**UNIVERZITET U ZENICI**

**POLITEHNIČKI FAKULTET**

**SEMINARSKI RAD IZ PREDMETA:**

Vještačka inteligencija

|  |  |
| --- | --- |
| **Tema rada:** | **Generisanje sistema preporuke za filmove** |

|  |  |
| --- | --- |
| **Predmetni nastavnik:** | van. prof. dr. Nermin Goran |

|  |  |
| --- | --- |
| **Student:** | Ajdin Bukvić |
| Broj indeksa: | II-89 |
| Usmjerenje: | Softversko inženjerstvo |
| Godina studija: | 1. godina, 2. ciklus |
| **Rezultat rada:** | Analiza skupa podataka i implementacija različitih metoda i algoritama vještačke inteligencije za generisanje sistema preporuke za filmove, uz poređenje dobijenih rezultata i pregled performansi |

Datum: 24.01.2024.

**SADRŽAJ**

[1. UVOD 2](#_Toc156798204)

[2. SISTEMI PREPORUKE 3](#_Toc156798205)

[2.1. Filtriranje zasnovano na sadržaju 4](#_Toc156798208)

[2.2. Kolaborativno filtriranje 5](#_Toc156798215)

[2.3. Ostali tipovi sistema preporučivanja 6](#_Toc156798220)

[3 PRAKTIČNI DIO 7](#_Toc156798221)

[3.1 Alati, biblioteke i radno okruženje 7](#_Toc156798222)

[3.2 Skup podataka (dataset) 8](#_Toc156798223)

[3.3 Priprema i analiza podataka 8](#_Toc156798224)

[4 IMPLEMENTACIJA 11](#_Toc156798225)

[4.1 Model neuronske mreže 11](#_Toc156798226)

[4.2 TF-IDF algoritam 14](#_Toc156798227)

# UVOD

U zadnjih nekoliko godina područje vještačke inteligencije je napravilo veliku ekspanziju u raznim sferama ljudskog života. Pošto se tehnologija sve brže razvija, u mnogim djelatnostima dolazi do potrebe za primjenom vještačke inteligencije. Upotreba vještačke inteligencije može znatno unaprijediti svakodnevne životne procese, kako u privatnom, tako i u poslovnom smislu. Mnoge grane nauke i tehnike se danas u visokoj mjeri oslanjaju na vještačku inteligenciju, kao što su auto industrija, medicina, vojna industrija, programiranje i druge. Iako su ovo relativno novi pristupi, daju znatna unaprijeđenja i mogu ubrzati sam proces rada. Kroz ovaj seminarski rad će biti prikazan primjer upotrebe vještačke inteligencije za generisanje sistema preporuke za filmove. Sistemi preporuke su posebna grana vještačke inteligencije koji koriste velike skupove podataka, kako bi korištenjem različitih algoritama predvidjeli ili klasifikovali, korisničke podatke i dali mu što precizniju sugestiju onoga što on inače preferira.

