



PREPORUKA FILMOVA BAZIRANA NA PREFERENCAMA KORISNIKA

UMJETNA INTELIGENCIJA

Martina Gaćina
Fran Vojković
Alen Živković

Uvod

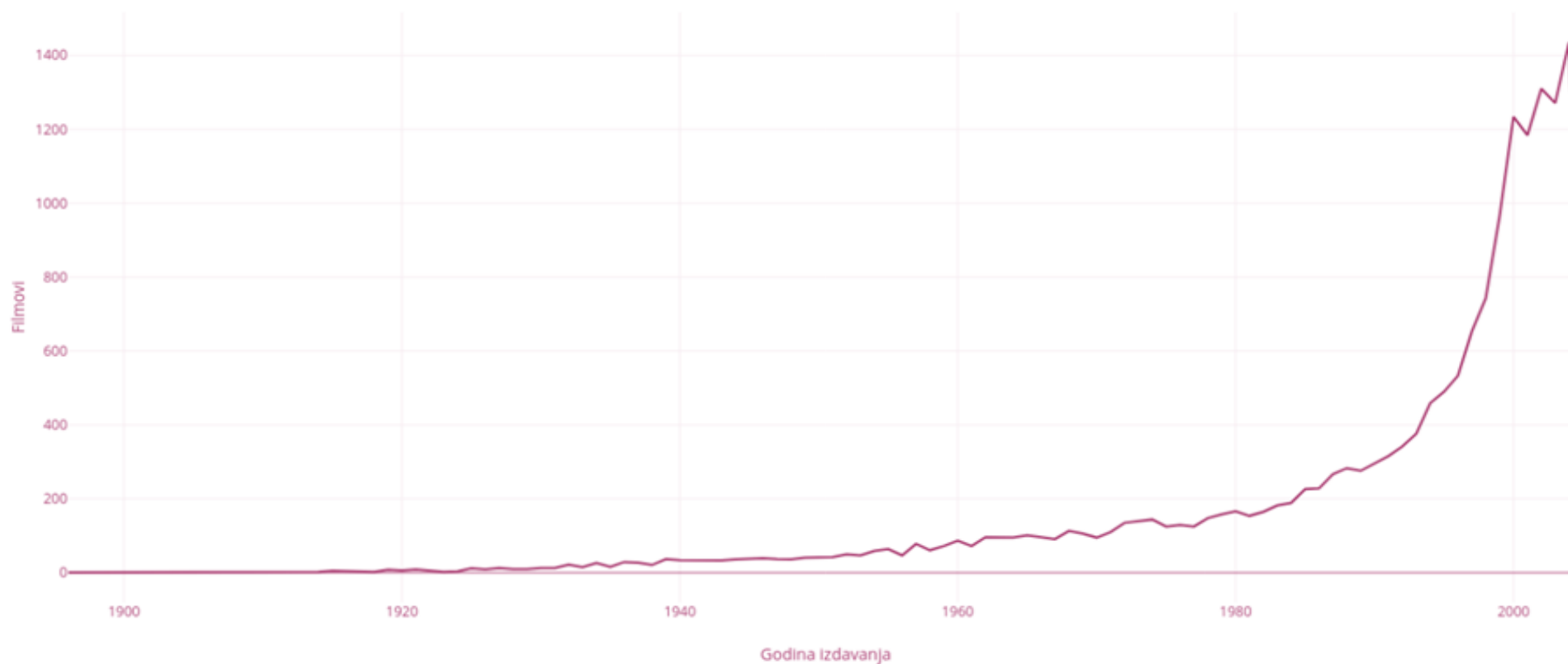
- ▶ Modeli za preporuku filmova korisniku na temelju vlastitih preferencija
- ▶ Temelje se na strojnom učenju, iz podataka o ocjenama filmova
- ▶ Data set- naslov filma, godina izdavanja, ID filma, ocjene korisnika za pojedini film (od 1 do 5)

Analiza i obrada podataka

- ▶ 17 770 filmova
- ▶ 24 053 764 ocjena korisnika
- ▶ Zbog jednostavnosti i brzine izvođenja koristimo samo jedan dana set

