

Zadatak za pripremu II kolokvijuma

Na tender za izgradnju stanova javilo se nekoliko ponuđača, sa sledećim karakteristikama:

	Brzina [zgrada/god] (max)	Kvalitet [1-10] (max)	Cena [mil. Rsd] (min)	RS Radnici [%] (max)	Pouzdanost [1-5] (max)
sBrda	3	7	80	50	5
Mrka	6	4	60	20	3
Blok70	2	8	80	70	3
naVodi	1	10	100	90	5
Z	7	8	80	30	1

1. Ekspertski tim sastavljen od Milana, Jelene i Đorđa, glasaju za najboljeg ponuđača (prikazati rezultat glasanja). Nakon toga, treba oceniti prosečan rang svakog ponuđača, kao i usklađenost mišljenja grupe. Prokomentarisati rezultate. Poznato je sledeće:
 - a. Đorđe je primetio da je kvalitet ponuđača varijabilan, te je ocenio da je standardna devijacija (normalne raspodele) za kvalitet ponuđača: 2, 4, 1, 0.5 i 3. On želi da bude 90% siguran da ukupna korist neće biti manja od procenjene.
 - b. Milan želi da izabere ponuđača kod koga su izbalansirani kriterijumi, tj. kod koga nema osobina koje su jako loše.
 - c. Za Jelenu, najbolji ponuđač je za nju onaj koji ima ukupno najbolje osobine.
 - d. Tumačenjem zakonskih akata od strane grupe, utvrđena je sledeća važnost kriterijuma:

	Brzina [zgrada/god] (max)	Kvalitet [1-10] (max)	Cena [mil. Rsd] (min)	RS Radnici [%] (max)	Pouzdanost [1-5] (max)
Težina	5	7	10	8	4

2. Komisija nije bila u potpunosti zadovoljna prethodnim rešenjem i smatra da je problem odlučivanja mogao da se bolje strukturise. U tom cilju, potrebno je analizirati atribute i predložiti bolju strukturu problema odlučivanja (i prokomentarisati rezultate), tako da:
 - a. Se maksimalno smanji broj zavisnih (korelisanih) atributa.
 - b. Se, ručno, napravi hijerarhija kategoričkih atributa (za diskretizaciju koristiti aritmetičku sredinu i standardnu devijaciju, 3 kategorije). Prikazati konačnu skalu za atribut Cena.

- c. Nad diskretizovanim atributima alternativa iz prethodnog koraka (pre kreiranja hijerarhije), indukovati stablo odlučivanja, ako je poznato da li je na prethodnim konkursima ponuđač već bio biran:

	BiranRanije?
sBrda	Ne
Mrka	Ne
Blok70	Da
naVodi	Da
Z	Ne

3. Prethodni tender je propao, te se pristupilo licitaciji. U postupku licitacije cena nije poznata, ali svi drugi kriterijumi jesu i isti su kao u polaznom slučaju.

- a. Ako je poznato da kvalitet gradnje, od svih kriterijuma, najviše diktira cenu, napraviti model iz podataka koji će dati procenu cene svakog od ponuđača, koristeći informacije o ceni sa prethodnog tendera. Ispitati kvalitet modela i prokomentarisati.
- b. Na licitaciju se javio novi ponuđač. Potrebno je, proceniti koliku ponudu će dati u odnosu na ukupno najbližeg ponuđača ako su poznate njegove karakteristike:

	Brzina [zgrada/god] (max)	Kvalitet [1-10] (max)	RS Radnici [%] (max)	Pouzdanost [1-5] (max)	Cena
NN	3	4	80	4	?
Težine	4	8	9	6	-

Rešenje zadatka

1. Da bi se zadatak uspešno rešio, potrebno je, pravilnim redosledom (unazad) rešiti potprobleme. Prvo je potrebno odrediti pondera (rešiti zadatak pod D):

d) Da bi se od težine stiglo do pondera, potrebno je koristiti L1 normu nad datim težinama:

Težine	5	7	10	8	4	SUMA: 34
Ponderi	0,147	0,206	0,294	0,235	0,118	SUMA: 1

Dalje, zadatak rešavamo zasebno za svaki od problema od c) do a). Svako od pojedinačnih rešenja ovih potproblema predstavlja rešenje odgovarajućeg donosioca odluke. Na kraju, rešenja metoda pomirićemo metodama grupnog odlučivanja.

c) Invertujemo *Cenu* kako bi svi kriterijumi bili tipa maksimizacije.

