

Ime: _____

Priimek: _____

Vpisna številka: _____

2. kolokvij pri predmetu

OSNOVE STROJNEGA UČENJA IN PODATKOVNEGA RUDARJENJA

Navodila:

- Na kolokvij ter dodatne pole je potrebno napisati **IME**, **PRIIMEK** in **VPISNO ŠT.**;
- Kolokvij sestavljajo **4 naloge**, ki so skupaj vredne **100 točk**;
- Čas reševanja je **80 minut**;
- Dovoljena je le uporaba kalkulatorja in enega A4 lista, kjer je lahko gor karkoli;
- Vse ostalo je prepovedano!!!

I	Date	E	F	G	Class
100	1913,555	0	C	3	X
101	1912,143	1	B	1	X
102	1917,455	?	B	2	Y
103	1920,000	?	B	1	X
104	1914,975	0	B	3	Y

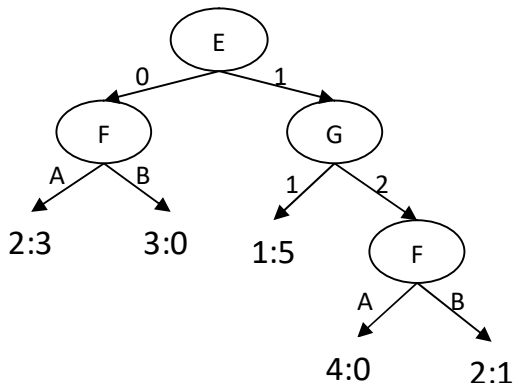
1. Odločitvena drevesa (30 točk)

- a. Podane imate podatke s štirimi atributi in razredom. Podani so informacijski prispevki za diskretne attribute F in G . Izračunajte informacijski prispevek zveznega atributa zveznega $Date$ in diskretnega atributa E .

- $\text{InfoGain}(F) = 0.275$
- $\text{InfoGain}(G) = 0.049$
- $\text{InfoGain}(E) =$
- $\text{InfoGain}(Date) =$

Kateri atribut je najboljši? Skicirajte prvi nivo odločitvenega drevesa!

- b. Klasificirajte spodnje primere na podlagi podanega odločitvenega drevesa. Porazdelitve so zapisane v obliki $\sigma(X) : \sigma(Y)$. Pripisite verjetnosti.



Date	E	F	G	Class	P(Class)
1910,111	0	A	?		
1900,121	1	?	2		
1999,890	1	A	?		

2. Odločitvena pravila (20 točk)

Poiščite odločitvena pravila za podatke iz tabele na prvi strani. Upoštevajte le atributa F in G.

3. Najbližji sosed (20 točk)

Višina (cm)	Spol	Obseg trebuha (cm)	Teža (kg)	D (Max)	D (Manhattan)
180	M	90	75		
160	M	140	120		
165	Ž	68	54		
175	M	110	80		
171	Ž	75	61		

Naj bodo *višina*, *spol* in *obseg trebuha* atributi, *teža* pa razred. S pomočjo regresijske metode **k-nn** z vrednostjo **k = 3** določite vrednost razreda spodnjega primera z uporabo:

- Maksimalne razdalje
- Manhattan razdalje

Razdalje in napovedana razreda vpišite v ustrezna polja v tabelah.

Višina	Spol	Obseg trebuha	Teža (Max)	Teža(Manhattan)
163	Ž	90		

4. Asociacijska pravila (30 točk)

Podane imate naslednje podatke o nakupih:

Račun 1	solata, rukola, paprika, bučke, kumare, fižol, zelje, redkvice
Račun 2	solata, rukola, paprika, radič
Račun 3	solata, kumare, paradižnik, paprika
Račun 4	solata, bučke, paprika, rukola, peteršilj, zelena
Račun 5	bučke, paradižnik

- Poiščite vsa pravila s podporo vsaj 60% in zaupanjem vsaj 80%
- Katera izmed dobljenih pravil z vsaj tremi artikli so zanimiva po metriki Lift (dvig) ?