Analiza i obrada podataka – filmovi po godini izdavanja

17770 filmova sortirani po godini izdavanja

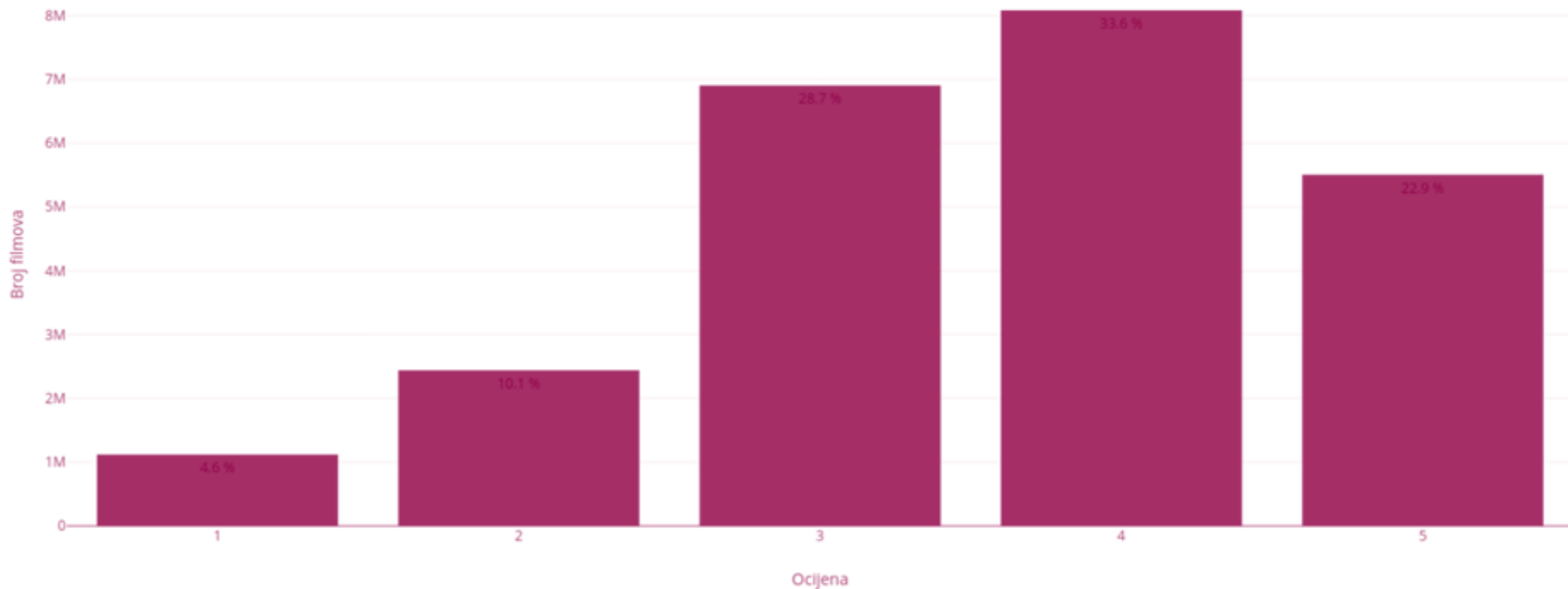


Analiza i obrada podataka – filmovi po godini izdavanja

- ▶ Rast približno eksponencijalan do 2004. godine
- ▶ Nakon 2004. godine slijedi pad
- ▶ Pad zanemariv- koristili samo jedan dana set, podaci nepotpuni za godine nakon 2004.

