

**Sažetak analize pristranosti u sustavu ocjenjivanja projekata
Projekt razvoja karijera mladih istraživača (DOK-2025-02)
Prijavljeni za znanstveno područje "Prirodne znanosti"**

*VAŽNO: Detaljni opisi metoda, rezultati i njihove interpretacije
nalaze se u dokumentu*

Analiza_sustava_evaluacije_DOK_2025_02.pdf

Branimir K. Hackenberger

5. kolovoza 2025.

1 Uvod

Analizirano je 95 projekata prijavljenih u području prirodnih znanosti s ukupno 285 jedinstvenih kombinacija projekt-recenzent. Cilj analize bio je utvrditi objektivnost i valjanost postupka vrednovanja kroz klasične statističke metode, napredne analize i Bayesov pristup.

2 Ključne metode i njihova svrha

2.1 Klasična statistička analiza

- **ICC analiza** - mjeri pouzdanost među recenzentima
- **Mixed Effects Model** - razdvaja izvore varijabilnosti ocjena
- **Klaster analiza** - identificira stilove recenziranja

2.2 Napredne statističke metode

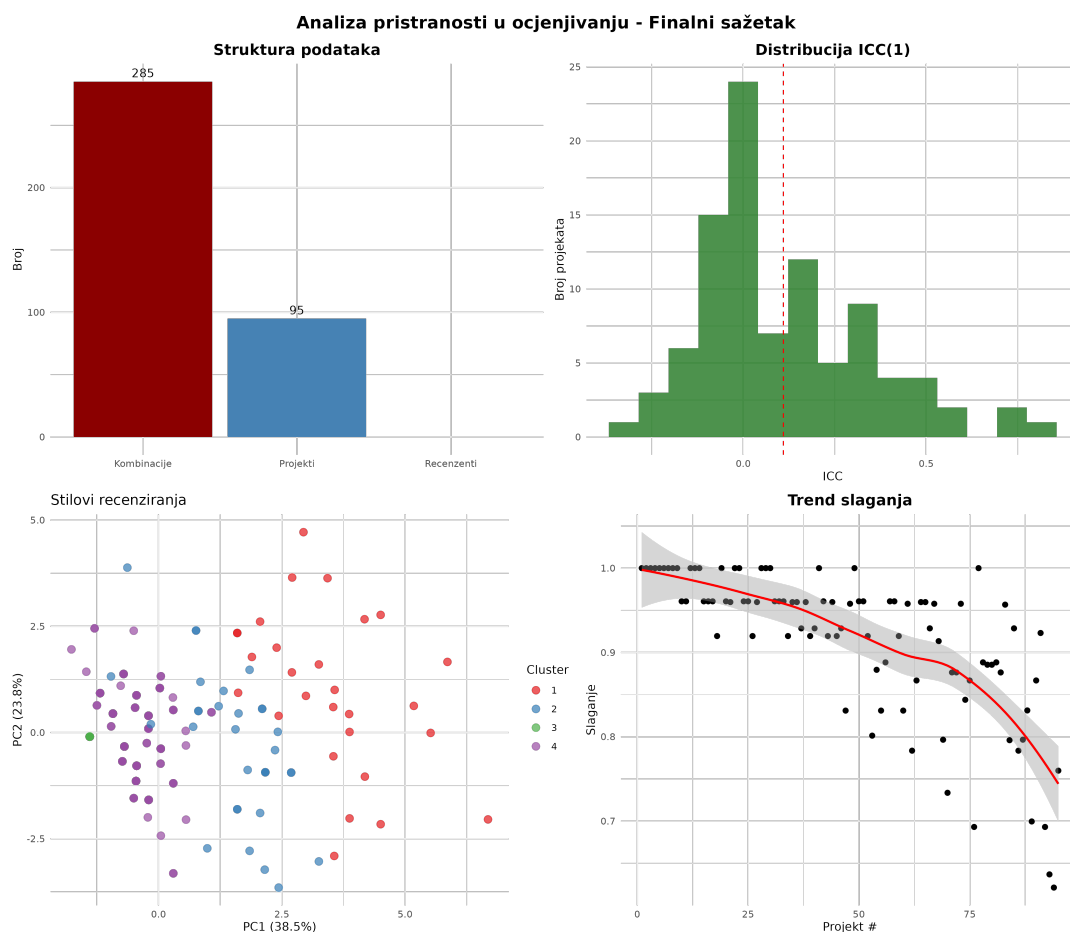
- **Kendall-ov W** - kvantificira slaganje recenzenata
- **Monte Carlo simulacija** - uspoređuje s nasumičnim ocjenjivanjem
- **Analiza ceiling učinka** - mjeri koncentraciju visokih ocjena