	Brzina	Kvalitet	Cena	RS Radnici	Pouzdanost
sBrda	3	7	0,013	50	5
Mrka	6	4	0,017	20	3
Blok70	2	8	0,013	70	3
naVodi	1	10	0,010	90	5
Z	7	8	0,013	30	1

Normalizujemo matricu L^∞ normom. Računamo očekivanu korist koristeći pondera iz rešenja pod d).

	Brzina	Kvalitet	Cena	RS Radnici	Pouzdanost	JAT
sBrda	0,429	0,700	0,750	0,556	1,000	0,676
Mrka	0,857	0,400	1,000	0,222	0,600	0,625
Blok70	0,286	0,800	0,750	0,778	0,600	0,681
naVodi	0,143	1,000	0,600	1,000	1,000	0,756
Z	1,000	0,800	0,750	0,333	0,200	0,634

Zaključujemo da je za Jelenu najprihvatljivija alternativa *naVodi*.

b) Nad matricom iz prethodnog koraka sprovodimo MAXIMIN metodu i dobijamo:

	MAXIMIN
sBrda	0,429
Mrka	0,222
Blok70	0,286
naVodi	0,143
Z	0,200

Zaključujemo da je, prema Milanovom mišljenju, najprihvatljivija alternativa *sBrda*.

- a) Iz polazne matrice, računamo rizik za kriterijum *Kvalitet*. Od rešenja pod c) oduzimamo $1,3 * \text{Rizik za } Kvalitet$ kako bismo dobili ukupan rizik. Dobijamo:

	Rizik Kvalitet	OK (izračunato pod c)	Ukupan Rizik (OK-1.3*Rizik Kvalitet)
sBrda	0,041	0,676	0,623
Mrka	0,082	0,625	0,518
Blok70	0,021	0,681	0,654
naVodi	0,010	0,756	0,743
Z	0,062	0,634	0,554

Zaključujemo da je, po Dorđevom mišljenju, najprihvatljivija alternativa *naVodi*.

Na kraju, potrebno je oceniti prosečan rang svakog ponuđača, kao i usklađenost mišljenja grupe. Kreiramo matricu rangova svakog DO-a i koristimo modele aditivnog rangiranja i minimalne varijanse.

	Djordje	Milan	Jelena	AS	VAR
sBrda	3	1	3	2,333	1,333
Mrka	5	3	5	4,333	1,333
Blok70	2	2	2	2,000	0,000
naVodi	1	5	1	2,333	5,333
Z	4	4	4	4,000	0,000

Zaključujemo da je alternativa *Blok70* ukupno najprihvatljivija jer, u proseku, ima najviši rang od svih alternativa. Takođe, neslaganje po pitanju ranga ove alternative ne postoji, jer je uvek bila na tom rangu kod svakog DO-a. Napomena: ovde se uočavaju neki od nedostataka metoda. Ako se pogleda da je *naVodi* u 2 od 3 slučaja bila prvorangirana, a samo je Milan drastično loše ocenio, te time „nepravredno pogoršao“ prosek i uneo ogromno neslaganje u grupu usled visoke varijanse, bilo bi prikladnije da se koristi model Umnoženog rangiranja kako bi se eliminisao negativan uticaj ovog DO.

2. Pristupamo rešavanju potproblema:

- a) Kako bismo ispitali korelisanost atributa, potrebne su nam matrice kovarijacije i korelacije:

cov	Brzina	Kvalitet	Cena	RS Radnici	Pouzdanost
Brzina	5,36	-2,72	-20	-55,6	-2,72
Kvalitet	-2,72	3,84	24	41,2	0,64
Cena	-20	24	160	280	8
RS Radnici	-55,6	41,2	280	656	23,2
Pouzdanost	-2,72	0,64	8	23,2	2,24

Matrica kovarijacije

cor	Brzina	Kvalitet	Cena	RS Radnici	Pouzdanost
Brzina	1	-0,60	-0,68	-0,94	-0,78
Kvalitet	-0,60	1	0,97	0,82	0,22
Cena	-0,68	0,97	1	0,86	0,42
RS Radnici	-0,94	0,82	0,86	1	0,61
Pouzdanost	-0,78	0,22	0,42	0,61	1

Matrica korelacije

Posmatrajući zavisnosti između atributa zaključujemo sledeće:

- Između kriterijuma *Brzina* i *RS Radnici* postoji visoka negativna korelisanost, što znači da će veći broj domaćih radnika na gradilištu gotovo uvek dovesti do sporije gradnje, što je loše za odabir alternative (jer generalno se želi što više domaćih radnika na gradilištu). Stoga je potrebno zadržati ovu informaciju, a ovaj par kriterijuma nije dobro grupisati zajedno ili izbaciti iz posmatranja jer bi se izgubila značajna informacija za donošenje odluke.
- Sa druge strane, između kriterijuma *Cena* i *Kvalitet*, *Cena* i *RS radnici* i *Cena* i *Kvalitet* i *RS radnici* nam govori da ove tri veličine međusobno daju istu informaciju, jer su visoko pozitivno korelisane, te bi se npr. *Kvalitet* mogao izbaciti iz posmatranja, jer bi se skoro u potpunosti mogao objasniti preko *Cene*. Ovde se potvrđuje prethodni zaključak da ne bi bilo dobro izbaciti *RS radnike* iz posmatranja, jer bi se izgubila važna informacija koja ide u korist izbora, tj. govori da domaći radnici uglavnom pružaju viši kvalitet gradnje.
- Ako bi bilo potrebno da se grupišu neki kriterijumi u hijerarhiji, prema prethodnoj logici, dobar izbor bi bila grupa: *Brzina*, *Kvalitet* i *Pouzdanost*. Ovi kriterijumi bi se mogli opisati preko kriterijuma-roditelja: *Performanse*, a da se pritom ne izgubi informacija koju nose dok se dimenzije problema na jednom nivou smanje.

b) Potrebno je kreirati DEX model (prateći smernice analize iz prethodnog koraka), uz prethodnu diskretizaciju tabele odlučivanja. Prvo vršimo diskretizaciju, koristeći pristup sa srednjom vrednošću i standardnom devijacijom. Parametar n je 1, kako bi se dobile tri grupe:

	Brzina	Kvalitet	Cena	RS Radnici	Pouzdanost
m	3,8	7,4	80	52	3,4
sd	2,32	1,96	12,65	25,61	1,50
m-1*sd	1,48	5,44	67,35	26,39	1,90
m+1*sd	6,12	9,36	92,65	77,61	4,90

Grupe sada pripadaju sledećim intervalima:

I	$X < m-1*sd$	Niska/Nizak
II	$m-1*sd \leq X \leq m+1*sd$	Srednja/Srednji
III	$X > m+1*sd$	Visoka/Visok

Konačno, kada zamenimo vrednosti, dobijamo sledeću tabelu:

	Brzina	Kvalitet	Cena	RS Radnici	Pouzdanost
sBrda	Srednja	Srednji	Srednja	Srednji	Visoka
Mrka	Srednja	Nizak	Niska	Nizak	Srednja
Blok70	Srednja	Srednji	Srednja	Srednji	Srednja
naVodi	Niska	Visok	Visoka	Visok	Visoka
Z	Visoka	Srednji	Srednja	Srednji	Niska

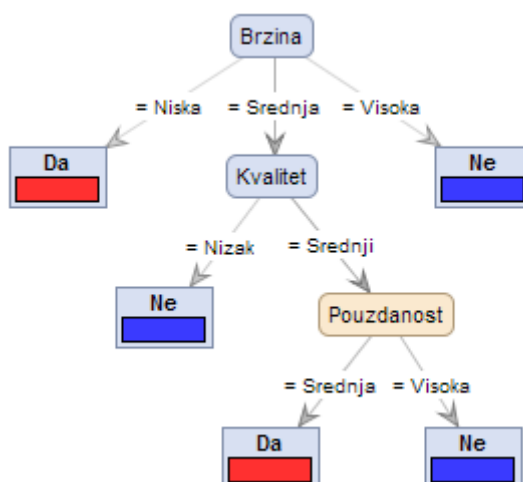
Na kraju, pravimo DEX model:

Atribut	Skala
GradnjaStana	NE; DA
└ PerformanseGradnje	Niske ; Srednje; Visoke
└└ Brzina	Niska ; Srednja; Visoka
└└ Kvalitet	Nizak ; Srednji; Visok
└└ Cena	Visoka ; Srednja; Niska
└ RSRadnici(DomRadSnaga)	Nizak ; Srednji; Visok
└ Pouzdanost	Niska ; Srednja; Visoka

Važno je obratiti pažnju da je *Cena* po značenju tipa minimizacije i da je potrebno da se raspored vrednosti prilagodi rastućoj orijentaciji DEX skale.