Analiza i obrada podataka – distribucija ocjena po filmovima

Distribucija 24053764 ocjena za filmove

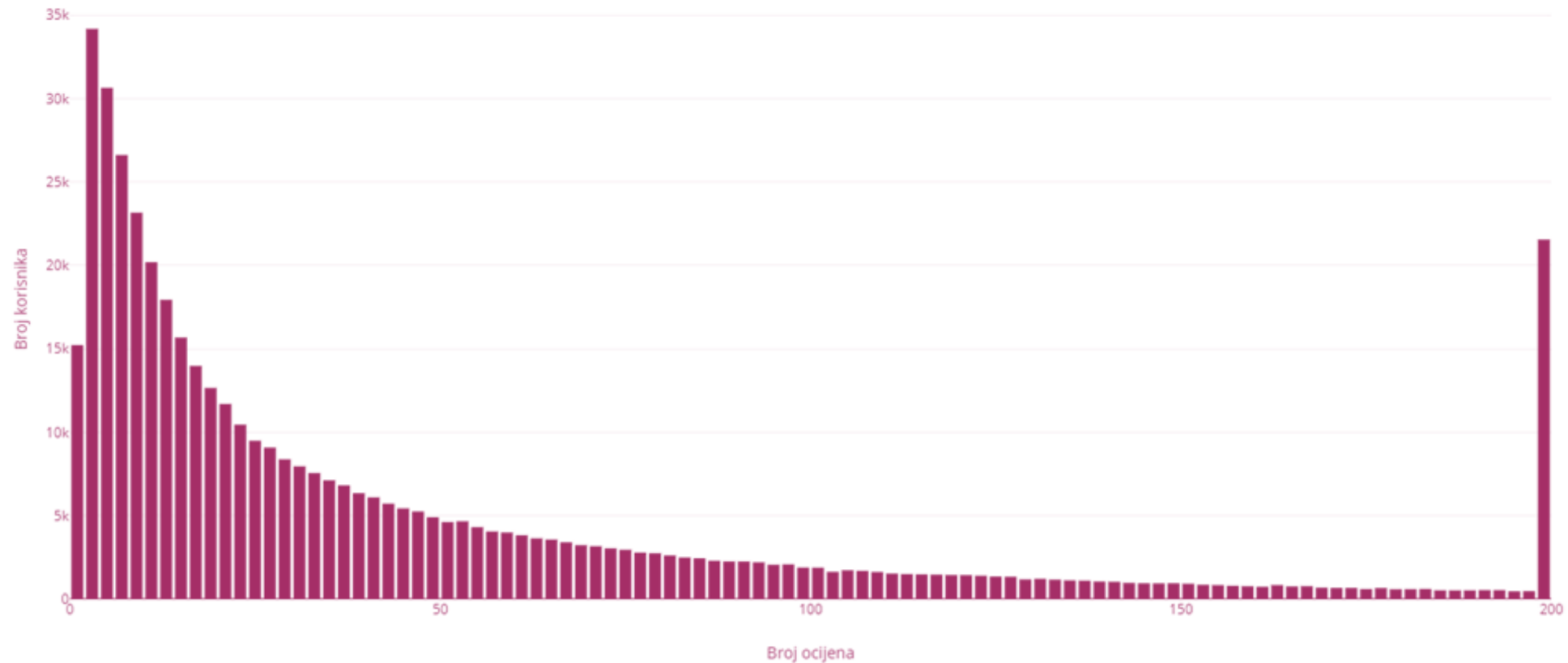


Analiza i obrada podataka – distribucija ocjena po filmovima

- ▶ 85,2% ocjena veće ili jednake 3
- ▶ Nezadovoljni korisnici napuštaju streaming servis
- ▶ 14,7% ocjena manjih od 3- zanemarivo

