

Statistička analiza podataka

Upute za rad na projektu i izradu izvještaja

Stjepan Begušić, Tessa Bauman, David Bojanić, Tomislav Kovačević, Andro Merćep

Studeni 2021.

Rad na projektu

Pri radu na projektu važna je komponenta timski rad – iako će se uvijek naći kolege koji će se znanjem i trudom istaknuti u svakoj grupi, svakako preporučujemo da uključite sve kolege u rad grupe, budući da je znanje cijele grupe bitno pri konačnoj predaji. U samom radu na dodijeljenoj temi važno je imati na umu sljedeće:

- Pregledajte datoteke s podacima i način na koji su sami podatci učitani u vašu razvojnu okolinu (RStudio) – uvijek provjerite jesu li ispravno učitani i postoje li kakvi sumnjivi unosi ili nedostajuće vrijednosti koje mogu utjecati na vašu analizu.
- Tijekom rada na projektu vizualizirajte podatke na različite načine i analizirajte ono što vidite – iako neće sve vizualizacije nužno ući u konačni izvještaj, mogu vam biti korisne u analizi i vlastitim istraživačkim pitanjima i hipotezama.
- Istraživačke smjernice dane u opisu projekta su samo osnovne natuknice za izradu projekta – svakako razmišljajte i analizirajte vlastita istraživačka pitanja i hipoteze u svom projektu.

Izvještaj

Sam izvještaj predajete u obliku .rmd datoteke koja sadrži kod, rezultate i tekst, te .pdf datoteke koju generira RStudio kad pokrenete *knit* vašeg .rmd izvještaja. Također je nužno predati sve datoteke koje sadrže međurezultate ili dodatne podatke potrebne za izvršavanje koda vašeg projekta. Potrebno je da izvještaj sadrži određene deskriptivne analize i vizualizacije kojima možete motivirati neke od vaših analiza, različite testove kojima biste neke od njih provjerili, te eventualne modele (npr. ANOVA ili regresija, ovisno o tome kakvim se podacima bavite) kojima objašnjavate različite efekte. Pri izradi izvještaja vodite se sljedećim smjernicama:

- Sam izvještaj sadrži izmjenjujuće *chunkove* koda, rezultata (slike, tablice, ispisi iz različitih testova i analiza) i teksta. Sam tekst je nužan kako biste povezali i objasnili ono što radite u kodu, ali nipošto ne treba pisati previše (po par rečenica kod svakog dijela koda ili analize koje objašnjavaju što i zašto radite te komentiraju rezultate).
- U izvještaju preporučujemo da pokušate što više povezati zajedničku "priču" cijelog projekta. Upravo forma izvještaja kroz deskriptivnu analizu, vizualizaciju i statističke analize vam daje priliku da slijedno predstavite svoj projekt.

- Sve što radite na projektu, kao i sve što ste pokušavali ili isprobavali, ne mora nužno biti u izvještaju – ključno je da odaberete one primjere koje želite prikazati u projektu (pritom nije dobro birati samo one analize gdje su vaše hipoteze "ispale" točne - dapače, dobro je pokazati i slučajeve koji su vas možda posebno iznenadili i dati neku interpretaciju tome).
- Iako nema minimalne i maksimalne duljine izvještaja (a i ona uvelike ovisi o količini slika, tablica i ispisa koje ćete imati), ako vaš izvještaj ima previše ispisa koji su redundantni (npr. ispisi ili upozorenja iz nekih funkcija u R-u koji nisu nužno dio analize), možete spriječiti njihov ispis u izvještaju, iako se funkcija izvršava u kodu.

Kontrolna točka i predaja projekta

U sklopu izrade projekta postoji kontrolna točka koja je obavezna za sve grupe. U tom trenutku će svaka grupa sa svojim asistentom proći kroz svoj projekt (odnosno ono što do tog trenutka imate), a asistenti će dati preporuke za proširenje ili popravak projekta. Za uspješnu kontrolnu točku svakako preporučujemo da se pripremite s konkretnim pitanjima za asistente.

NAPOMENA: Kontrolna točka se ne ocjenjuje i sasvim je dopušteno doći s potpuno praznim projektom (iako je i dalje obavezna) – no u tom slučaju gubite priliku dobiti dubinski uvid u vaš projekt i savjete kako ga proširiti ili poboljšati.

Osim kontrolne točke, asistenti su vam na raspolaganju za upite tijekom cijelog semestra – svakako preporučujemo da iskoristite i tu priliku (no, na takvim konzultacijama asistenti mogu odgovarati na konkretna pitanja – iz praktičnih razloga ne mogu pregledavati vaše cijelo rješenje i tražiti greške).

Pri predaji projekta asistenti će zajedno sa svim članovima tima prolaziti kroz projektni izvještaj i naizmjenice postavljati pitanja pojedinim članovima tima. Svi članovi tima moraju znati sve aspekte rješenja. Pritom se ispituju i boduju: sam izvještaj, poznavanje koda (što se događa u određenim dijelovima koda i zašto radite određene analize – što njima želite pokazati) te poznavanje teorije u pozadini metoda koje koristite (zašto smijete ili ne smijete koristiti određeni test, kako radi određena metoda itd.). Studenti često znaju proširiti osnovni skup metoda i alata iz SAP-a i koristiti i neke naprednije metode u svojim analizama – to nipošto nije zabranjeno, ali imajte na umu da vas asistenti mogu ispitati sve što koristite u projektu, pa tako i teoretske pozadine svih metoda koje koristite.

Za sva pitanja tijekom rada na projektu svakako se obratite asistentima.

Želimo vam uspješan rad na projektu.