



DATAKOLEKTIV

ADVANCED ANALYST Foundations for Advanced Data Analytics in R



Kurs **ADVANCED ANALYST: Foundations for Advanced Analytics in R** u celini traje tri meseca i obuhvata (a) 12 nastavnih sesija, koje se organizuju subotom, u trajanju od po 1.5 - 2h svaka, (b) grupni sastanak za diskusiju važnih pitanja jednom nedeljno u trajanju od 45 minuta do 1h, (c) komunikaciju u realnom vremenu na platformi Slack u vezi svih pitanja, i (4) *ad hoc* 1:1 sastanke u trajanju do 45 minuta sa predavačem ukoliko je polazniku neophodna pomoć koja zahteva neposredan, zajednički rad.

Preporučena priprema za pohađanje ovog kursa je besplatni kurs uvoda u R programiranje koji je Goran S. Milovanović održao u saradnji sa Startit: [YouTube plejlista kursa Uvod u programski jezik R](#).

Svim zainteresovanima DataKolektiv savetuje da se pre odluke o pohađanju ovog kursa jave kako bismo organizovali razgovor u kome možemo da objasnimo sve detalje ove ponude.

Cenovnik kursa

	Nezaposleni i studenti	Zaposleni, firme, i organizacije
Advanced Analytics: Foundations for Advanced Analytics in R	EUR 1200	EUR 1500

Goran Milovanović PR Data Kolektiv

ul. Breza 4/7, 11000 Beograd, Serbia, PIB: 109890695, APR: 64498339
e-mail: goran.s.milovanovic@datakolektiv.com, telefon: +381691686298
www.datakolektiv.com

Program kursa

ADVANCED ANALYST: Foundations for Advanced Data Analytics in R

Modul 1: Bootcamp za R programiranje

- **Nedelja 01**
 - Instalacije i tehnički preduslovi
 - Uvod u R i RStudio
 - Osnove R markdown
 - Navigacija kroz okruženje
 - Osnovni tipovi i podaci (vektori, matrice, liste, data.frames)
 - Rad sa promenljivama i osnovne operacije
 - Kontrola toka (if-else odluke, petlje)
- **Nedelja 02**
 - Funkcije u R
 - Uvod u koncepte funkcionalnog programiranja
 - Porodica funkcija "apply" (apply, lapply, sapply, mapply)
 - Prilaz tidyverse: manipulacija i eksploracija podataka
 - Čitanje i pisanje podataka u fajlove (CSV, Excel, itd.)
 - Dublje u R markdown za izveštavanje

Modul 2: Vizualizacija podataka i eksploratorna analitika

- **Nedelja 03**
 - Osnove ggplot2: gramatika grafike
 - Osnovne vizualizacije: linijski i stubičasti grafikoni, dijagrami raspršenja (skaterplot), histogrami, bubble grafikoni, box-plot grafikoni, linije trenda i paneli u ggplot2
 - Deskriptivna statistika i njena interpretacija
 - Eksploratorna analitika u ggplot2
 - Detekcija neuobičajenih vrednosti (autlejera) u podacima
 - Napredne ggplot2 vizualizacije
 - Izvoz i formatiranje ggplot2 vizualizacija
- **Nedelja 04**
 - Osnove Plotly: jednostavna interaktivna vizualizacija
 - Osnovne vizualizacije: linijski i stubičasti grafikoni, dijagrami raspršenja (skaterplot), histogrami, bubble grafikoni, box-plot grafikoni, linije trenda i paneli u Plotly
 - Iz ggplot2 u Plotly
 - Eksploratorna analitika u Plotly
 - Osnove A/B testiranja
 - ggplot2 i Plotly u R markdown: fundamentalno izveštavanje

Modul 3: Osnovna prediktivna analitika i forekasting

- **Nedelja 05**
 - Jedan prediktor: Jednostavna linearna regresija
 - Više prediktora: Višestruka linearna regresija
 - Studija slučaja 1: Izveštaj o prediktivnoj analitici linearnom regresijom: predikcija cena nekretnina
- **Nedelja 06**
 - Osnove analize vremenskih serija
 - Komponente vremenskih serija
 - Prognoza vremenskih serija koristeći ARIMA model
 - Studija slučaja 2: Elementarni forkasting izveštaj

Modul 4: Prikupljanje podataka iz API-ja i PDF-a, automatizovana produkcija u MS Office iz R, i OpenAI ChatGPT iz R

- **Nedelja 07**
 - R ekosistem podataka: paketi za pristup različitim izvorima podataka
 - Uvoz podataka iz API-ja: jednostavni API pozivi
 - Razumevanje JSON i XML formata
 - Napredni API pozivi i obrada podataka
 - Ekstrakcija podataka iz PDF fajlova
 - Studija slučaja 3: Izveštaj o podacima Svetske banke
- **Nedelja 08**
 - Interakcija sa MS Excelom iz R
 - Automatizacija produkcije PowerPointa iz R
 - Izveštavanje u MS Wordu iz R
 - OpenAI API: automatizovana interakcija sa ChatGPT iz R
 - Studija slučaja 4: Izveštaj o analizi sentimenta teksta