2.3 Bayesova analiza

- **Hijerarhijski modeli** - procjena komponenti varijance s nesigurnošću
- **Bootstrap analiza** - procjena stvarnog broja recenzenata

3 Ključni rezultati s grafovima

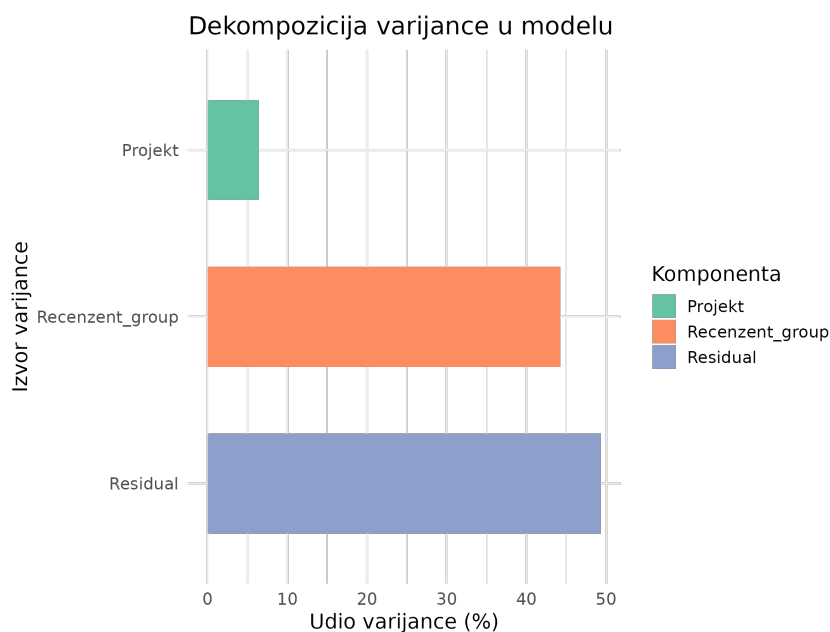
3.1 Struktura podataka i pouzdanost



Slika 1: Sažetak analize pristranosti. Prosječni $ICC(1) = 0.11$ pokazuje izrazito nisku pouzdanost među recenzentima. Identificirana su 4 različita stila recenziranja, a slaganje recenzenata dramatično pada kroz vrijeme.

Rezultat: Samo 11% varijance u ocjenama može se pripisati stvarnim razlikama između projekata.