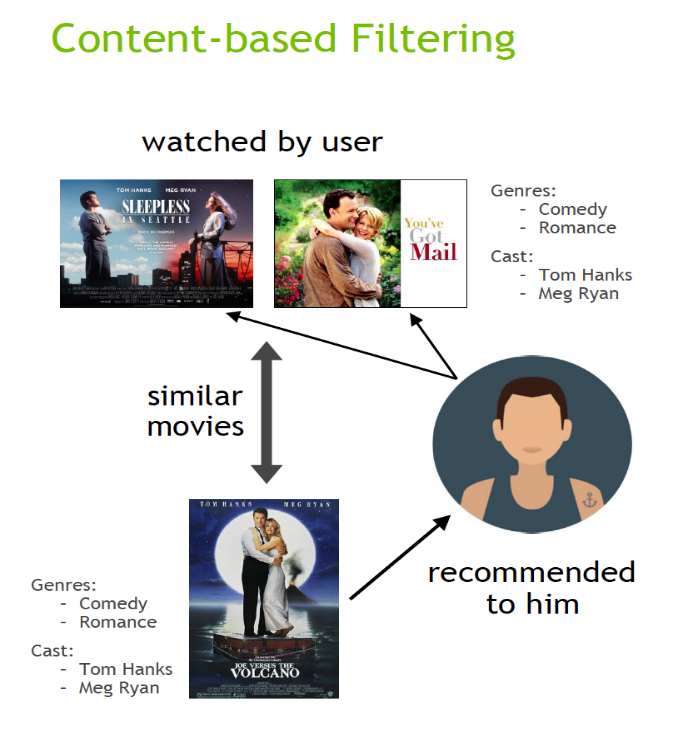
# SISTEMI PREPORUKE

Sistemi preporuke su bazirani na praćenju ponašanja pri pregledavanju i historiji pregledavanja od strane korisnika kako bi kreirali različite obrasce i pronašli sličnosti u podacima. Na osnovu toga se korisniku preporučuje sadržaj koji bi ga mogao zanimati. Sistemi preporuke se koriste svugdje gdje bi mogli unaprijediti korisničko iskustvo. Poznato je da sve velike kompanije i organizacije koriste ovaj pristup kako bi svojim članovima ili kupcima preporučili sadržaje koje ih zanimaju, a sve u svrhu povećanja profita i zadržavanja korisnika. Najjednostavniji pristup na kojem rade ovi sistemi jest da korisnik nakon korištenja nekog softvera ili sistema, ostavlja dosta svojih privatnih podataka, o svojim navikama ili proizvodima i uslugama koji ga interesuju. Organizacije zatim koriste ove prikupljene podatke kako bi iz tih podataka izvukli relevantne podatke i igradili jedinstveni opis korisnika, koji im pomaže da poboljšaju svoju ponudu na temelju stečenih saznanja. Neke od vodećih kompanija koje koriste ove sisteme u svojim aplikacijama su Netflix, Amazon, Spotify, Google (YouTube), Pandora, Apple Music i mnogi drugi. Podaci koji su prikupljeni, te koji se koriste se još nazivaju i „Big Data“. To su veliki skupovi podataka, višemilionski, koji služe kako bi filtrirali različite korisnike, proizvode, usluge, te ih međusobno uparili na osnovu sličnosti. Faktori koji su sastavni dijelovi ovih podataka, mogu varirati od različitih tipova informacija, kao što je već rečeno historija pregledavanja ili kupovine, ali i dosta dublji podaci koji mogu dati preciznije rezultate, koji uključuju demografske podatke. Ovim pristupom se mogu postići preporuke, koje se ne odnose samo na korisnikove lične preferencije, nego i na njegov status, porijeklo, kulturu, jezik i mnoge druge parametre do kojih se može doći. Na ovaj način korisnik može doći i do proizvoda ili usluga koje inače ne bi sam pronašao, ali sistem može zaključiti da bi mu se takvi rezultati preporuke mogli svidjeti. S druge strane kada su u pitanju sami proizvodi ili usluge, za njihovo filtriranje i klasifikaciju koriste se parametri kao što su broj pregleda, klikova, lajkova, komentara, recenzije ili kupovina. Područja primjene sistema za preporuku su široko rasprostranjena u različitim industrijama, koje uključuju lance za nabavku, e-trgovinu, medije i zabavu, društvene mreže, finansijske usluge, putovanja i ugostiteljstvo. Prednosti korištenja sistema preporuke u aplikacijama i sistemima mogu značajno povećati prihode, kao i broj pregleda (klikova), te razne druge metrike. Ključni faktor je zadovoljstvo i zadržavanje već postojećih članova i kupaca.



## Filtriranje zasnovano na sadržaju

Filtriranje zasnovano na sadržaju je jedan od tipova za generisanje sistema preporuke, koji se kako sam naziv kaže bazira na samom sadržaju. Pitanje je šta u konkretnom kontektstu predstavlja „sadržaj“. Sadržaj su zapravo stavke ili karakteristike pojedinačnog proizvoda ili usluge. Ovaj princip uzima sve karakteristike nekog proizvoda ili usluge i upoređuje ih s onim već poznatim podacima (proizvodima ili uslugama), s kojima je korisnik već ranije imao neku interakciju. Naprimjer, ako se uzmu u obzir korisnikove osobine, broj godina, spol, status i drugi specifični atributi, ovaj tip će pronaći sve proizvode ili usluge koji odgovaraju gore navedenim uslovima. U konkretnom primjeru, ako se radi o muškoj osobi od 20. godina koja je student i dolazi iz Zenice, sistem bi mogao preporučiti samo proizvode koji su namjenjeni mlađim osobama (između 18 i 25 godina), muškog spola, koje pritom studiraju, a da su dostupni u njihovoj radnji u Zenici. Posebno, ako je već od ranije poznato da je korisnik zainteresovan za ovakav tip proizvoda, obzirom da je već jednom kupovao slične proizvode. Naravno, ovo je samo grubi (naivni) primjer primjene ovog tipa preporuke. U praksi bi ovaj postupak bio znatno napredniji i kompleksniji. Glavni segment kod ovog tipa preporuke jeste izdvajanje ključnih karakteristika iz samog sadržaja. Modeliranje ovog sistema zasniva se na određivanju relevantnih atributa, koji bi mogli biti zanimljivi za proučavanje. Jedan od klasičnih primjera primjene ovog tipa je model baziran na ključnim riječima, koji se još naziva i model vektorskog prostora s ponderom frekvenzije inverzne frekvencije dokumenta. Ovaj algoritam će biti detaljnije objašnjen u nastavku rada.