Modul 5: Klasifikacija (nadgledana i nenadgledana)

- **Nedelja 09**
 - Uvod u binomijalnu logističku regresiju u R
 - Uvod u multinomijalnu regresiju u R
 - Studija slučaja 5: Izveštaj o rešavanju churn problema binomijalnom logističkom regresijom
- **Nedelja 10**
 - Analiza glavnih komponenti (PCA) u R
 - Studija slučaja 6: Izveštaj o segmentaciji tržišta kroz PCA

Modul 6: Model drveta odlučivanja i algoritmi klasterizacije u R

- **Nedelja 11**
 - Uvod u teoriju informacija za klasifikaciona drveta
 - Pojmovi informacije i entropije, informacione dobiti (Information Gain) i Gini dobiti (Gini Gain)
 - Drvo odlučivanja za klasifikacioni problem u rpart paketu
 - Predikcija drvetom odlučivanja
 - Regresiona drveta odlučivanja
 - Studija slučaja 7: Izveštaj o predikciji cena na tržištu nekretnina regresionim drvetom odlučivanja
- **Nedelja 12**
 - k-Means Clustering algoritam u R
 - Određivanje broja klastera u rešenju
 - Studija slučaja 8: Izveštaj o segmentaciji korisnika kroz k-Means Clustering

O predavaču

Dr Goran S. Milovanović



[Goran S. Milovanović, PhD](#), je profesionalni Data Scientist sa velikim iskustvom u fundamentalnim i primenjenim istraživanjima, istraživanjima tržišta, medijskoj analitici, oblasti pronalazaženja informacija (*Information Retrieval*), izveštavanju, vizuelizaciji, i vođenju Data Science projekata i softverskih sistema i timova. Predavao je i kao saradnik učestvovao u nastavi psihologije, metodologije, statistike, R programiranja (i rada sa drugim statističkim sistemima) od ranih 90-ih godina, onlajn kao i u klasičnim setinzima, na srpskom i engleskom jeziku, u akademskom, nevladinom i profitnom sektoru, godinama pružajući mentorstvo za R u Data Science polaznicima obrazovnih programa američkog startapa *Springboard*. Sa akademskim bekgroundom u društvenim naukama (Phd psihologija u oblasti komputacionih kognitivnih nauka: teorija odlučivanja, kauzalno statističko učenje, suđenje, distribuciona semantika, kategorizacija i učenje koncepata) programer je od 80-ih godina. Od 2017 do 2022 godine preko svoje konsultantske kuće DataKolektiv pruža Data Science usluge za [Wikidata](#), najveću svetsku otvorenu bazu znanja, koristeći R kao deo tehnološkog steka koji obuhvata još i Python, Apache Hadoop, Apache Spark, RBDS, XML, JSON, RDF, SPARQL, RStudio Shiny, R Markdown, CloudVPS, i Docker. Njegov rad obuhvata full-stack Data Science, od rada sa Big Data sistemima, preko mašinskog učenja, do razvoja interaktivnih dašborda i izveštaja u produkciji u virtuelizovanim okruženjima. Autor je i urednik nekoliko knjiga u oblasti razvoja Informatičkog društva, vodeći timove koji su razvili prva istraživanja Internet ponašanja i stavova prema Internetu u regionu Jugoistočne Evrope 2002 - 2005. Ponekad piše za [R-bloggers](#) i redovan je učesnik (i ponekad član organizacionih odbora i timova) [European R Users Conference](#). Vodi malu konsultantsku kuću [DataKolektiv](#), u Beogradu, Srbija, a trenutno vodi LABS team u srpsko-holadnskom media-tech startapu [smartocto](#) kao Lead Data Scientist. Keynote speaker je na globalnim događajima poput Webit i Data Science Conference.

Kontakt: goran.milovanovic@datakolektiv.com

Tech Stack

Uglavnom koristim [R](#), ali ceo zoo bi obuhvatao:

- Front-end: ([Shiny](#), [R Markdown](#) - koristim ili [Open Source Shiny Server](#), ili ravno u produkciju sa [golem](#) skalirajući sisteme sa [ShinyProxy](#) i [Docker](#))
- vizuelizacije: [ggplot2](#), [Bokeh](#), [Plotly](#), [VisNetwork](#), [igraph](#);
- Back-end/ML: [data.table](#), [text2vec](#), [maptpx](#), [Rtsne](#), [XGBoost](#), [Matrix](#), [mclust](#), [scikit-learn](#) i verovatno još mnogo toga;
- [Python/Pyspark](#)/[Apache Spark](#), [Hadoop/HiveQL](#), and [MariaDB](#), tipično za Data Engineering i ETL poslove;
- [XML](#), [JSON](#), [RDF](#);
- naravno da je [RStudio](#) moj omiljeni IDE, ali koristim i [Visual Studio Code](#) kao i ponekad [PyCharm](#).

GitHub

- Goran: github.com/GoranMilovanovic
- DataKolektiv: github.com/datakolektiv