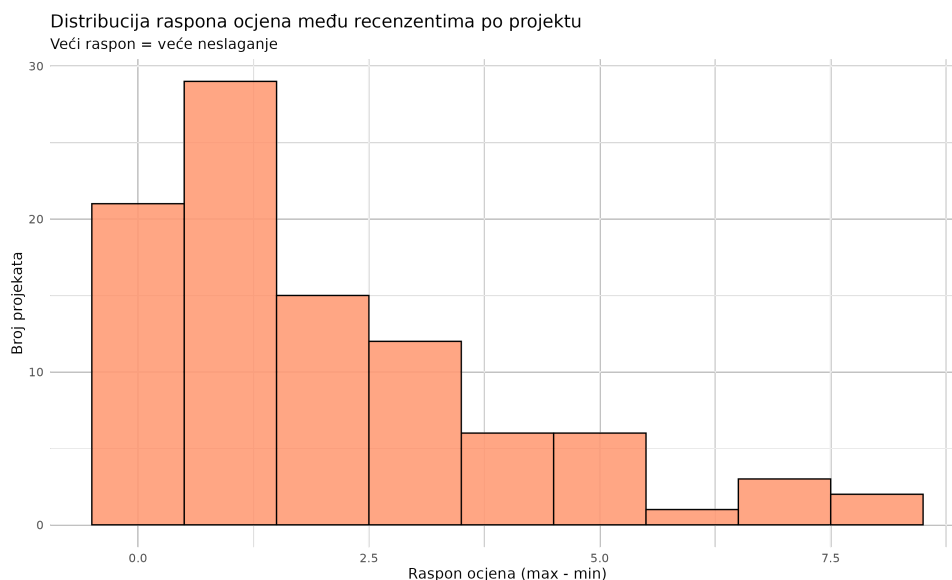
3.2 Dekompozicija varijance



Slika 2: Dekompozicija varijance pokazuje da samo 6.4% ukupne varijance proizlazi iz stvarnih razlika između projekata. Dominantni izvori su grupe recenzenata (44.2%) i rezidualna varijanca (49.3%).

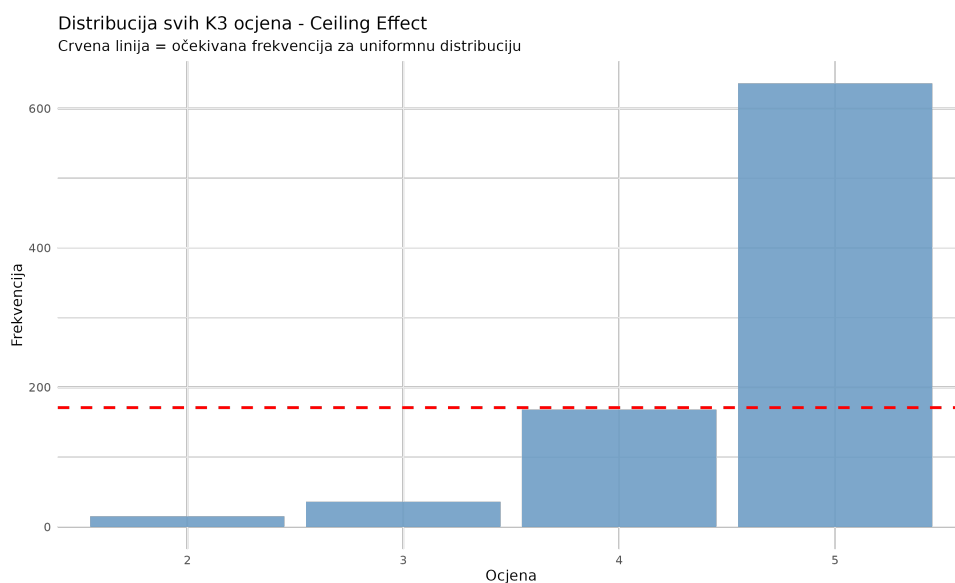
Ključni nalaz: 93.6% varijance nema veze s kvalitetom projekata!

3.3 Distribucija neslaganja



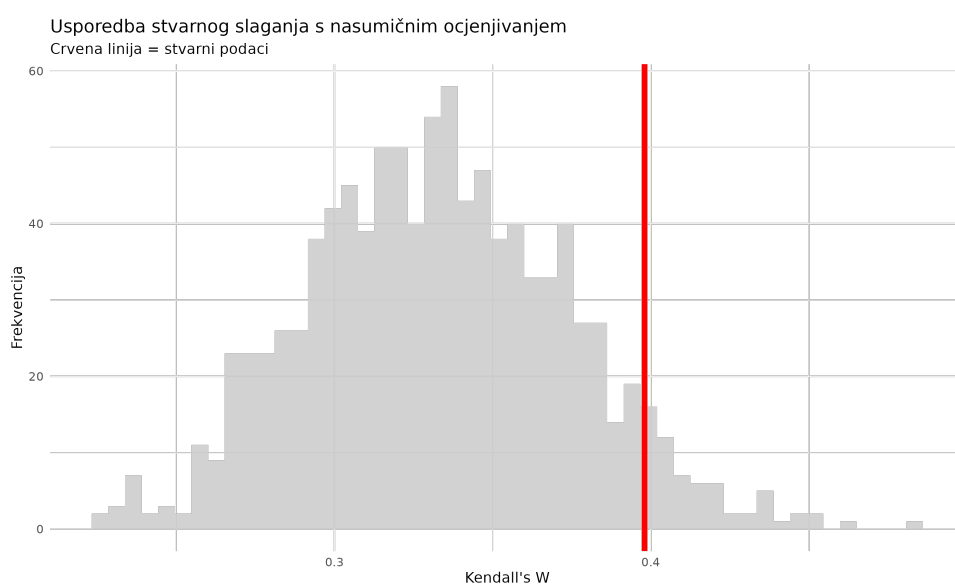
Slika 3: Distribucija raspona ocjena među recenzentima. Značajan broj projekata ima raspon 2-8 bodova, što predstavlja više od 50% ukupnog raspona skale.