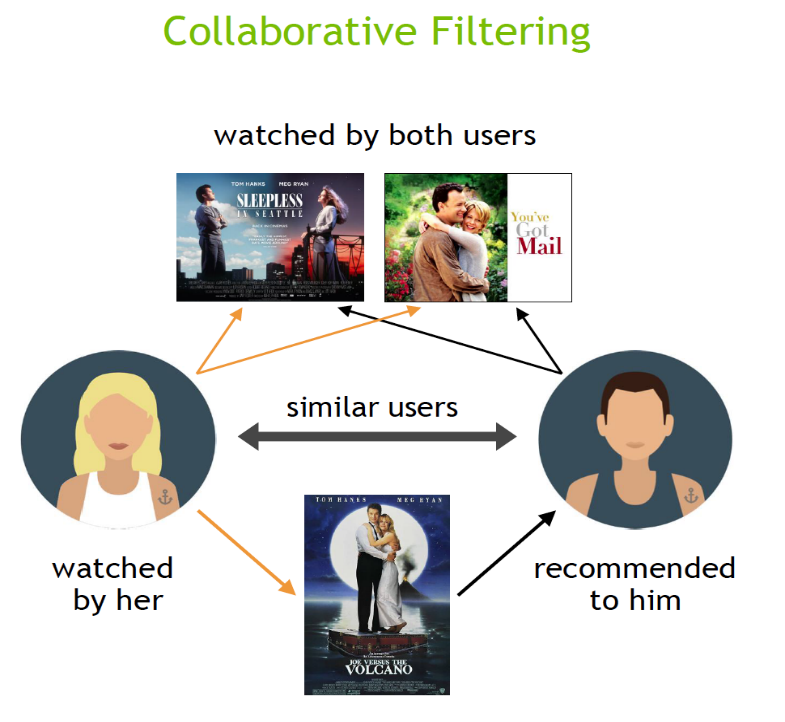


Slika 1



## Kolaborativno filtriranje

Jedan od najefikasnijih tipova koji se koriste u sistemima preporuka je kolaborativno filtriranje. Metoda kolaborativnog filtriranja se zasniva na analizi ponašanja korisnika, te pronalasku sličnosti u njihovim preferiranim izborima. U ovom pristupu u obzir se uzima koncept koji pokazuje da se će ljudi koji su u prošlosti pokazali slično ponašanje (preferencije), isto tako pokazati i u budućnosti (vjerovatno). Sam naziv ovog tipa sastoji se od dva dijela: kolaborativno, koje se odnosi na „saradnju“ (prepoznavanje sličnosti između korisnika), te filtriranje, odnosno preporučivanje stavki (sadržaja, proizvoda, usluga) na osnovu prikupljenih informacija o korisnicima. Interakcije između korisnika i stavki, odnosi se na kreiranje modela koji će biti u stanju predvidjeti buduće interakcije korisnika, na osnovu onih koje su imali slični korisnici. Sličnosti između korisnika mogu biti određene, naprimjer proizvodima koje su neki korisnici kupili ili ocjenama koje su korisnici ostavili na neki proizvod. Jednostavan primjer, jeste da ako postoje korisnici koji dijele slične ukuse prema hrani, tako da su već ranije kupovali i ocjenjivali proizvode od iste kompanije. Naprimjer, korisnici su probali neki broj istih proizvoda, te podijeli mišljenje da su ti proizvodi za njih zaista prihvatljivi (ukusni, zanimljivi ili bilo koji drugi opisni atribut). Tako će sistem vrlo lahko ove korisnike označiti, kao „slične“. Naredni korak jeste da, ako postoji neki proizvod koji je jedan korisnik probao (ocijenio, kupio), a da drugi korisnik nije. Ovdje ovaj sistem dolazi do izražaja tako što se „pretpostavlja“ da će se ovaj proizvod svidjeti i drugom korisniku, te mu se zbog toga i preporučuje. Također, kao i u prethodnom tipu, ovo je samo jedan banalan primjer, s dva korisnika i nekoliko proizvoda, ali se u realnosti radi o velikom broju različitih korisnika i proizvoda. Ovaj tip preporuke se još naziva i „korisnik-korisnik“ (user-user) kolaborativno filtriranje, jer se zasniva na sličnostima samih korisnika. Također, postoji i drugi tip koji se naziva „stavka-stavka“ (item-item) koji se odnosi na sličnost između dvije stavke (šta god da one predstavljaju). Analogno prvom primjeru, drugi tip određuje sličnosti između stavki, tako da se neka stavka preporučuje korisniku koji je već ranije pokazao interes za slične stavke. Postoje razne metode i metrike, za mjerenje „sličnosti“ između korisnika (ili stavki), koje se koriste za predikciju rezultata. Najčešći pristup je kreiranje matrica korisnik-stavka-korisnik ili stavka-korisnik-stavka. Preporuke nastale ovim pristupom je potrebno još dodatno filtrirati, dodavanjem nekih težinskih vrijednosti, kao što su srednje ocjene, kako bi se izvršila normalizacija rezultata. Primjene ovih algoritama će biti prikazane u nastavku rada.