Analiza i obrada podataka – distribucija ocjena po korisniku

Distribucija ocjena po korisniku



Analiza i obrada podataka – distribucija ocjena po korisniku

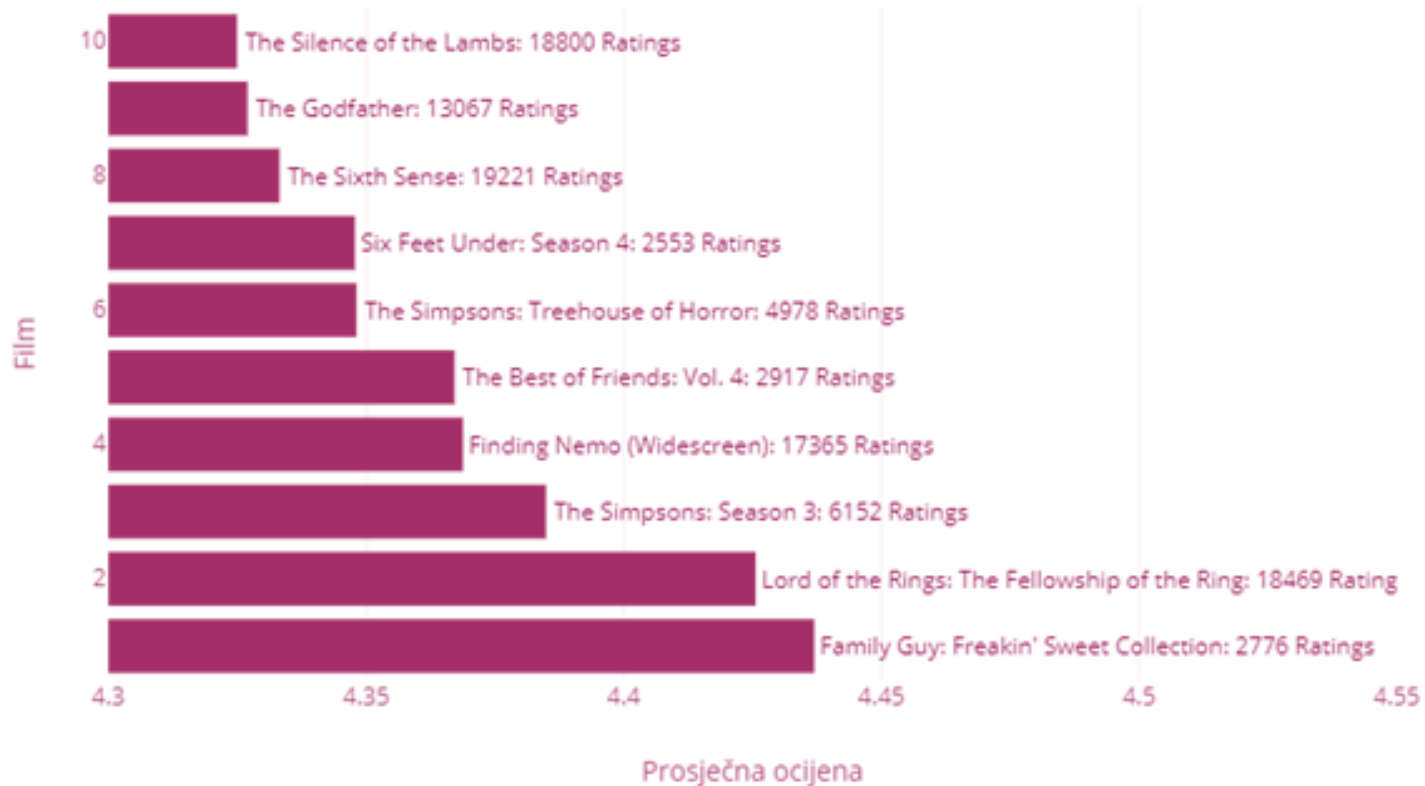
- ▶ Pad ocjena po korisniku približno eksponencijalan
- ▶ Većina korisnika ocijeni do 50 filmova
- ▶ Zbog toga ćemo „očistiti” podatke filmova koji imaju malo ocjena te podatke korisnika koji su ocijenili manji broj filmova
- ▶ To model čini pouzdanijim

Analiza i obrada podataka – matrični prikaz

- ▶ Rec i - filmovi, stupci - korisnici
- ▶ (i, j) -to mjesto – ocjena korisnika i za film j
- ▶ Ako korisnik i nije ocijenio film j , mjesto označimo kao prazno
- ▶ Matrica u početku rijetka
- ▶ Želimo popuniti prazna mjesta u matrici predviđanjem potencijalne ocjene korisnika
- ▶ Korisnicima koji imaju mali broj ocijenjenih filmova, filmove preporučujemo na osnovu generalnog „top score-a” filmova dobivenog iz ocjena svih korisnika

