

Uvod u R za Data Science
Jesen 2018
Startit, Beograd

Goran S. Milovanović

Wikimedia Deutschland, Data Scientist for Wikidata
DataKolektiv, Owner

- Kurs *Uvod u R za Data Science* će biti održan u deset sesija.
- **Materijali** za kurs su sledeći:
 - [GitHub repozitorijum](#) sa R kodom i RMarkdown notebooks sa prethodnih kurseva u Startitu;
 - knjiga R za statističku obradu podataka (O'Reilly izdanje na srpskom jeziku, Mikroknjiga, 2017, odn. [R for Data Science](#), na engleskom, kompletno otvorena online, aut. Hadley Wickham & Garrett Grolemund);
 - knjiga za one koji žele da prodube svoje poznavanje programiranja u R: [THE ART OF R PROGRAMMING](#), aut. Norman Matloff.
- **Od polaznika se očekuje sledeće:**
 - da prisustvuju sesijama kursa;
 - da se pripremaju svaku sesiju unapred (čitanje + kodiranje odn. vežbe u R koje će biti zadavane); ovo je neophodno pošto će kurs biti prilično intenzivan ali zato i efikasniji (nadam se) od prethodnih; fokus će biti na knjizi R for Data Science, ali pošto ona prati na linearan način učenje R programiranja polaznici će jezičke ekspresije u R savladavati uporedo sa poglavljima knjige, kroz vežbe i kod na GitHub;
 - da pronađu, u dogovoru sa predavačem, skup podataka na kome će vežbati i razvijati svoje znanje programskog jezika R tokom kursa i na kraju kursa predstaviti mini-projekat na osnovu analize i vizuelizacije tog skupa podataka.
- **Polaznici mogu da očekuju sledeće:**
 - poznavanje i razumevanje logike rada sa R paketima iz {tidyverse};
 - I/O funkcije samog jezika, paketa poput {readr} i {data.table};
 - transformisanje i uređivanje skupova podataka sa paketima {dplyr} i {tidyr};
 - vizuelizaciju podataka statističkim dijagramima u paketu {ggplot2};
 - poznavanje toka rada za eksploratornu analizu podataka (EDA), deskripciju podataka, i elementarne neparametrijske metode u R;
 - elementarno matematičko modeliranje u programskom jeziku R koje će se proučavati na primeru linearnog regresionog modela i nekih veoma korisnih modela iz klase Generalizovanih linearnih modela (binarna i multinomijalna logistička regresija).

- **Polaznici na kraju kursa ne treba da očekuju:**
 - poznavanje R za produkciju (Rmarkdown, RStudio Shiny, etc);
 - poznavanje paralelizacije R koda;
 - rad sa Big Data kroz R;
 - znanje razvoja interaktivnih, dinamičkih vizuelizacija putem R interfejsa ka D3 funkcijama i sl. JS front-end alatima;
 - elaborirano znanje matematičkog modeliranja u određenoj oblasti (to je granica koju samo oni sami sebi mogu da zacrtaju i osvajaju je).

O predavaču

Dr Goran S. Milovanović se više od petnaest godina profesionalno bavi analitikom podataka, a radi kao Data Scientist od 2013. godine, posle akademske karijere u kognitivnim naukama (suđenje i rezonovanje, odlučivanje u uslovima rizika i neizvesnosti, semantičko pamćenje, kategorizacija, statističko kauzalno učenje). Ima akademsko, profesionalno, i volontersko iskustvo u nastavi statistike u društvenim naukama, te eksperimentalne i ne-eksperimentalne metodologije, na srpskom i engleskom jeziku.

Trenutno je angažovan kao Data Scientist na semantičkom veb projektu Wikidata od strane Wikimedia Deutschland, i kao mentor u oblasti Data Science za američki startup Springboard. Vlasnik je konsultantske firme DataKolektiv u Beogradu. Njegova ekspertiza u Data Science je u Information Retrieval problemima (u eksploratornoj analitici) i predikciji ljudskih izbora (u prediktivnoj analitici). Više o njegovom radu i interesovanjima se može naći na stranicama bloga [The Exactness of Mind](#). Član je [Data Science Srbija](#).