3.4 Ceiling učinak



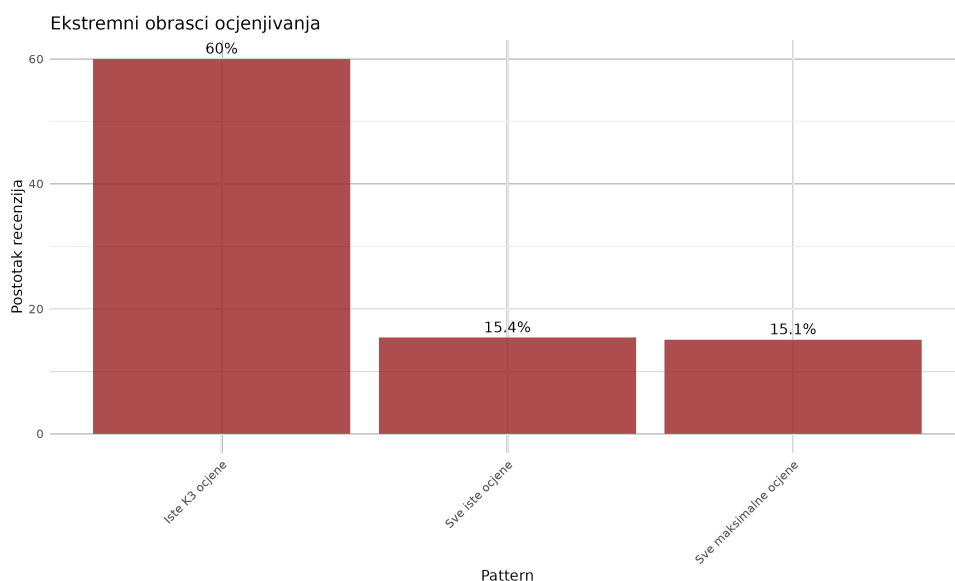
Slika 4: Distribucija K3 ocjena pokazuje ekstremni ceiling učinak - 94% svih ocjena je 4 ili 5, čineći većinu skale (1-3) praktički neupotrebljivom.

3.5 Usporedba s nasumičnim ocjenjivanjem



Slika 5: Usporedba stvarnog slaganja recenzenata ($W = 0.398$, crvena linija) s nasumičnim ocjenjivanjem. Samo 6% nasumičnih simulacija proizvelo je jednako ili bolje slaganje ($p = 0.06$).

3.6 Ekstremni obrasci ocjenjivanja



Slika 6: 60% recenzija daje identične ocjene za sve tri K3 komponente, 15.4% daje identične ocjene za sve komponente, a 15.1% daje maksimalne ocjene za sve.

