SVEUČILIŠTE JOSIPA JURJA STROSSMAYERA U OSIJEKU

FAKULTET ELEKTROTEHNIKE, RAČUNARSTVA I

INFORMACIJSKIH TEHNOLOGIJA OSIJEK

Sveučilišni studij

Sustav preporuka za video igre

Projektni zadatak

Marko Budak

Osijek, 2024.

Sadržaj

[1. UVOD 3](#_Toc158288673)

[2. OPIS KORIŠTENIH SKUPOVA PODATAKA 4](#_Toc158288674)

[2.1. Filtriranje na temelju sadržaja 4](#_Toc158288675)

[2.1.1 Eksplorativna analiza 4](#_Toc158288676)

[2.1.2 Predobrada podataka 5](#_Toc158288677)

[2.2. Suradničko filtriranje 7](#_Toc158288678)

[2.2.1 Eksplorativna analiza 7](#_Toc158288679)

[2.2.2 Predobrada podataka 8](#_Toc158288680)

[3. OPIS KORIŠTENIH METODA/ALGORITAMA 9](#_Toc158288681)

[4. EVALUACIJA IZGRAĐENIH MODELA/ALGORITAMA 11](#_Toc158288682)

[5. DEMONSTRACIJA RADA 12](#_Toc158288683)

[5.1 Filtriranje na temelju sadržaja 12](#_Toc158288684)

[5.1.1 Izgled sučelja 12](#_Toc158288685)

[5.1.2 Primjer korištenja 13](#_Toc158288686)

[5.1.3 Prikaz rezultata 13](#_Toc158288687)

[5.2 Suradničko filtriranje 16](#_Toc158288688)

[5.2.1 Izgled sučelja 16](#_Toc158288689)

[5.2.2 Primjer korištenja 16](#_Toc158288690)

[5.2.3 Prikaz rezultata 17](#_Toc158288691)

[6. ZAKLJUČAK 18](#_Toc158288692)

# 1. UVOD

Cilj ovog projektnog zadatka je kreiranje sustava za preporuke video igara. Motivacija za pravljenje ovog projekta je da većina osoba koja igra video igre dolazi u period kad jednostavno ne znaju što više igrati. Ova web aplikacija je napravljena da riješi taj problem, jer daje korisnicima mogućnost da dobiju preporuku novih igara za igrati prema dosada odigranoj igri. Postoje različiti principi rješavanja ovog problema poput:

* Suradničko filtriranje: Ovaj pristup preporučuje stavke (u ovom slučaju igre) korisnicima na temelju preferencija drugih korisnika koji imaju sličan ukus. Ne zahtijeva eksplicitne oznake za ciljne varijable, ali se oslanja na sličnost između korisnika ili stavki.
* Filtriranje na temelju sadržaja: Ova metoda preporučuje stavke korisnicima na temelju atributa stavki i korisničkih preferencija. Analizira sadržaj stavki (npr. oznake igara, opise) i povijesne preferencije korisnika kako bi dao preporuke.
* Hibridne metode: hibridni sustavi preporuka kombiniraju suradničko filtriranje i pristupe filtriranja temeljene na sadržaju kako bi pružili preciznije i raznolikije preporuke. Ovi sustavi iskorištavaju interakcije korisnika i stavki i atribute stavki za generiranje preporuka.
* Implicitne povratne informacije: U sustavima za preporuku igara, interakcije korisnika s igrama (npr. vrijeme igranja, povijest kupovine, recenzije) mogu poslužiti kao implicitni povratni signali. Ti se signali mogu koristiti za zaključivanje korisničkih preferencija i generiranje personaliziranih preporuka.

Steam je najveća platforma za prodaju igara te su korišteni njegovi podatkovni skupovi.

Za ovaj projekt su napravljena dva za sustava za preporuku, jedan sa suradničkim filtriranjem, a drugi sa filtriranjem na temelju sadržaja.

