

PRIMENA MAPPER ALGORITMA NA NETFLIX PODACIMA

NIKOLA VESELINOVIC

KOSTA GRUJCIĆ

OPIS SKUPA PODATAKA

- KOMPANIJA NETFLIX JE 2009. GODINE ZAPOČELA TAKMIČENJE, ČIJI JE CILJ BIO DIZAJNIRANJE MODELA ZA PREPORUČIVANJE FILMOVA KORISNICIMA NETFLIX SERVISA
- JAVNO DOSTUPNI SKUP PODATAKA SE SASTOJI OD BAZE PODATAKA FILMOVA I KORISNIKA
- KONAČNA REŠENJA SU PROVERENA NA PRIVATNOM SKUPU PODATAKA NA OSNOVU KOJIH JE VRŠENO OCENJIVANJE TIMOVA

OPIS SKUPA PODATAKA

- BAZA PODATAKA FILMOVA SE SASTOJI OD IDENTIFIKATORA FILMA, NAZIVA FILMA I DATUMA NJEGOVOG OBJAVLJIVANJA
- BAZA PODATAKA KORISNIKA SE SASTOJI OD IDENTIFIKATORA KORISNIKA, IDENTIFIKATORA FILMA, OCENE FILMA I DATUMA KADA SE OCENA DESILA
- POSTOJI UKUPNO 17770 FILMOVA I 480189 KORISNIKA
- GLAVNA METRIKA JE KOREN SREDNJEKVADRATNOG ODSUPANJA (ENG. RMSE)

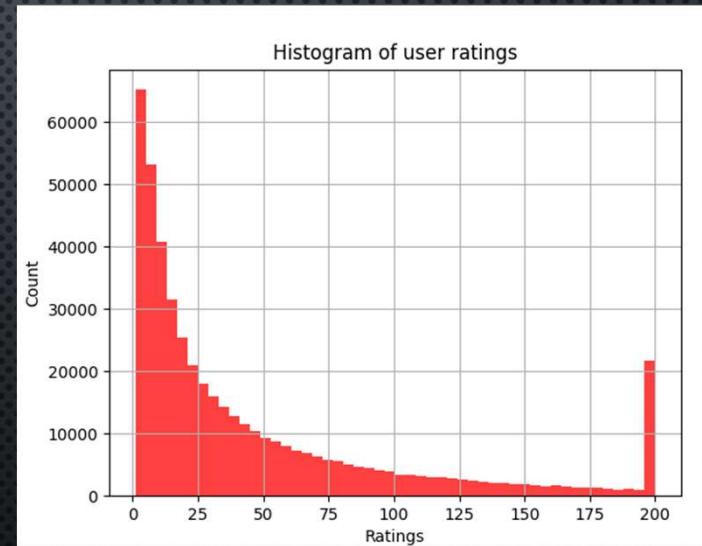
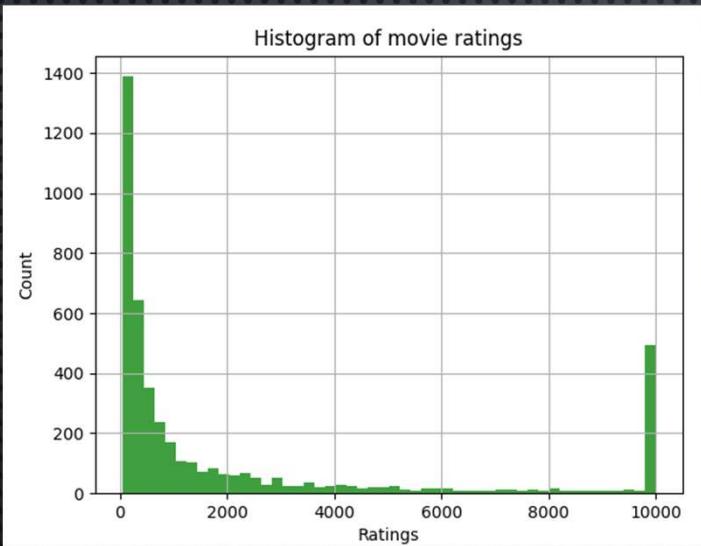
PRIступ