Slika 2



## Ostali tipovi sistema preporučivanja

Postoje još neki tipovi na kojima se temelje sistemi preporučivanja, koji će ovdje biti samo ukratko predstavljeni. Poseban tip je hibridno filtriranje, koje kobinuje koncepte prethodno objašnjenih tipova i još nekih dodatnih tipova. Pošto svaki od ovih tipova, sadrži određeni broj problema, ova metoda kombinira sve prednosti i uzima pojedinačne segmente ovih tipova, kako bi se kreirao napredniji sistem, koji prevazilazi njihove nedostatke. Na ovaj način se postižu bolje performanse i dosta bolji i precizniji rezultati. Naprimjer, ako kolaborativnom filtriranju nedostaju informacije o zavisnostima korisnika (ili stavki), dok filtriranju zasnovanom na sadržaju nedostaju informacije o preferencijama korisnika, hibridni sistem će iskoristiti poznate podatke iz oba tipa kako bi generisao preporuke. Kontekstno filtriranje je tip koji uključuje kontekstualne informacije korisnika u procesu preporuke. Ovdje se koristi niz kontekstualnih radnji korisnika, zajedno s trenutnim kontekstom, kako bi se predvidjela vjerovatnoća sljedeće akcije. Naprimjer, na osnovu nekih parametara kao što su zemlja, uređaj, vrijeme, datum prethodnih radnji (akcija), model može predvidjeti korisnikove naredne radnje (akcije). Isto tako jedan od tipova je i sistem zasnovan na znanju, kod kojeg se preporuke zasnivaju na uticaju na potrebe korisnika. To mogu biti neki kriteriji koji definiraju, kada bi neki proizvod (usluga) mogla biti od koristi za korisnika. Tehnike dubokog učenja su napredne tehnike koje koriste umjetne neuronske mreže za proučavanje svih vrsta podataka povezanih s određenim domenima vezanim za tematiku samog sistema. Postoje različite vrste neuronskih mreža koje se koriste za ovu namjenu (RNN, CNN, DNN).

# PRAKTIČNI DIO

Praktični dio se odnosi na implementaciji sistema preporuke za filmove. Pošto su domen ovog sistema filmovi, koristit će se već gotovi (prikupljeni) skup podataka, koji će u nastavku biti detaljnije objašnjen. Također, iskoristit će se prethodno navedeni tipovi, zajedno s pratećim algoritmima za preporuku. Iako se radi o sistemu preporuke za filmove, na sličan način bi se mogao implementirati i neki drugi sistem preporuke, naravno s adekvatnim podacima, kao što su muzika, knjige, online trgovina i drugi.

## Alati, biblioteke i radno okruženje

Implementacija je urađena koristeći Python programski jezik (verzija 3.10.0). Python je najbolji izbor za ovakve projekte, jer ima jako velik ekosistem biblioteka, mogućnosti vizualizacije, fleksibilnost, podršku zajednice, te nezavisnost platforme. Iz tog razloga je idealan kandidat za potrebe mašinskog učenja i analize podataka. Korištene su napredne Python-ove biblioteke koje pružaju efikasnost i jednostavnost u svojim primjenama, što karakteriše čitljivost i lahka sinatksu. U nastavku je lista korištenih biblioteka:

* numpy
* pandas
* mathplotlib
* seaborn
* scikit-learn
* tensorflow/keras
* surprise