# 2. OPIS KORIŠTENIH SKUPOVA PODATAKA

Korišteni podatkovni skupovi pronađeni su na Kaggle, pod nazivima „Steam Video Games“ i „Steam Store Games (Clean dataset)“. Iz navedenog podatkovnog skupa obrisani su stupci poput „Broj kupaca“ jer ti podatci nisu korišteni, a podatkovni skup je već sam po sebi jako velik.

## 2.1. Filtriranje na temelju sadržaja

### 2.1.1 Eksplorativna analiza

Analiza podatkovnog skupa napravljena je u Google Collab-u radi jasnijeg prikaza analize.

Skup se sastoji od 13 stupaca. Stupac „appid“ predstavlja unikatni identifikacijski broj svake igre prema kojemu možemo dohvatiti ostale podatke o igri. Stupci „name“, „release\_date“, „developer“, „publisher“, „genres“, „steamspy\_tags“, „price“, „positive\_ratings“, „negative\_ratings“ i „short\_description“ su stupci koji su razumljivi samo po sebi i prikazuju nam određene podatke igara. Stupac „score“ je omjer pozitivnih i negativnih ocjena koje kod računa i ubacuje kao novi stupac u podatkovni skup.

A screenshot of a computer

Description automatically generated

Slika 1: Informacije skupa

A screenshot of a computer code

Description automatically generated

Slika 2: Dodavanje stupca za ocjenu igre

### 2.1.2 Predobrada podataka

Zbog velike veličine skupa, za očekivati je da će skup sadržavati null vrijednosti.

Dolje navedenim kodom prvo provjeravamo koji stupci imaju te null vrijednosti te koliko null vrijednosti svaki stupac ima.

A screenshot of a computer

Description automatically generated

Slika 3: Provjera null vrijednosti

Zbog principa rada sustava za preporuku, opisne riječi igre i izvođača igara spajamo u jedan stupac.

Sustav preporuka će raditi baziran na imenu razvojne firme video igre i na opisnim riječima svake igre. Probalo se također raditi i sa imenima igara i kratkim opisima svake igre, no to je smanjilo točnost preporuka te je također i uzrokovalo mnoga crashanja algoritma jer TF-IDF i vektoriziranje matrice nije moglo s toliko podataka raditi.

A screen shot of a computer code

Description automatically generated

Slika 4: Kombinirani stupac

Pomoću samo jedne linije koda iz podatkovnog skupa možemo ukloniti sve null vrijednosti unutar svakog stupca.

A screenshot of a computer program

Description automatically generated

Slika 5: Brisanje null vrijednosti

## 2.2. Suradničko filtriranje

### 2.2.1 Eksplorativna analiza

Skup se sastoji 5 stupaca, no nazivi stupaca nam uopće ne ukazuju na što ti stupci predstavljaju. Ovaj problem je riješen jednostavnim dodavanjem naziva za svaki stupac.

A screenshot of a computer

Description automatically generated

Slika 6: Informacije skupa

A screenshot of a computer

Description automatically generated

Slika 7: Izmjena imena stupaca

### 2.2.2 Predobrada podataka

Zadnji stupac nam ne predstavlja nikakvu vrijednost pa je stoga uklonjen. Također su se obrisali svi duplikati korisnika i igara te smo se riješili svih null vrijednosti podatkovnog skupa.

A screen shot of a computer

Description automatically generated

Slika 8: Brisanje zadnjeg stupca, duplikata i null vrijednosti.

# 3. OPIS KORIŠTENIH METODA/ALGORITAMA

Kao i prije navedeno, koriste se filtriranje na temelju sadržaja i suradničko filtriranje.