- c) Tabeli iz prethodnog koraka pridodajemo informaciju o prethodnim izborima izvođača i indukujemo ID3 stablo:

	Brzina	Kvalitet	Cena	RS Radnici	Pouzdanost	BiranRanije?
sBrda	Srednja	Srednji	Srednja	Srednji	Visoka	Ne
Mrka	Srednja	Nizak	Niska	Nizak	Srednja	Ne
Blok70	Srednja	Srednji	Srednja	Srednji	Srednja	Da
naVodi	Niska	Visok	Visoka	Visok	Visoka	Da
Z	Visoka	Srednji	Srednja	Srednji	Niska	Ne



3. U ovom zadatku potrebno je da donesemo odluke iz podataka.

a) Potrebno je da napravimo model linearne regresije koji će u zavisnosti od kvaliteta gradnje predviđati vrednost ponuđene cene u milionima dinara:

Prvo računamo koeficijent $b(X)$:

	Kvalitet X	Cena Y	Xc	Yc	Xc*Yc	Xc^2
sBrda	7	80	-0,4	0	0	0,16
Mrka	4	60	-3,4	-20	68	11,56
Blok70	8	80	0,6	0	0	0,36
naVodi	10	100	2,6	20	52	6,76
Z	8	80	0,6	0	0	0,36
AVG	7,4	80			24	3,84

$b(x) = 6,25$. Naš model ima oblik $Y = 6,25 * X$. Dakle, za jediničnu promenu kvaliteta, cena, u milionima dinara, će se povećati 6,25 puta.

Dalje, predviđamo vrednosti ovim modelom da vidimo koliko greši.

	b(X)*X	Y-b(X)*X	Y-(a+b(X)*X)	Y-(a+b(X)*X)^2
sBrda	43,75	36,25	2,5	6,25
Mrka	25	35	1,25	1,56
Blok70	50	30	-3,75	14,06
naVodi	62,5	37,5	3,75	14,06
Z	50	30	-3,75	14,06
AVG		33,75	0	

Posmatrajući kolonu $Y-b(X)*X$, tj. grešku inicijalnog modela, primećujemo da u proseku greši (potcenjuje) za 33,75 jedinica. To znači da je potrebno da dodamo konstantu u model kako bismo objasnili tu količinu informacije koja nije objašnjena direktnom vezom cene i kvaliteta.

Stoga je konstanta $a = 33,75$. Konačan model ima oblik: $Y = 6,25 * X + 33,75$

Na kraju, potrebno je da izračunamo mere kvaliteta modela:

SSE	50
RMSE	3,16
R^2	0,69

Zaključujemo da prosečna devijacija procenjenih vrednosti (standardna greška procene) iznosi 3,16 miliona dinara. S druge strane, naš model objašnjava 69% cene.

b) U ovom slučaju *Cena* postaje izlazni atribut, jer želimo da donesemo zaključak o njenoj visini tražeći najbližeg suseda novoj alternativni.

Kako je *Cena* izbačena iz posmatranja, a DO su zadali nove težine za ostale kriterijume, potrebno je da prvo izračunamo pondera koristeći L1 normu:

	Brzina	Kvalitet	RS Radnici	Pouzdanost	
Težine	4	8	9	6	Suma: 27
Ponderi	0,148	0,296	0,333	0,222	Suma: 1

Računamo L^∞ normu na osnovu starih slučajeva:

	Brzina	Kvalitet	RS Radnici	Pouzdanost
MAX	7	10	90	5

Vrednostima iz prethodne tabele normalizujemo stare slučajeve, kao i novi slučaj:

	Brzina	Kvalitet	RS Radnici	Pouzdanost
NN	3	4	80	4
NN normaliz	0,429	0,400	0,889	0,800

	Brzina	Kvalitet	RS Radnici	Pouzdanost
sBrda	0,429	0,700	0,556	1,000
Mrka	0,857	0,400	0,222	0,600
Blok70	0,286	0,800	0,778	0,600
naVodi	0,143	1,000	1,000	1,000
Z	1,000	0,800	0,333	0,200

Računamo udaljenost novog slučaja od starih, primenom Euklidske udaljenosti:

	Brzina	Kvalitet	RS Radnici	Pouzdanost
sBrda	0,000	0,090	0,111	0,040
Mrka	0,184	0,000	0,444	0,040
Blok70	0,020	0,160	0,012	0,040
naVodi	0,082	0,360	0,012	0,040
Z	0,327	0,160	0,309	0,360

Na kraju, otežavamo udaljenosti i dobijamo agregatnu udaljenost:

	agregatna udaljenost
sBrda	0,269
Mrka	0,429
Blok70	0,252
naVodi	0,363
Z	0,528

Kako je novi slučaj najbliži alternativni *Blok70*, to znači da očekujemo cenu od 80 miliona dinara.