4 Bayesova analiza - dodatna potvrda

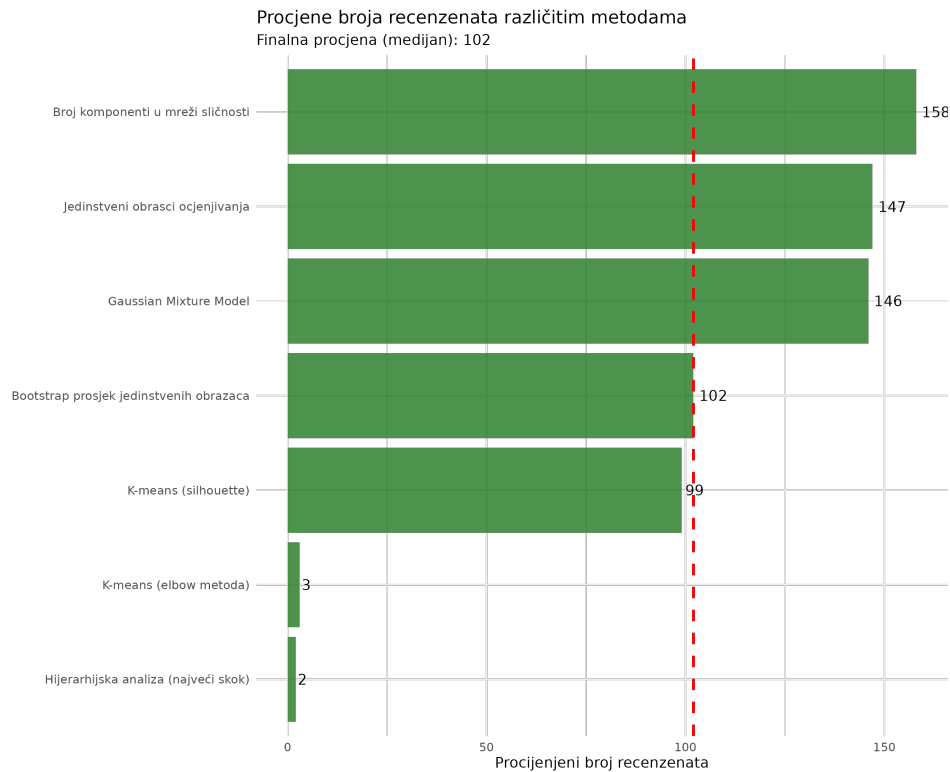
Tablica 1: Bayesova dekompozicija varijance

Izvor	Proporcija varijance
Projekti	9%
Recenzenti	32%
Rezidualna	59%

Ključni rezultati Bayesove analize:

- Kendall-ov $W = 0.045$ (20 puta niži od izvrsnog)
- P-vrijednost = 0.451 (sustav se ne razlikuje od nasumičnog)
- 45.3% recenzija ima sve maksimalne ocjene

5 Procjena broja recenzenata



Slika 7: Distribucija procjena broja recenzenata različitim metodama. Medijan = 102 recenzenta.

Rezultat: Procijenjeno je da je približno 102 recenzenta ocijenilo 95 projekata (prosječno 2.8 projekata po recenzentu).

6 Glavni zaključci

6.1 Totalna nefunkcionalnost sustava

1. Sustav ne mjeri kvalitetu projekata

- Samo 6-9% varijance povezano s projektima
- 91-94% varijance zbog recenzenata i slučajnosti

2. Ekstremna pristranost prema visokim ocjenama

- 94% ocjena ≥ 4
- 70% svih ocjena = 5
- Ceiling učinak onemogućava diskriminaciju

3. Nedostatak slaganja među recenzentima

- ICC < 0.2 (daleko ispod prihvatljivog)
- 30% projekata ima negativne ICC vrijednosti
- Razlike do 8 bodova za isti projekt

4. Statistička nerazlučivost od nasumičnosti

- $p = 0.451$ u usporedbi s nasumičnim ocjenjivanjem
- Sustav nije bolji od bacanja kocke

6.2 Preporuke

1. **Hitno:** Rezultati ovog vrednovanja moraju biti odbačeni kao nevaljani
2. **Kratkoročno:** Potpuni redizajn sustava s:
 - Kalibracijom recenzenata
 - Jasnim kriterijima
 - Kontrolom kvalitete
3. **Dugoročno:** Razmotriti alternativne sustave (npr. ponderirana lutrija)

7 Završna poruka

Ovaj sustav vrednovanja predstavlja potpuni kolaps znanstvenih standarda objektivnosti i pravednosti. Sustav koji samo 6.4% do najviše 9% svoje varijance temelji na stvarnim razlikama između evaluiranih objekata je fundamentalno nefunkcionalan i mora biti hitno reformiran.

Stoga ovaj sustav i rezultate vrednovanja treba u potpuosti odbaciti i dodijeliti doktorante SVIM predlagateljima u ovom krugu!

WWW verzija dostupna na poveznici <https://branimir-k-hackenberger.github.io/>