

# Introduction to Data Mining

[Home](#) / [Courses](#) / [Undergraduate Programs](#) / [University study Computer and Information Science](#) / [3rd year](#) / [Information systems](#) / [uozp](#)  
/ General / [5. domača naloga: priporočilni sistemi](#)

## 5. domača naloga: priporočilni sistemi

Na voljo imate podatke o ocenah glasbenih izvajalcev s strani poslušalcev. Vaša naloga je, da na podlagi podatkov zgradite priporočilni sistem, s katerim boste na podlagi profila poslušalca rangirali oziroma priporočali glasbene izvajalce. Naloga je podana v obliki [tekmovalnja](#), kjer je cilj za dane pare izvajalcev in poslušalcev čim bolj natančno napovedati ocene.

### Naloga:

- [50 %] Zgradite priporočilni sistem z metodo skupinskega filtriranja. Priporočilni sistem lahko zgradite glede na podobnost (s tako implementacijo pridobite največ 30% točk) ali z uporabo matričnega razcepa (pravilna implementacija te metode vam prinese 50% točk). Priporočilni sistem naj bo implementiran v objektu, ki ob inicializaciji sprejme učno množico, zgradi model, in implementira funkcijo, ki za dano kombinacijo uporabnika in izvajalca napove uporabnikovo oceno.
- [30 %] Vaš sistem ovrednotite na podatkih o ocenah izvajalcev, ki so vam na voljo. In sicer tako, da oblikujete učno množico (70% podatkov) in testno množico (preostalih 30% podatkov). Priporočilni sistem zgradite na učni množici in testni množici oceniti točnost z oceno RMSE (Root mean squared error). Postopek ponovite nekajkrat (recimo 5-krat) ter poročajte o povprečni točnosti.
- [20 %] Iz seznama vseh izvajalcev izberite njih 20, ki jih poznate. Vaša izbira naj bo čimbolj raznolika. Ocenite jih z oceno od 0 do 10, kjer 0 pomeni, da izvajalca nebi želeli poslušati, 10 pa, da bi izvajalca z veseljem poslušali. Na podlagi ocen, s pomočjo vašega priporočilnega sistema, predlagajte nove izvajalce. Komentirajte predloge.

### Podatki:

Vsebujejo 1892 uporabnikov, 17632 izvajalcev, 83550 ocen v učni množici, 9284 parov uporabnik-izvajalec v testni množici na strežniku.

Podatki so na voljo na [tekmovalnem strežniku](#).

### Datoteke:

- artists.dat:** Vsebuje ime izvajalca ter povezavo do njegovega profila na Last.fm
- user\_artists\_training.dat:** Učna množica, ki vsebuje povezave med uporabnikom in izvajalcem v obliki ocene izvajalca od 0 do 10.
- user\_artists\_test.dat:** Testna množica, ki vsebuje povezave med uporabnikom in izvajalcem v obliki ocene izvajalca od 0 do 10.

### Za izboljšanje napovedi lahko uporabite tudi:

- user\_friends.dat:** Označuje prijateljstva med uporabniki na strani Last.fm
- user\_taggedartists.dat:** Določa oznake (tage), ki so jih uporabniki dali izvajalcem
- tags.dat:** Določa imena oznak

### Poročilo naj vsebuje naslednja poglavja:

- Uvod** (enostavni opis cilja domače naloge)
- Opis metod**, kjer opišete metode priprave podatkov in napovedovanja, ki ste jih za oblikovanje napovedi uporabili. Vsako metodo poimenujte (akronim), njen opis pa naj ne bo daljši od 5 vrstic. Te opise postavite v latex okolju `description (\item[ime] Opis.)`. Predvidevamo, da boste razvili nekaj metod, katerih rezultate boste preverili na strežniku. Med njimi izberite do tri najbolj zanimive ali najbolj točne, in te vključite v poročilo o domači nalogi.
- Rezultati**, ki vsebuje tabelo z rezultati (ime metode, ocena s preverjanjem na učnih podatkih, ocena na tekmovalnem strežniku). V tabeli z znakom "\*" označite končno oddajo. Komentirajte predloge, ki vam jih poda priporočilni sistem (tretji del naloge).

### Oddaja:

Na [strani tekmovalnja](#) oddajte vmesne in končne napovedi (na voljo imate štiri oddaje na strežnik na dan), na spletni učilnici pa poročilo in izvorno kodo. Poročilo mora biti napisano s predpisano predlogo. V posameznem dnevu lahko oddate največ deset napovedi testnih podatkov.

Dodatne točke:

- 1. mesto na tekmovalnem strežniku: +50%.
- 2.-5. mesto: +35%.
- 6.-10. mesto: +25%.
- 11.-15. mesto: +15%
- 15.-20. mesto: +10%
- Povečanje natančnosti napovedi z uporabo podatkov iz user\_friends.dat, user\_taggedartists.dat: do + 20 %

Štejejo mesta na tekmovalnem strežniku v trenutku zaključka domača naloge. Kasnejših oddaj ne bomo upoštevali.

[◀ 4. domača naloga: logistična regresija](#)

Jump to...

You are currently using guest access (Log in)

uozp

Get the mobile app