Filtriranje na temelju sadržaja (Content-based filtering) je tehnika preporuka koja koristi karakteristike proizvoda ili stavki kako bi predložila slične stavke korisnicima. Općeniti koraci algoritma za filtriranje na temelju sadržaja su:

1. Prikupljanje podataka: Prikupe se podaci o stavkama koje se preporučuju (npr. filmovi, knjige, proizvodi) i njihove karakteristike. Karakteristike mogu uključivati ključne riječi, žanrove, ocjene korisnika, opise i slično.
2. Izgradnja profila korisnika: Za svakog korisnika izračunaju se preferencije na temelju stavki koje su im se svidjele u prošlosti. To se može postići analizom sadržaja stavki koje su korisnici ocijenili ili konzumirali.
3. Izračunavanje sličnosti: Koristi se metrika sličnosti (npr. kosinusna sličnost, Jaccardova sličnost) kako bi se utvrdilo koliko su stavke slične jedna drugoj na temelju njihovih karakteristika.
4. Izrada preporuka: Za svakog korisnika, algoritam identificira stavke koje su slične onima koje su im se svidjele u prošlosti. Preporučuju se stavke s najvećom mjerom sličnosti.
5. Prilagođavanje preporuka: Preporuke se mogu prilagoditi korisničkim preferencijama ili kontekstu. Na primjer, ako korisnik preferira određeni žanr, algoritam može dati prednost stavkama tog žanra.
6. Ažuriranje modela: Periodično se ažuriraju podaci o stavkama i profilima korisnika kako bi se algoritam prilagodio promjenama u preferencijama korisnika ili novim stavkama koje se dodaju.
7. Evaluacija performansi: Performanse algoritma mogu se evaluirati kroz metrike kao što su preciznost, odziv, F-mjera ili korisničko zadovoljstvo putem povratnih informacija korisnika.

Suradničko filtriranje je tehnika koja se koristi u sustavima preporuka kako bi se korisnicima preporučili proizvodi, usluge ili sadržaji koji bi ih mogli zanimati. Ova tehnika se temelji na ideji da će korisnicima biti zanimljivi proizvodi koje su prethodno voljeli drugi korisnici sličnog ukusa ili interesa. Umjesto analize samih stavki ili njihovih karakteristika, suradničko filtriranje analizira ponašanje korisnika i koristi tu informaciju za generiranje preporuka. Postoje dvije osnovne metode suradničkog filtriranja:

1. Suradničko filtriranje temeljeno na korisnicima: Ovaj pristup koristi se za pronalaženje sličnosti između korisnika. Sustav analizira ponašanje korisnika kako bi pronašao druge korisnike sličnih interesa ili preferencija. Kada se identificiraju slični korisnici, sustav preporučuje proizvode ili stavke koje su se svidjele tim sličnim korisnicima, a koje su možda korisniku u pitanju nepoznate.
2. Suradničko filtriranje temeljeno na stavkama: Ovaj pristup usredotočuje se na sličnost između stavki, a ne korisnika. Sustav analizira ponašanje korisnika kako bi identificirao stavke koje su se svidjele istim korisničkim skupinama. Na temelju tih informacija, sustav preporučuje slične stavke korisnicima koji su pokazali interes za određene stavke.

Ove tehnike često koriste algoritme strojnog učenja, poput k-najbližih susjeda (k-NN) ili faktorizacije matrica, kako bi identificirali slične korisnike ili stavke. Suradničko filtriranje je popularna metoda u sustavima preporuka jer ne zahtijeva duboko razumijevanje stavki ili korisnika, već se oslanja na uzorke ponašanja korisnika kako bi generirao relevantne preporuke.

# 4. EVALUACIJA IZGRAĐENIH MODELA/ALGORITAMA

Sustav preporuka igara nije moguće evaluirati pomoću točnosti, preciznosti, odziva i sl. jer je nam performanse modela ovise o ljudskoj percepciji, mišljenjima, subjektivnom mišljenju i sl., pa je stoga evaluacija odrađena pomoću ankete. Anketa provjera je li korisnik mogao normalno pokrenuti aplikaciju, je li mu aplikacija preporučila igre, njihovo zadovoljstvo preporukama te opcija da sami unesi kako bi poboljšali sustav. Uspio sam dobiti povratnu informaciju od 12 korisnika.

