

I	D	A	B	E	F	C
438	<b>12. 03. 2040</b>	5	3.49	14	dobro	y
450	24. 04. 1934	3	58.48	32	slabo	z
461	05. 01. 1989	5	47.23	12	slabo	y
466	07. 08. 1945	1	31.40	21	dobro	y
467	<b>21. 07. 2028</b>	5	79.60	20	slabo	y
469	30. 04. 1966	3	19.88	3	slabo	w
485	28. 02. 2015	5	59.13	4	slabo	w
514	19. 03. 2033	3	27.05	2	slabo	x
522	13. 03. 2022	2	80.14	16	dobro	y
529	28. 07. 2037	4	65.02	20	slabo	z
534	05. 10. 1986	2	99.17	13	dobro	z

566	20. 04. 1982	4	43.97	24	dobro	
578	15. 05. 2012	2	13.02	2	dobro	
600	30. 11. 1943	1	32.43	10	slabo	

I: identifikator, numerična vrednost  $[0, \infty)$

D: datum v obliki DD.MM.LLLL

A: diskretna vrednost  $\{1, 2, 3, 4, 5\}$

B: numerična vrednost  $(0, 100]$

E: numerična vrednost  $[0, 32]$

F: diskretna vrednost  $\{\text{dobro, slabo}\}$

C: razred, diskretna vrednost  $\{w, x, y, z\}$

1. Pretvori vrednosti atributa D v KSP format pri instancah, kjer vrednost atributa I ni deljiva z 2. Vrednost zaokroži na 4 decimalne vrednosti.
2. Skiciraj box-plot za vrednosti atributa B.
3. Diskretiziraj atribut B na 4 enako "visoke" intervale (odstopanje upoštevaj pri prvem podintervalu) Podintervale označi z B1, B2, B3 in B4. Skiciraj histogram.
4. Diskretiziraj atribut E na 3 enako "široke" intervale. Podintervale označi z E1, E2 in E3. Skiciraj histogram.
5. Uporabi algoritem ZeroR za klasifikacijo instanc neznanega razreda. Kolikšna je pri tem napaka?
6. Uporabi algoritem OneR za klasifikacijo instanc neznanega razreda. Upoštevaj attribute A, diskretiziran atribut B, diskretiziran atribut E in atribut F. Skiciraj vsa odločitvena drevesa. Kolikšna je napaka pri klasifikaciji?
7. Uporabi algoritem Naivni Bayes za določitev verjetnosti pripadanja neznanih instanc posameznim razredom. Upoštevaj enake attribute, kot pri nalogi 6. ter dodaj eno pojavitev k vsaki frekvenci. Verjetnosti normaliziraj.