Analiza i obrada podataka - filmovi po prosječnoj ocjeni

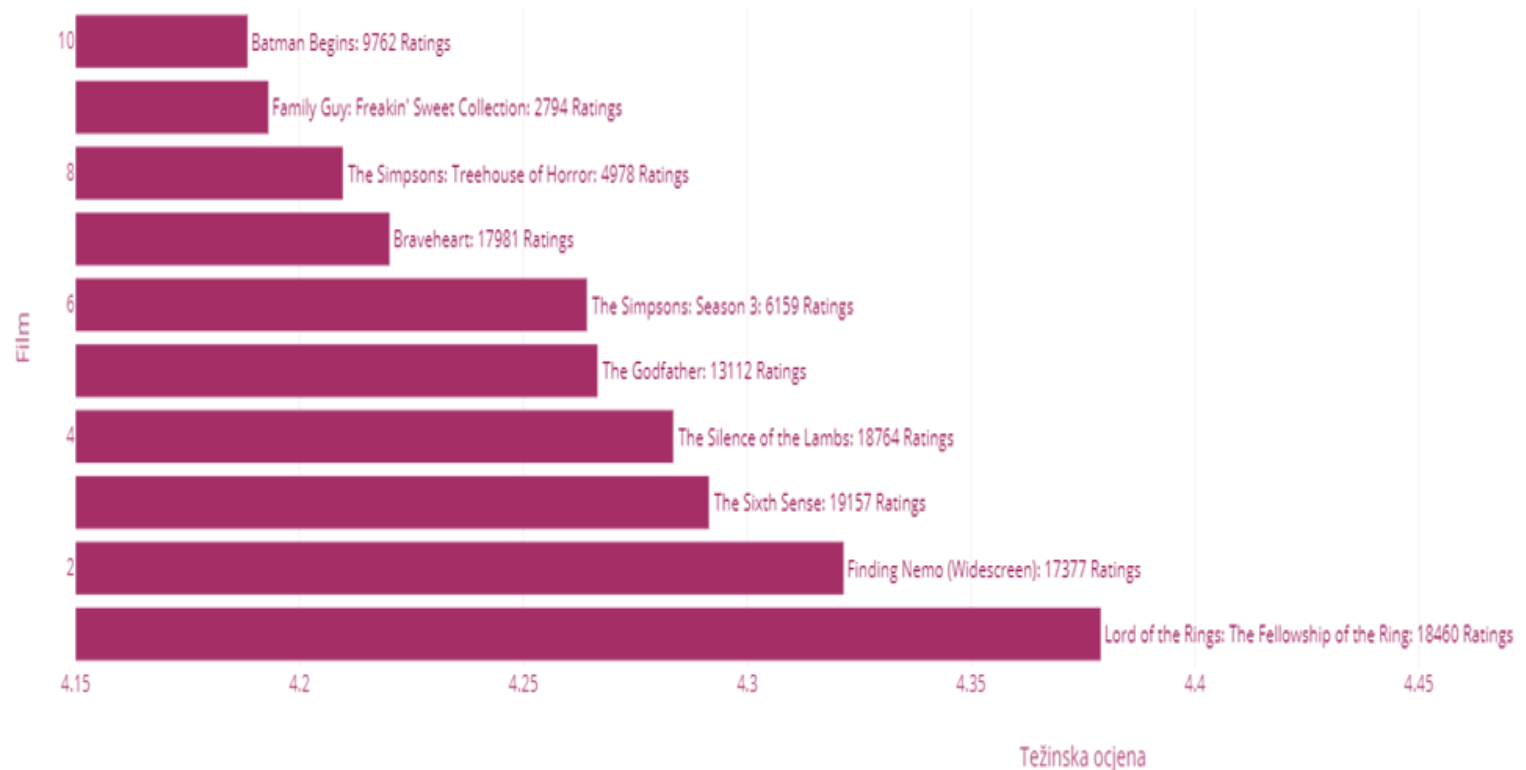
Rang lista top 10 filmova po prosječnoj ocjeni