- OSNOVNI MODEL JE PLITKA NEURONSKA MREŽA
- NAKON OBУČAVANJA NEURONSKE MREŽE VRŠI SE TOPOLOŠKA ANALIZA PODATAKA

RAD SA PODACIMA

- PODACI SE PREČIŠĆAVAJU U CILJU ELIMINISANJA FILMOVA SA MALO OCENA I KORISNIKA KOJI SU MALO OCENJIVALI
- PODACI O FILMOVIMA I KORISNICIMA SE POTOM UTAPAJU U NISKODIMENZIONE PROSTORE JER PRETPOSTAVLJAMO DA SU KORISNICI KAO I FILMOVI MEĐUSOBNO ZAVISNI
- ZA TESTIRANJE JE KORIŠĆENO 100000 INSTANCI, DOK SU OSTALE KORIŠĆENE ZA OBUČAVANJE

RASPODELA OCENA KORISNIKA I FILMOVA



UTAPANJE

- ZA UTAPANJE PODATAKA O FILMOVIMA I KORISNICIMA KORIŠĆEN JE POSEBAN SLOJ U KERAS BIBLIOTECI – *EMBEDDING*
- ZA DIMENZIJE PROSTORA JE UZET ČETVRTI KOREN IZVORNOG BROJA ATRIBUTA

REZULTATI OBUČAVANJA

- ZA KONSTRUKCIJU NEURONSKE MREŽE JE KORIŠĆENA BIBLIOTEKA KERAS
- PORED ULAZNOG I IZLAZNOG SLOJA, POSTOJI SAMO JEDAN SKRIVENI SLOJ OD 256 NEURONA
- MODEL JE TRENIRAN JEDNU EPOHU UPOTREBOM ADAM OPTIMIZATORA MINIMIZUJUĆI SREDNJEKVADRATNO ODSTUPANJE

REZULTATI OBUČAVANJA

Training set	Validation set	Test set
0.8962 (MSE)	0.8193 (MSE)	0.8221 (MSE)
0.9467 (RMSE)	0.9052 (RMSE)	0.9067 (RMSE)

TOPOLOŠKA ANALIZA PODATAKA

- NEFORMALNO: TOPOLOGIJA NAD PODATCIMA PREDSTAVLJA ODNOS ELEMENATA BAZE PODATAKA U PROSTORU (MOŽE BITI I FORMALIZACIJA KLASTERA)
- ČESTO, DISKRETNNA REPREZENTACIJA TOPOLOŠKIH PODATAKA JE SVEDENA NA SIMPLICIJALNE KOMPLEKSE