A screenshot of a computer

Description automatically generated

Slika 9: Anketa korisničkih povratnih informacija

# 5. DEMONSTRACIJA RADA

## 5.1 Filtriranje na temelju sadržaja

### 5.1.1 Izgled sučelja

Izgled sučelja se prikazuje pomoću streamlit-a. Streamlit je otvoreni alat za brzo razvojno sučelje koji omogućuje lako kreiranje interaktivnih web aplikacija koristeći Python. Omogućuje programerima da brzo pretvore svoje Python skripte u funkcionalne web aplikacije s minimalnim naporom, što ga čini popularnim alatom u područjima poput znanstvenog računanja, analize podataka i strojnog učenja. Korisnik upisuje igrica prema kojoj želi preporuke. Unutar algoritma napisana je funkcija koja dopušta korisniku da iako krivo unese ime igre, algoritam će prepoznati o kojoj igri se radi. Funkcija je napravljena pomoću biblioteke fuzzywuzzy. Korisnik također može unijeti ključnu riječ, da ako npr. korisnik želi samo „Batman“ igre, popunjavanjem te ključne riječi ostvariti će to. Prozor se također može ostaviti prazan. Također postoji i opcija unošenja minimalne godine izlaske i minimalne ocijene igre ako korisnik ne želi starije ili loše igre. Nakon popunjavanja, klikom na tipku „Recommend Games“ korisniku će biti preporučeno 5 igara.

A screenshot of a computer

Description automatically generated

Slika 10: Izgled sučelja

### 5.1.2 Primjer korištenja

Primjer korištenja gdje korisnik kao naslov igre prema kojoj želi biti preporučen je „Batman arkham city game of the year editio“, ključna riječ je „Batman“ jer korisnik želi samo batman igre, godina izdanja 2002. dok je minimalna ocjena igre ostavljena netaknuta.

A screenshot of a computer

Description automatically generated

Slika 11: Primjer korištenja

### 5.1.3 Prikaz rezultata

Korisniku se prikazuje 5 preporučenih igara. Pored svake igre piše ime, postotak sličnosti između originale i preporučene igre, kratki opis igre, ime razvojne firme video igre, cijena, platforme na kojima se igra može igrati, žanrovi igre, ocjena igre koja je od 1 do 100 prema korisničkim ocjenama te datum izlaska igrice na tržište.

A screenshot of a computer

Description automatically generated

Slika 12: Prikaz ispisa 1/2

A screenshot of a computer

Description automatically generated

Slika 13: Prikaz ispisa 2/2

## 5.2 Suradničko filtriranje

### 5.2.1 Izgled sučelja

Poput filtriranja na temelju sadržaja, koristi se streamlit za prikazivanje sučelja te je također unutar algoritma napisana funkcija koja dopušta korisniku da iako krivo unese ime igre, sustav prepozna o kojoj igri se radi. Funkcija je napravljena pomoću biblioteke fuzzywuzzy.

A screenshot of a computer

Description automatically generated

Slika 14: Izgled sučelja

### 5.2.2 Primjer korištenja

Primjer korištenja gdje korisnik kao naslov igre prema kojoj želi biti preporučen je „The witcher 3“.

A black background with white text

Description automatically generated

Slika 15: Primjer korištenja

### 5.2.3 Prikaz rezultata

Korisniku se prikazuje 10 preporučenih igara. Pored svake preporučene igre nalazi se ime igre te postotak sličnosti u ocjenjivanju igre s drugim korisnicima.

A screenshot of a computer program

Description automatically generated

Slika 16: Prikaz rezultata

# 6. ZAKLJUČAK

U okviru projektnog zadatka, napravljena su dva algoritma za preporuku igara, suradničko filtriranje i filtriranje na temelju sadržaja. Korisničko sučelje smo postigli pomoću streamlit web aplikacije gdje korisnik može unijeti sve dostupne podatke. Postoje razni načini unaprjeđenja, poput korištenja hibridnih metoda kako bismo povećali postotak sličnosti ili ubrzali sveukupni proces preporuke. Hibridna metoda koristi obje prijašnje navedene algoritme kako bi korisniku što kvalitetnije preporučili igre.