Za pisanje samog koda korišten je Visual Studio Code, korištenjem posebnog formata (ekstenzije) prilagođenog Python-u koji se naziva Jupyter Notebook (.ipynb). Jupyter Notebook pruža podršku za kreiranje i dijeljenje dokumenta koji sadrži „živi“ (izvršeni) kod, koji je popraćen vizualizacijom i opisnim dijelom teksta. Kada je u pitanju sama implementacija prvo je odrađen kratak osvrt na skup podataka, te njihova kratka analiza i obrada za daljnju upotrebu. Prva metoda (algoritam) jeste kreiranje i testiranje modela neuronske mreže, a zatim primjena filtriranja baziranog na sadržaju korištenjem TF-IDF algoritma. Na kraju je prikazano poređenje dva algoritma KNN i SVD na primjeru korištenja kolaborativnog filtriranja. Svi podaci, modeli i metode su organizovani u zasebnim folderima i fajlovima, te sadrže kratka uputstva i objašnjenja samog koda.

## Skup podataka (dataset)

Skup podataka koji je korišten za implementaciju praktičnog dijela je MovieLens Dataset. Iza ovog skupa podataka stoji organizacija GroupLens Research koji su prikupili i stavili na raspolaganje podatke o recenzijama (ocjenama, rejtingu) s web stranice MovieLens (<https://movielens.org>). Podaci su prikupljeni u različitim vremenskim periodima, te postoje skupovi različitih veličina (broja podataka). Podaci su preuzeti s web stranice Kaggle (<https://www.kaggle.com>). Kaggle je stranica koja omogućava korisnicima da pronađu skupove podataka za korištenje u izgradnji modela vještačke inteligencije. Također, moguće je i objaviti svoje skupove podataka, raditi s drugim naučnicima i inženjerima, kao i učestvovati u raznim takmičenjima i izazovima koji se odnose na rješavanje problema iz sfere „nauke o podacima“ (data science). Sam MoviLens skup podataka sadržava ocjene korisnika na određene filmove (još dodatno i dodjeljene oznake korisnika na filmove, ali oni neće biti razmatrani u ovom radu). Period u kojem su kreirani podaci je između 1995-2015. godine, a skup podataka je generisan 2016. godine. Korisnici su odabrani nasumično, te njihovi podaci ne uključuju nikakve demografske podatke, već samo jedinstveni identifiaktor (ID). U nastavku će detaljnije biti proučen i prikazan skup podataka.

