



Sadržaj

1. Uvod	1
1.1. Kako čitati ovu knjigu?	3
2. Ciljevi i zadatci rudarenja podataka	7
2.1. Što (ni)je rudarenje podataka?	7
2.2. Proces rudarenja podataka	10
2.3. Podjela metoda rudarenja podataka	12
2.4. Primjene, primjeri, izazovi	14
3. Upoznavanje s R-om	17
3.1. Zašto naučiti R?	18
3.2. Prvi koraci	19
3.3. Naredbe za samopomoć (a.k.a. “help”)	21
3.4. R kao skriptni jezik	22
3.5. Objekti u R-u	23
3.6. Ispis objekta	24
3.6.1. Napomena o imenovanju objekata	25
Pitanja i zadatci za ponavljanje	26
Smjernice i korisne naredbe	26
4. Vektori, nizovi, faktori i tablice	27
4.1. Vektor kao temeljni objekt	28
4.2. Vektorizacija i operacije nad vektorima	32
4.3. Množenje vektora	34
4.4. Nizovi	37
4.5. Faktori i tablice	42
Pitanja i zadatci za ponavljanje	47
Smjernice i korisne naredbe	48
5. Indeksiranje, matrice i polja	49
5.1. Indeksiranje	50
5.2. Matrice i njihovo množenje	55
5.3. Indeksiranje matrica	59



5.4. Polja	66
Pitanja i zadatci za ponavljanje	69
Smjernice i korisne naredbe	69
6. Liste i okviri	71
6.1. Liste	72
6.2. Rad s listama	75
6.3. Okvir ili “data frame”	79
6.4. Rad s okvirima	84
Pitanja i zadatci za ponavljanje	90
Smjernice i korisne naredbe	90
7. Dohvat i pohrana podataka	91
7.1. Zapis tabličnih podataka u datoteku	92
7.2. Učitavanje tabličnih podataka – osnovne naredbe	95
7.3. Rad s putanjama i direktorijima	98
7.4. Rad s datumima i vremenima	100
7.4.1. Datumi	101
7.4.2. Datum-vrijeme formati – POSIXct i POSIXlt	103
Pitanja i zadatci za ponavljanje	105
Smjernice i korisne naredbe	105
8. Još neke korisne naredbe	107
8.1. Naredba attach() i maskiranje objekta	108
8.2. Naredbe detach() i with()	111
8.3. Naredba plot()	113
8.4. Operatori i operacije nad skupovima	117
8.5. Rad sa znakovima	120
8.6. Sortiranje i spajanje	125
Pitanja i zadatci za ponavljanje	130
Smjernice i korisne naredbe	130
9. Funkcije, petlje i uvod u F programiranje	131
9.1. Funkcije	132
9.2. Zašto se ne koristimo petljama u R-u?	137
9.3. Bezemne funkcije i F programiranje	140
9.4. Kompozicija funkcija i operatori	144



Pitanja i zadatci za ponavljanje	147
Smjernice i korisne naredbe	148
10. Izviđanje podataka	149
10.1. Statističke informacije o skupu podataka	152
10.2. Prikaz razdiobe podataka	155
10.3. Tablični i grafički prikaz	157
10.4. Istraživanje odnosa među varijablama	158
10.5. Zbirna obradba podataka	159
10.6. Jednostavniji prikazi podataka	161
10.7. Napredniji prikazi podataka	164
Pitanja i zadatci za ponavljanje	168
Smjernice i korisne naredbe	168
11. Primjeri dodatnih biblioteka za prikaz podataka	169
11.1. Paket <code>scatterplot3d</code>	170
11.2. Paket <code>lattice</code>	171
11.3. Paket <code>rgl</code>	173
11.4. Paket <code>MASS</code>	173
11.5. Paket <code>ggplot2</code>	174
12. Manipuliranje podatcima i tidyverse	177
12.1. Paket <code>tibble</code>	179
12.2. Paket <code>dplyr</code>	181
12.2.1. Odabir stupaca i redaka, preimenovanje stupaca – naredbe <code>select()</code> , <code>rename()</code> i <code>filter()</code>	182
12.2.2. Sortiranje i kreiranje novih varijabli - naredbe <code>arrange()</code> , <code>mutate()</code> , <code>summarise()</code> i <code>group_by()</code>	193
12.2.3. Pripajanje stupaca i redaka – naredbe <code>bind_cols</code> i <code>bind_rows</code>	199
12.2.4. Operacije nad skupovima – naredbe <code>intersect()</code> , <code>union()</code> i <code>setdiff()</code>	203
12.2.5. Kombiniranje skupova podataka – naredbe <code>*_join</code>	205
12.3. Paket <code>tidyverse</code>	208



