Univerzitet u Istočnom Sarajevu
Filozofski fakultet

PROGRAMSKI JEZICI R I PYTHON

Jovana Forcan

Univerzitet u Istočnom Sarajevu - Filozofski fakultet
Akademska misao

PROGRAMSKI JEZICI R I PYTHON

Autor:
Jovana Forcan

Recenzenti:
Dr Dragoljub Krneta, vanredni profesor
Univerzitet u Istočnom Sarajevu
Filozofski fakultet

Dr Miloš Ljubojević, vanredni profesor
Univerzitet u Banjoj Luci
Elektrotehnički fakultet

Dizajn korica:
Boris Popović

Izdavač:
Univerzitet u Istočnom Sarajevu, Filozofski fakultet
Akademska misao, Beograd

Štampa:
Akademska misao, Beograd

Tiraž:
200 primjeraka

ISBN:
978-86-7466-999-0

Mjesto i godina izdanja:
Beograd, Istočno Sarajevo, 2024.

Odlukom Senata Univerziteta u Istočnom Sarajevu broj 01-C-75-LXIII/24 od 29.2.2024. godine, rukopis „Programski jezici R i Python" autora Jovane Forcan, odobren je za objavljivanje kao univerzitetski udžbenik.

S ljubavlju i zahvalnošću
mojim dragim roditeljima, koji su mi uvijek bili oslonac

Predgovor

R i Pajton (*Python*) su dva istaknuta programska jezika u domenu analize podataka i statističkih istraživanja. Oba jezika se odlikuju širokim spektrom naprednih biblioteka i alata, čime omogućavaju jednostavno i efikasno rješavanje složenih analitičkih izazova u svijetu analize podataka. Ovi izazovi uključuju statističko modelovanje, mašinsko učenje i vizualizaciju podataka.

Ovaj udžbenik kreiran je sa ciljem da pruži sveobuhvatan pregled programskih jezika R i Pajton. Namijenjen je kako početnicima, tako i iskusnim korisnicima. Udžbenik je strukturiran u poglavlja koja pokrivaju različite aspekte analize podataka, prateći korake od osnovnih koncepata R i Pajton jezika prema naprednijim temama. Svako poglavlje sadrži brojne primjere kako bi koncepti i tehnike bili ilustrovani i omogućili čitaocima bolje razumijevanje.

Sadržaj udžbenika je organizovan u dva dijela.

Prvi dio udžbenika detaljno istražuje karakteristike programskog jezika R. Ovdje se obrađuju teme kao što su strukture podataka, manipulacija podacima, vizualizacija i statističko modelovanje. Čitalac će se takođe upoznati sa nekim od najpopularnijih R paketa i biblioteka, uključujući dplyr, ggplot2 i druge, uz praktične smjernice o njihovoj primjeni za različite zadatke analize podataka. Prvi dio udžbenika sastoji se od 6 poglavlja.

- Uvod u R: ovo poglavlje pruža osnovni uvod u programski jezik R, njegovo integrisano razvojno okruženje, proces preuzimanja i instalacije R-a i RStudio-a, te uvodi različite programske paradigme.

- Osnove programiranja u R-u: u ovom poglavlju razmatra se sintaksa R-a, osnovne operacije, rad sa stringovima, upotreba paketa i različite strukture podataka. Takođe se obrađuje kontrola toka programa, rad sa funkcijama, te koncepti objektno orijentisanog programiranja.

- Obrada podataka: ovo poglavlje fokusira se na upotrebu paketa dplyr i data.table za obradu podataka, kao i na različite pakete za povezivanje R-a sa bazom podataka.

- Grafika u R-u: poglavlje razmatra kreiranje različitih vrsta grafikona pomoću paketa graphics i ggplot2.
- Uvod u statistiku: u ovom poglavlju detaljno se istražuju koncepti deskriptivne i inferencijalne statistike, sa posebnim fokusom na njihovu primjenu unutar programskog jezika R.
- Regresiona analiza: poglavlje razmatra linearnu i nelinearnu regresiju, istražujući postupke i tehnike za izradu efikasnih regresionih modela unutar R-a.