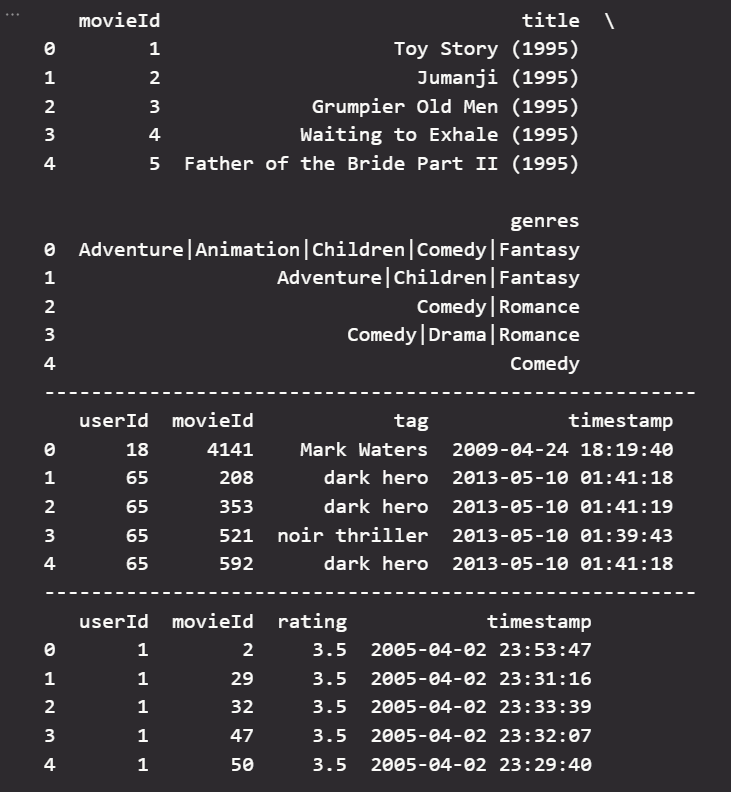
## Priprema i analiza podataka

Preuzeti podaci su raspoređeni unutar zasebnih .csv datoteka, a glavni skup podataka se sastoji od 6 datoteka, ali će se u trenutnom primjeru koristiti samo podaci iz dvije datoteke i to: movie.csv i rating.csv. Podaci o filmovima unutar movie.csv sastoje se od: movieId (jedinstvenog identifikatora filma), title (puni naziv filma), genres (lista pripadajućih žanrova filma). Podaci o ocjenama (rejtingu) unutar rating.csv sadrže četiri kolone: userId (jedinstveni identifikator korisnika), movieId (jedinstveni identifikator filma), rating (ocjenu korisnika za dati film u rasponu od 0.5 – 5), timestamp (datum i vrijeme kreiranja/davanja ocjene). Za generisanje sistema preporuke prvi korak jeste istraživanje podataka, kako bi se utvrdi odnos između podataka, ali kako bi se i razumjela svrha i vrijednost svakog pojedinačnog atributa. U ovom primjeru radi se o vrlo malom broju atributa, tako da je vrlo intuitivno zaključiti povezanost između ovih podataka. Radi se o tome da su ove dvije datoteke jednostavno povezane putem jedinstvenog identifikatora filma movieId (koji se pojavljuje u obje datoteke). Kada se radi o samoj količini (broju) podataka unutar datoteka radi se o:

* 27278 različitih filmova
* 20000264 različitih recenzija

U nastavku analize može se utvrditi da je ukupno 138494 različitih korisnika ostavilo ocjenu (minimalno jednu). Prije generisanja sistema preporuke važno je pripremiti same podatke, kako bi rezulati preporuke bili što tačniji. Ovo se odnosi na uklanjanje ili čišćenje suvišnih ili nepotpunih podataka. Ove operacije se mogu odnositi na:

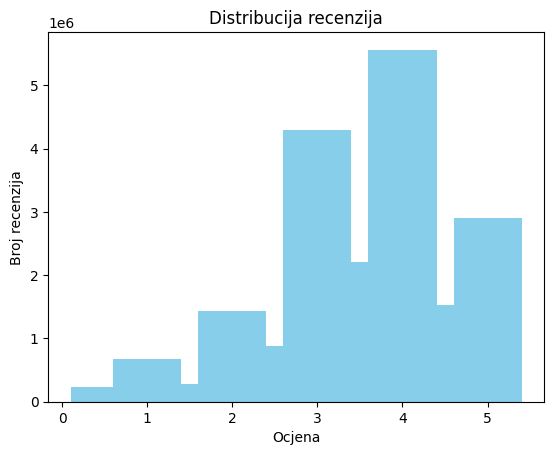
* uklanjane duplikata (ako postoje)
* uklanjanje N/A vrijednosti (ako postoje)
* izostavljanje filmova ili korisnika koji nisu „aktivni“ (korisnici koji su ostavili mali broj recenzija ili filmovi koji su dobili mali broj recenzija)



Slika 3

Naredni korak jeste vizualizacija dostupnih podataka, kako bi se lakše mogli uvidjeti određeni trendovi i poveznice. Naprimjer, za potpuno razumijevanje je korisno saznati neke grupisane podatke koji se ne mogu vrlo lahko saznati samim pregledom „sirovih“ podataka. Neki od zanimljivih grafičkih prikaza mogu uključivati:

* distribuciju recenzija (pregled broja recenzija po kategoriji)
* računanje srednje i medialne vrijednosti recenzija
* prosjek recenzija po korisniku i filmu
* prosjek količine recenzija po korisniku i filmu
* distribucija žanrova filmova
* stanje recenzija po godinama



Slika 4

# IMPLEMENTACIJA

Nakon pripreme i analize podataka može se započeti s kreiranjem sistema preporuke za filmove. Pošto je skup podataka istražen, filtriran i spreman za primjenu, može se krenuti s implementacijom različitih algoritama. Kroz ove algoritme bit će prikazana njihova tačnost i preciznnost i performanse, koje se odnose na predviđanje ocjena koje bi korisnik ostavio na filmove koje nije gledao (ocijenio), te generisanje preporuka na osnovu filmova koje je korisnik gledao (ocijenio), odnosno na osnovu sličnosti između korisnika s drugim korisnicima i filmova s drugim filmovima.

