uporabnik B	1	0	1	0	1
uporabnik C	0	0	1	1	0
uporabnik D	1	1	0	1	1

(b) (20 točk) V tabeli smo označili nadaljevanke, ki so uporabnikom všeč (npr. so si jih ogledali) .

Vrednost 0 pomeni, da jim niso všeč (si niso ogledali).

- a. Izračunajte podporo (support) , zaupanje (confidence) in dvig (lift) povezovalnega pravila
IF Prava kri THEN Sherlock

$$\text{support} = 2/4$$

$$\text{confidence} = 2/3$$

$$\text{lift} = (2/3) / (1/2) = 4/3$$

- b. Kakšna je Jaccardova mera podobnosti med nadaljevankama Prijatelji in Seks v mestu?
1/3

- c. Kateri nadaljevanki sta si **najbolj** in kateri **najmanj** podobni glede na Jaccardovo mero podobnosti?

Najmanj: Sherlock in Prijatelji (J=0)

Najbolj: Igra prestolov in Prava kri (J=1)

- d. Izpišite vsa povezovalna pravila, katerih podpora je vsaj 0.5, zaupanje vsaj 0.75 in omenjajo največ dve nadaljevanki (povezovalnih pravil s tremi nadaljevankami torej ne pišite).

IF Igra Prestolov THEN Prava Kri

IF Prava Kri THEN Igra Prestolov

IF Sherlock THEN Prava kri

IF Sherlock THEN Igra Prestolov

4. (10 točk) Uporaba vsebine pri priporočanju nadaljevank

Za nadaljevanke iz prejšnjega vprašanja imamo na voljo precej tekstovnih vsebin, npr. opisov, kritik, itd. Radi bi uporabili naivni Bayesov klasifikator, ki bi pri priporočanju uporabljal to tekstovno vsebino. Kako bi se tega lotili? Kako bi določili attribute? Kako bi pripravili podatke?