12.4. Paket <code>purrr</code>	215
Pitanja i zadatci za ponavljanje	217
Smjernice i korisne naredbe	219
Za one koji žele znati više	220
Dodatna literatura	220
13. Gramatika vizualizacije <code>ggplot2</code> paketom	221
13.1. Temeljni slojevi	223
13.2. Oblikovni slojevi	225
13.2.1. Odnos statističkih i geometrijskih slojeva	240
13.2.2. Razdvajanje skupina naredbama <code>facet_wrap()</code> i <code>facet_grid()</code>	241
13.3. Slojevi detalja	242
13.3.1. Naredbe <code>scale()</code> i <code>labs()</code>	244
13.3.2. Funkcija <code>theme()</code>	246
13.4. Pohrana grafova	249
Pitanja i zadatci za ponavljanje	250
Smjernice i korisne naredbe	251
Za one koji žele znati više	252
Dodatna literatura	252
14. Proširenje paketa <code>ggplot</code> dodatnim bibliotekama	253
14.1. Paket <code>ggrepel</code>	253
14.2. Paket <code>ggradar</code>	254
14.3. Paketi <code>gganimate</code> i <code>gifski</code>	256
14.4. Paketi <code>plotly</code> i <code>shiny</code>	257
15. Izvještavanje kodom	259
15.1. <i>Rmarkdown</i>	260
15.1.1. YAML zaglavje	262
15.1.2. Osnove formatiranja teksta	263
15.1.3. Okruženje za kôd	264
15.2. <i>Web-aplikacije</i> uz paket <code>shiny</code>	267
Pitanja i zadatci za ponavljanje	268
Smjernice i korisne naredbe	268
Za one koji žele znati više	268
Dodatna literatura	268



16. Stabla odluke	269
16.1. Razvrstavanje kao model	270
16.2. Implementacija u R-u	273
16.3. Evaluacija i generalizacija	279
16.4. Mjere kvalitete detektora	284
16.5. Kontinuirane varijable u stablima	291
Za one koji žele znati više	295
Dodatna literatura	295
17. Grupiranje	297
17.1. Grupa i grupiranje	299
17.2. Algoritmi iz obitelji k -centara	304
17.3. Hiperarhijsko grupiranje	308
17.4. Grupiranje gustoćom	310
17.5. Grupiranje ili razvrstavanje	313
Pitanja i zadaci za ponavljanje	316
Smjernice i korisne naredbe	316
Za one koji žele znati više	316
Dodatna literatura	317
18. KNN klasifikator	319
18.1. Klasifikacija ili razvrstavanje	320
18.2. Algoritam KNN klasifikatora	321
18.3. Primjer: Klasificiranje vrsta perunike	323
18.4. Primjer: Određivanje pozicija NBA igrača na temelju statistike s utakmica	325
18.5. Prokletstvo dimenzionalnosti	329
18.6. Odabir značajki	331
Pitanja i zadaci za ponavljanje	337
Smjernice i korisne naredbe	337
Za one koji žele znati više	337
Dodatna literatura	338
19. Linearna regresija	339
19.1. Jednostavna regresijska analiza	340
19.2. Implementacija u R-u	342
19.3. Višestruka linearna regresija	346



19.4. Poopćenje linearog modela na nelinearne ovisnosti	348
19.4.1. Polinomijalna regresija	352
19.4.2. Logaritamska transformacija nezavisne varijable	354
19.5. Pretreniranje	356
Pitanja i zadatci za ponavljanje	361
Smjernice i korisne naredbe	361
Za one koji žele znati više	362
Dodatna literatura	362
20. Neuronske mreže	363
20.1. Model neurona	364
20.2. Učenje i optimizacija	367
20.3. Višeslojni perceptron	371
20.4. Arhitekture neuronskih mreža i dubinske neuronske mreže	375
Za one koji žele znati više	377
Dodatna literatura	379
Rješenja zadataka	381
Zaključak	397
Kazalo pojmova	399