## Model neuronske mreže

Za kreiranje modela neuronske mreže bilo je neophodno još dodatno pripremiti podatke. Zbog velikog broja podataka (preko 20 miliona recenzija), proces treniranja neuronske mreže je spor, odnosno performanse znatno zavise od samih specifikacija hardvera. Iz tog razloga ocjene su dodatno filtrirane na način da se „odbace“ manje ocjene, tako da ostaju samo ocjene 4 i 5. Ovim pristupom ostalo je 12 miliona recenzija. Na ovaj način se može simulirati binarna klasifiakcija, tako da se ocjenom 5 smatra film koji bi se korisniku svidio (odnosno kojeg bi on kao takvog ocijenio), dok ocjena 4 predstavlja film koji mu se manje „sviđa“. U ovom kontekstu radi se o filmu koji mu se „ne sviđa“, jer su ostale ocjene samo 4 i 5, mogu se mapirati u vrijednosti 0 i 1. Dodatni korak jeste filtriranje korisnika, kako bi se broj recenzija još smanjio. Za ovo se koristi filtriranje kojim se izuzima 30% slučajno odabranih korisnika, a tako i njihovih pratećih recenzija. Na samom kraju konačno se došlo do „zadovoljavajućeg“ broja podataka za treniranje modela (oko 3.6 miliona recenzija). Prije kreiranja modela potrebno je još transformisati podatke koristeći LabelEncoder. Ovaj postupak je jako bitan, jer se podaci transformišu u normalizirane vrijednnosti, kako bi ih model lakše i bolje procesirao. Zatim se podaci dijele u dvije grupe: trening i test podaci (80% trening i 20% test podataka). Model neuronske mreže se kreira uz pomoć tensorflow i keras biblioteka, a sastoji se od sljedećih dijelova:

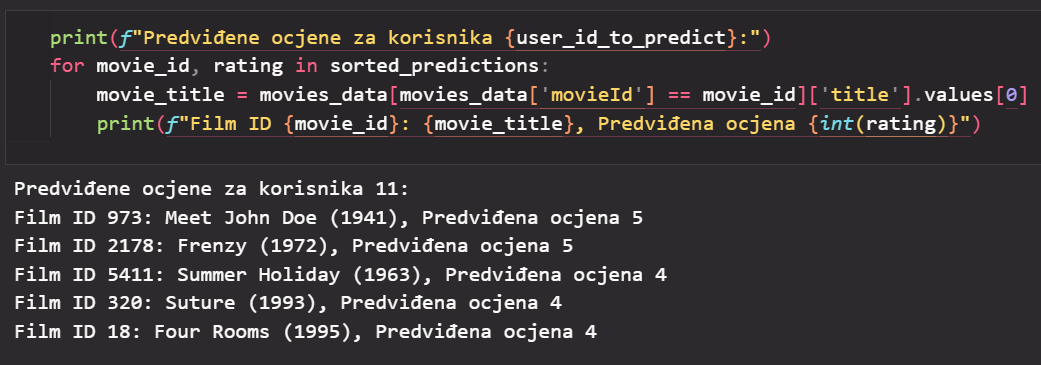
* ulazni slojevi - user\_input i movie\_input, koji predstavljaju jedinstvene identifikatore korisnika i filmova
* embedding slojevi – user\_embedding i movie\_embedding, koji se koriste za učenje reprezentacija korisnika i filmova, odnosno transformaciju identifikatora u gusto raspoređene vektore
* flatten slojevi – user\_flatten i movie\_flatten, koji služe kako bi se „izravnali“ rezultati iz embedding slojeva, što se postiže pretvaranjem matrica u vektore
* sloj konkatenacije – služi za spajanje rezultata iz flatten slojeva u jedan vektor
* dense slojevi – potpuno povezeni slojevi (128 neurona i „relu“ aktivacijska funkcija)
* izlazni sloj – sadrži 1 neuron (što predstavlja jedan izlaz) i ima linearnu aktivacijsku funkciju

Na kraju model je definisan prosljeđivanjem prethodno kreiranih ulaznih i izlaznih slojeve.



