

# Sistem za preporuku Steam igrice

Rajnović Teodora

Fakultet Tehničkih Nauka – Novi Sad



## Uvod

Glavni problem projekta je preporuka igrice sa platforme Steam korisniku na osnovu korisnikovih prethodno igranih ili kupljenih igrice. Za rešenje problema korišćen je algoritam kolaborativnog filtriranja baziran na proizvodu.

## Rešenje Problema

Algoritam generiše preporuke tako što za kupljenu ili igranu igricu pronade njoj najbližnju igricu i preporuči je korisniku.

Za izračunavanje sličnosti igrice korišćene su informacije o: datumu izlaska igrice, developeru, izdavaču, popularnim tagovima, detaljima o igrici i žanru. Navedene informacije se konvertuju u numeričke vrednosti i najčešće korišćene engleske reči se uklanjaju iz teksta.

Za konvertovanje teksta u numeričke vrednosti korišćen je CountVectorizer iz biblioteke sklearn ( [link](#) ). Dalje se dobijene numeričke vrednosti koriste za izračunavanje matrice sličnosti pomoću cosine\_similarity metode iz biblioteke sklearn ( [link](#) ).

Zadnji korak je da se pomoću dobijene matrice sličnosti generišu preporuke za igrice koje su slične onima koje je korisnik prethodno kupio ili igrao. Za ovaj korak korišćena je weighted sum metoda za predikciju nedostajućih sati u user-item matici.

## Obrada Podataka

Kako bi se smanjio broj redova u tabeli korisnika, za svakog korisnika koji je kupio i igrao određenu igricu izbrisan je red gde je igrice kupljena, a ostavljen je red sati igranja igrice. Ako je igrice samo kupljena, sati su stavljeni na 0.1. Uklonjene su nepotrebne kolone i nisu korišćeni korisnici koji su igrali samo jednu igricu.

Kod tabele igrice uklonjeni su duplikati, igrice sa lošim ocenama i nepotrebne kolone.

Broj jedinstvenih korisnika je 12393 i igrice 27487.

## Rezultati

Za evaluaciju rešenja je korišćeno 150 nasumično odabranih korisnika iz dataseta.

Za evaluaciju su korišćeni MAE i RMSE. MAE poredi odstupanja predviđenih sati igranja i stvarnih sati ciljanog korisnika. Što je manja razlika između predviđenih sati i stvarnih, to je bolje predviđanje. RMSE je sličan MAE, jedina razlika je što daje veću važnost velikim greškama.

MAE i RMSE su date za pojedine korisnike zbog velikih razlika u satima igranja između korisnika. U nastavku je data tabela za različite korisnike za N = 20.

UserID	MAE	RMSE
159077305	8.46	12.34
160461131	0.51	0.61
59234901	190.01	241.06
96504332	5.17	5.17

## Zaključak

Korisnicima su preporučene igrice za koje model predviđa najveće sate igranja. Za N najbližnjih igrice trenutnoj igrici koju je korisnik igrao računa se weighted sum i od N igrice vrši se odabir jedne za preporuku. Rezultati nisu mnogo varirali za različito N.

Za igrice koje se mogu preporučiti korišćene su one sa najboljim ili mešovitim ocenama i to je predstavilo bolje rešenje u odnosu na samo najbolje ocenjene igrice.

Rezultat varira u zavisnosti od količine igrice koje korisnik ima, za mali broj igrice daje gore rezultate nego sa većim brojem igranih igrice. Ako su sati igranja nekonzistentni MAE i RMSE će davati loš rezultat.

## Korišćene Reference

[Evaluating collaborative filtering recommender systems\(researchgate.net\)](#)

[Calculating String Similarity in Python | by Dario Radečić](#)

[Item-to-Item Based Collaborative Filtering - GeeksforGeeks](#)