Analiza i obrada podataka - „krive” ocjene

- Problem „krivih” ocjena
- Uvodimo težinske ocjene

Rang lista top 10 težinski ocijenjenih filmova

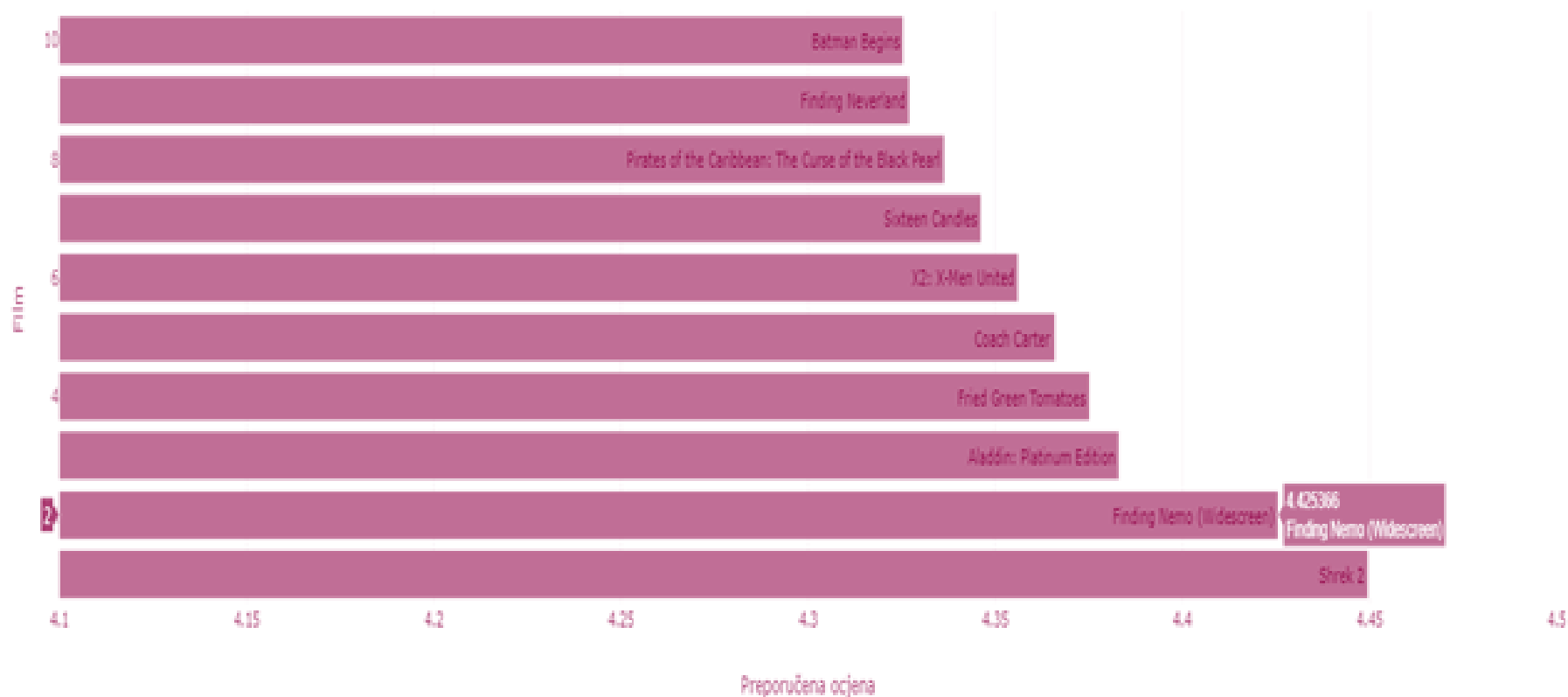


Analiza i obrada podataka – sličnost među korisnicima

- ▶ Svaki redak matrice interpretiramo kao vektor
- ▶ Računamo sličnost među svim korisnicima- vektorima
- ▶ Korisnicima preporučujemo visoko ocijenjene filmove sličnih korisnika
- ▶ Prazna mjesta u matrici popunjavamo prosječnom vrijednošću ocjena filmova svakog korisnika
- ▶ Ocjene sličnih korisnika skaliramo faktorom sličnosti i računamo njihovu srednju vrijednost
- ▶ Filtriranje neocijenjenih filmova nekog korisnika pokazuje najbolje rezultate
- ▶ Analogno možemo raditi za sličnost između filmova