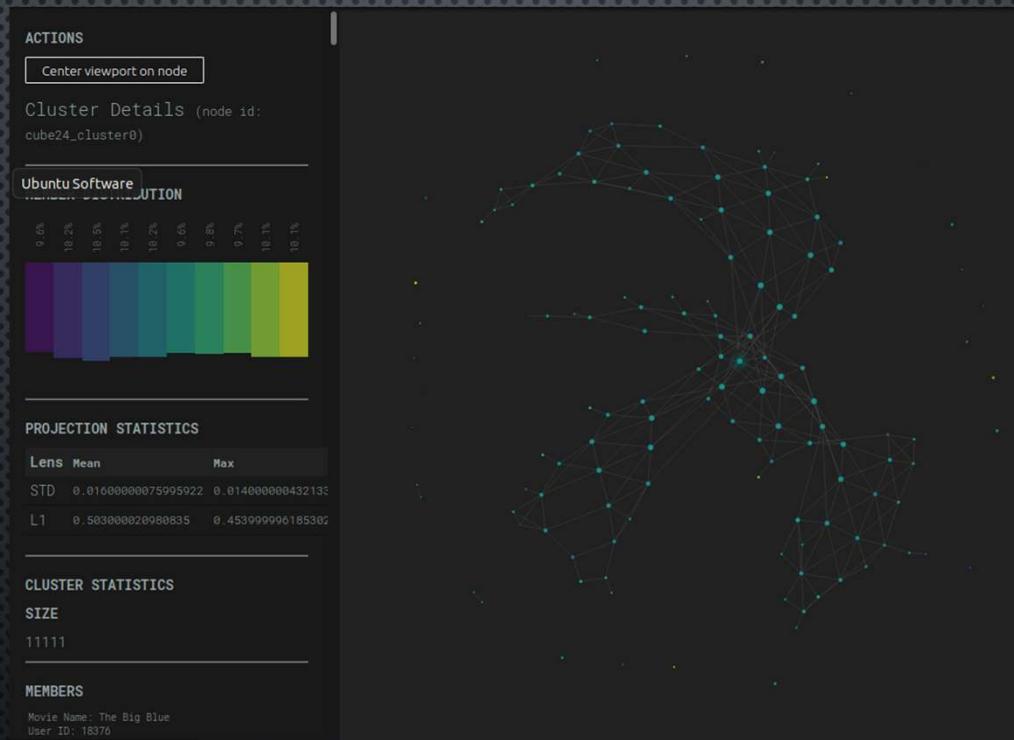
MAPPER

- MAPPER JE JEDNA OD, AKO NE I, NAJČEŠĆE KORIŠĆENA TOPOLOŠKA REPREZENTACIJA PODATAKA
- ON INTERPRETIROBILO KAKVU LISTU PODATAKA U \mathbb{R}^N KAO „OBLAK TACAKA“, I PRETVARA GA U UPROŠĆEN TOPOLOŠKI GRAF

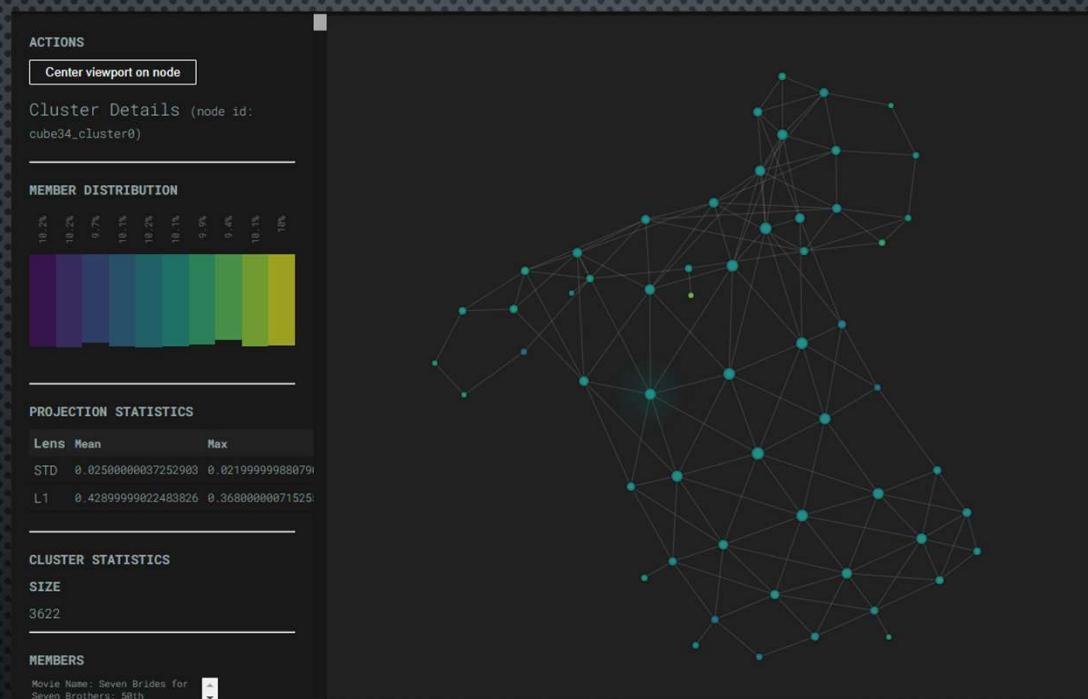
MAPPER

- ZA MAPIRANJE PODATAKA KORISTIMO KEPLERMAPPER
- KONSTRUIŠEMO I PROSLEDJUJEMO POKRIVANJEM I PODSKUP INVERYA GLAVNOG SKUPA
- KAO POSLEDNJI PARAMETAR PROSLEDJUJEMO ALGORITAM ZA KLASTEROVANJE U NASEM SLUCAJU KORISTIMO 2 KAKO BI MOGLI DA IH UPREDIMO

ALGORITAM ZA KLASTEROVANJE: AGGLOMERATIVECLUSTERING



ALGORITAM ZA KLASTEROVANJE: DBSCAN





HVALA NA PAŽNJI

memegenerator.net