Drugi dio udžbenika fokusira se na osnovne karakteristike jezika Pajton i obuhvata najpopularnije Pajton biblioteke i module, uključujući NumPy, Pandas, Matplotlib i Scikit-Learn, pružajući korisne smjernice za njihovu primjenu u različitim analitičkim zadacima. Drugi dio udžbenika sastoji se od 5 poglavlja.

- Uvod u Pajton: ovo poglavlje pruža uvod u Pajton, njegovu instalaciju i predstavlja platforme Anaconda i Google Colab, koje olakšavaju rad sa Pajtonom.
- Kurs programiranja u Pajtonu: ovo poglavlje obuhvata različite tipove podataka u Pajtonu, kontrolu toka programa (if-else, petlje), module, funkcije, rad sa izuzecima, objektno orijentisanu paradigmu, rad sa datotekama i povezivanje sa bazom podataka.
- Moduli NumPy, Pandas i SciPy: poglavlje obrađuje tri ključna modula u Pajtonu koji se široko koriste u naučnim i analitičkim aplikacijama.
- Grafika u Pajtonu: poglavlje se bavi bibliotekama Matplotlib, Seaborn i drugim alatima za vizualizaciju podataka.
- Regresiona analiza i osnove klasifikacije u Pajtonu: u ovom poglavlju istražuju se ključne tehnike analize podataka koje se koriste za modelovanje i predviđanje.

Glavni cilj ovog udžbenika je osigurati čvrsto razumijevanje jezika R i Pajton i omogućiti čitaocima efikasno korišćenje ovih jezika za raznovrsne analitičke zadatke povezane sa podacima. Knjiga je namijenjena studentima prirodnih, računarskih i tehničkih nauka, kao i studentima ekonomije i društvenih nauka. Takođe je idealna za one koji pohađaju interdisciplinarne programe koji kombinuju elemente matematike, informatike, statistike, nauke o podacima i drugih disciplina. Ova knjiga pruža uvide i vještine koje su primjenljive u različitim oblastima, čineći

je idealnim izvorom za raznovrsnu akademsku zajednicu, posebno za one koji teže da integrišu napredne analitičke i programerske vještine u svoje akademsko i profesionalno usavršavanje.

Zahvaljujem se profesoru dr Dragoljubu Krneti i profesoru dr Milošu Ljubojeviću na angažovanju i stručnosti koje su unijeli u recenziju mog djela, čime su značajno doprinijeli njegovom kvalitetu.

Sadržaj

Dio I

R

Poglavlje 1

Uvod u R

R je moćan i svestran programski jezik sa širokom primjenom u analizi podataka, statističkom modelovanju i vizualizaciji podataka. Njegova funkcionalnost obuhvata obilje ugrađenih funkcija, opsežan broj paketa i aktivnu zajednicu korisnika i programera koji neprekidno doprinose njegovom razvoju i održavanju. To ga čini popularnim izborom među stručnjacima u oblasti analize podataka i statističarima. Pored toga, R se ističe svojom paradigmom funkcionalnog programiranja, obiljem grafičkih mogućnosti i širokim spektrom primjena u različitim oblastima, što ga čini neizostavnim alatom za rješavanje različitih izazova.

1.1 Šta predstavlja R?

R je programski jezik, interpreter i platforma.

R, kao programski jezik omogućava pisanje programa za različite svrhe.

R, kao interpretirani jezik omogućava izvršavanje R programa liniju po liniju iz konzole ili pokretanje skripata iz .R datoteka pomoću ugrađenog interpretera.

R, kao platforma otvorenog koda pruža bogat skup alata za manipulaciju podacima, izvođenje statističkih proračuna i vizualizaciju rezultata.

R je kompatibilan s operativnim sistemima poput Windows-a, Mac OS-a i raznih UNIX platformi, uključujući Linux. Većina platformi distribuira R u obliku binarnih datoteka kako bi se olakšala instalacija. Ovaj softverski projekat prvobitno su pokrenuli Ros Ihak (*Ross Ihak*) i Robert Džentlmen (*Robert Gentleman*). R se smatra implementacijom jezika S, kojeg su razvili Rik Beker (*Rick Becker*), Džon Čejmbers (*John Chambers*) i Alan Vilks (*Allan Wilks*).