Analiza i obrada podataka – sličnost među korisnicima

Lista top 10 preporučenih filmova za korisnika na osnovu sličnosti



Pearsovova R-korelacija

- ▶ Računamo linearnu korelaciju između ocjena svih filmova
- ▶ Dobivamo listu od 10 filmova s najvećom korelacijom
- ▶ Na osnovi korelacije između filma koji se korisniku svidio i dobivenih filmova, korisniku možemo predložiti filmove koji bi ga mogli zanimati
- ▶ Pomoću funkcije preporuke (recommend) dobivamo listu „sličnih” filmova
- ▶ Za upit: `recommend(„Batman Beyond: The Movie”)` dobivamo sljedeću listu preporuka

Pearsonova R-korelacija

Za film (Batman Beyond: The Movie)

- Top 10 preporuka temeljam Pearsonove R-korelacije -

PearsonR	Name
1.000000	Batman Beyond: The Movie
0.756045	Regular Guys
0.752263	Batman Beyond: Return of the Joker
0.693876	11:14
0.685877	Morvern Callar
0.680686	Look at Me
0.652110	Sordid Lives
0.651295	Fast
0.651104	Batman: The Animated Series: Tales of the Dark...
0.649219	Batman: Mask of the Phantasm

Pearsonova R-korelacija

- ▶ Prvi stupac – Pearsonov koeficijent korelacije
- ▶ Drugi stupac – nazivi filmova
- ▶ Vidimo da je za film Batmand Beyond: The Movie Pearsonov koeficijent korelacije 1.000, kao što bi i očekivali
- ▶ Za upit: recommend(„Justice League”) dobivamo sljedeću listu preporuka

Pearsonova R-korelacija

Za film (Justice League)

- Top 10 preporuka temeljam Pearsonove R-korelacije -

PearsonR	Name
1.000000	Justice League
0.775745	Justice League: Justice on Trial
0.658475	Regular Guys
0.654393	Testosterone
0.643469	Batman: The Animated Series: Tales of the Dark...
0.631749	Batman Beyond: The Movie
0.614198	Batman Beyond: Return of the Joker
0.577765	Batman: Mask of the Phantasm
0.567673	Nine Dead Gay Guys
0.567410	Yi Yi

Random Forest i XGBoost modeli

- ▶ Imamo listu pogledanih filmova nekog korisnika i njegove ocjene
- ▶ Predviđamo koje bi od nepogledanih filmova korisnik ocijenio najvišom ocjenom kako bi mu ih preporučili
- ▶ Modele baziramo na ID-u i godini izlaska filma
- ▶ Treniramo model na trening setu (podacima o korisnikovim ocjenama filmova)

Random Forest i XGBoost modeli

- ▶ dataframe `df_title` kopiramo i očistimo od filmova koji nisu u našem `df-u` i onih koje smo odlučili izbaciti
- ▶ Izbacimo još i filmove koje je korisnik već pogledao te dobijemo popis svih filmova koje možemo preporučiti
- ▶ Provodimo predviđanja
- ▶ Filmove sortiramo silazno po predviđenim ocjenama
- ▶ Za korisnika 785314 dobivamo preporuke:

Random Forest i XGBoost modeli

Year		Name	Estimate_Score
392	1998.0	The Replacement Killers	5.0
329	1998.0	Wild Things	5.0
359	1999.0	Lies	5.0
344	1998.0	Star Trek: Voyager: Season 5	5.0
44	1999.0	The Love Letter	4.9
288	1998.0	The Avengers	4.9
303	1997.0	King of the Hill: Season 1	4.9
284	1997.0	The Devil's Own	4.9
361	2000.0	The Flintstones in Viva Rock Vegas	4.8
315	1999.0	Futurama: Monster Robot Maniac Fun Collection	4.8

Random Forest i XGBoost modeli

	XGB_Estimate
392	3.794509
329	3.794509
359	3.794509
344	3.794509
44	3.794509
288	3.794509
303	3.794509
284	3.794509
361	3.572643
315	3.794509



Hvala na
pažnji !