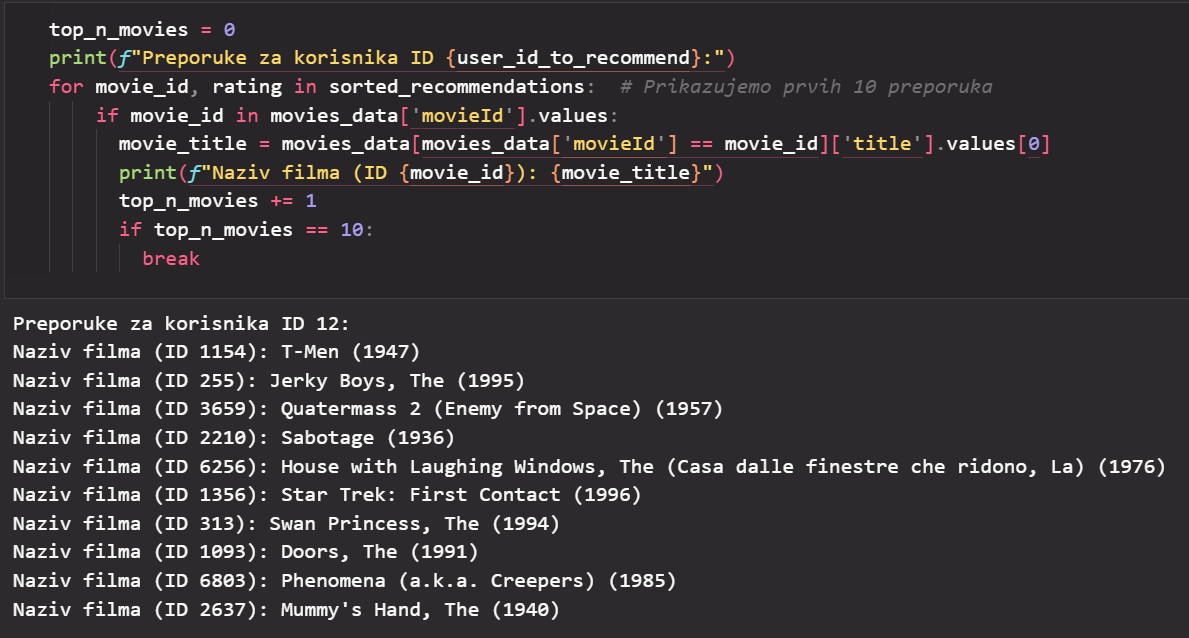
Slika 5

Sljedeći korak je „kompajliranje“ modela koji koristi „adam“ optimizator, za gubitak koristi „mean squared error“, te kao metriku ima „mean average error“. Model se zatim trenira korištenjem prethodno podijeljenih trening podataka na ulazne (korisnik i film) i izlazni (rejting). Proces treniranja je podijeljen u 10 epoha, koje su otprilike trajale 3 i pol sata. Naravno, uz različite specifikacije hardvera moguće su varijacije u odnosu na vrijeme izvođenja procesa treniranja. Nakon treniranja, model kao takav se može spremiti (odnosno istrenirane težine) na sam disk. Kako bi se u budućnosti ove istrenirane težine mogle ponovo koristiti, te kako bi se model mogao unaprijeđivati. Nakon učitavanja spremljenog modela (ili nastavka rada na trenutnom modelu) vrše se predikcije na osnovu testnih podataka. Na osnovu ovih predikcija mogu se izračunati „srednja kvadrata greška“ i „preciznost“ modela, koji u ovom slučaju iznose 0.33 (zaokruženo na dvije decimale), odnosno 0.48 (zaokruženo na dvije decimale). Kako bi se model testirao na nekim „stvarnim“ podacima, odnosno koristeći podatke iz rating.csv datoteke koji nisu uključeni u proces treniranja modela, mogu se odabrati neki nasumični identifikatori za korisnike i filmove. Cilj ovoga je osnovu slučajno odabranog identifikatora korisnika, prvo odrediti filmove koje on nije ocijenio, a zatim na osnovu identifikatora tih filmova predvidjeti koje bi ocjene korisnik dao tim filmovima. U konkretnom primjeru odabran je ID 11 za korisnika, te su uzeti slučajni ID-ovi njegovih neocjenjenih filmova (2178, 18, 973, 320, 5411 – njih 5 u ovom slučaju). Kada se modelu proslijede ovi podaci rezultat su predviđene ocjene za te filmove od strane odabranog korisnika. Pošto su na početku ostavljene samo ocjene 4 i 5, tako i rezultati variraju između 4 i 5 (u vidu decimalnih brojeva zaokruženih na nekoliko decimala), a radi jednostavnosti prikaza zaokružene su na približne vrijednosti.



Slika 6

Na osnovu ovih podataka sada je potrebno predvidjeti ocjene od svih neocjenjenih filmova za nekog korisnika, te onda jednostavno sortirati te rezultate od najvećeg do najmanjeg. Preporuka filmova korisniku se u ovom slučaju bazira na prikaz prvih N rezultata (u većini primjera koristi se prvih 10 rezultata). Na identičan način kao i u prethodnom slučaju odabire se ID korisnika, određuju se ID-ovi njegovih neocjenjenih filmova, te se nakon predikcije vrši sortiranje na osnovu ocjene, te se ispisuje prvih 10 rezultata.



Slika 7

## TF-IDF algoritam

## KNN i SVD algoritmi

## Poređenje rezultata i pregled performansi

# ZAKLJUČAK

# LITERATURA