# Sistem za preporuku filmova

Članovi tima: Luka Nikolić SW64/2017

## Motivacija

Pronaći film koji nam se sviđa u današnje vreme sa izobiljem sadržaja je monumentalan zahvat, pogotovo za filmofile koji su već pogledali sve poznatije naslove. Potrebno je razviti aplikaciju koja će korisniku na osnovu njegovih preferencija i istorije pregleda predložiti najprikladnije filmove.

## Opis problema

Problem predlaganja sadržaja su probale da reše mnoge kompanije i platforme. Netflix je razvio kompleksan machine learning algoritam sa velikim brojem parametara i podataka[[1](#_Literatura)]. Koristeći Poasonovu korelaciju između filmova i regresione modele određuju jedinstvene predloge za svakog korisnika. Sa druge strane, rad [[2](#_Literatura)] uvodi rule based model koji predviđa preferencije grupe korisnika na osnovu njihovih istorija i ocena.

Predlažem moj sistem za preporuku filmova koji će, koristeći Drools rule engine pored korisničkih omiljenih žanrova i ocena, uzeti u obzir ključne reči koje dublje opisuju filmove. Ključne reči su generalan pojam koji slobodnije i preciznije mogu da opišu atmosferu, iskustvo i osećanja dok gledamo film od drugih metrika. Iz tog razloga smatram da mogu da se koriste za kvalitetnije predviđanje filmova koji bi se korisniku svideli.

## Metodologija rada

Sistemu je potrebna kolekcija filmova koja se može prikupiti sa online baza podataka filmova. Takođe svaki film sadrži žanrove i ključne reči. Rule engine će učitati sve činjenice prilikom pokretanja aplikacije i formiranja baze.

Za rešavanje ovog problema koristiće se sledeći entiteti:

* User: Predstavlja korisnika, sadrži lične podatke , sviđanja, nesviđanja, pogledani filmovi, watchlist, omiljene ključne reči, omiljeni žanrovi
* Movie: Predstavlja film iz baze koji se učitava na startupu. Sadrži informacije o režiseru, glumcima, trajanju, prosečnoj oceni, žanr, itd.
* Genre: Predstavlja žanr.
* Review: Ostavljene od strane korisnika. Ocena utiče na weight odgovarajućeg filma
* Recommendation: Izlaz iz sistema, konkretni predlozi filmova sa bodovima

Ulaz u sistem su istorija pogledanih filmova, sviđanja, nesviđanja, recenzije, *watchlist*. Izlaz iz sistema su konkretni predlozi filmova sa bodovima (Recommendation(movie, score))

### Baza znanja:

Pravila su grupisana po nivoima rezonovanja. Kada korisnik označi da mu se nešto sviđa, ubacuje se činjenica u obliku UserLikes(Movie m), isto tako za nesviđanja i recenzije(UserDislikes(Movie m) ili UserReviews(Movie m,value)).

A1 – pravila atomskih akcija

Grupa pravila koja će od postojećih činjenica nad filmovima da naprave činjenice nad ključnim rečima koje sadrže taj film, kao na primer:

* Like(User, Movie) ->LikeKeyword(User, Movie, keyword)
* Dislike(User, Movie) ->DislikeKeyword(User, movie, keyword)
* Review(User, Movie, value) -> ReviewsKeyword(User, Movie,keyword, value)

A2 – pravila evaluacije ključnih reči

Ovo pravilo će koristiti accumulate nad prethodnim nivoom i formirati KeywordStats(keyword, likes, dislikes,reviewAvg, reviewCount)

A3 – Pravilo evaluacije filmova na osnovu ocena ključnih reči. Filtrira pogledane filmove.

Ovo pravilo će koristiti accumulate nad KeywordStats() i za svaki film kreirati MovieStats(score)

A4 – Pravila generisanja predloga

Ova pravila generišu konkretne Recommendation(Movie, score) predloge na osnovu MovieStats() i preferenci korisnika(Ako je u watchlist-i, povećaj *score*).

## Forward Chaining:

Koristi se za preporuku filmova na osnovu aktivnosti kao što su ostavljanje ocene, sviđanje, nesviđanje ili dodavanje filma u *watchlist*. Pravila se ulančavaju i izvršavaju po grupama gde svaki nivo koristi činjenice kreirane u prethodnom nivou.

LikeMovie -> likeMovieKeyword -> KeywordStats -> Movie stats -> Recommendation

## Konkretan primer

Ulazi korisnika

* Omiljeni žanrovi: Horror, Mystery, Drama
* Omiljene ključne reči: scary, slasher, serial killer
* Korisniku se sviđa Silence of the lambs i ostavio je recenziju sa ocenom 8

Baza znanja sadrzi sledeće filmove: Silence of the Lambs (7.5; scary, horror, mystery, killer; genre: Horror), Inception(8.2, mystery, dream,suspenseful ;genre: Mystery), Scary Movie(7.9; scary, comedy, killer; genre: Comedy)

### Izvršavanje pravila:

A1 – Za svaku korisnikovu akciju se kreira cinjenica za svaki keyword iz tog filma.

* LikesKeyword(User, Silence of the lambs, scary)
* LikesKeyword(User, Silence of the lambs, horror)
* LikesKeyword(User, Silence of the lambs,mystery)
* LikesKeyword(User, Silence of the lambs,killer)
* ReviewKeyword(User, Silence of the lambs, scary, 8)
* ReviewKeyword(User, Silence of the lambs, horror, 8)
* ReviewKeyword(User, Silence of the lambs, mystery, 8)
* ReviewKeyword(User, Silence of the lambs, killer, 8)

A2 – Akumuliraju se akcije i računaju procene ključnih reči

* KeywordStats(User, scary, likes-dislikes=1 , reviewAvg = 8, reviewCount=1)
* KeywordStats(User, horror, likes-dislikes=1 , reviewAvg = 8, reviewCount=1)
* KeywordStats(User, mystery, likes-dislikes=1 , reviewAvg = 8, reviewCount=1)
* KeywordStats(User, killer, likes-dislikes=1 , reviewAvg = 8, reviewCount=1)

A3 – Kreiraju se procene filmova na osnovu procena ključnih reči koje dati film sadrži. Filtriraju se već pogledani filmovi.

* MovieStats(User, Inception, score=8) // 1\*8
* MovieStats(User, Scary Movie, score=16) // 1\*8 + 1\*8

A4 – Kreiraju se preporuke filmova . Nagrađuju se filmovi koji sadrže omiljene ključne reči ili žanr. Inception dobija jos 5 poena jer sadrži omiljeni žanr mystery. Scary movie dobija 5 jer sadrži omiljenu ključnu reč scary

* Recommendation(User, Inception, score=13) // 8 + 5 jer se poklapa žanr mystery
* Recommendation(User, Scary Movie, score=21) // 16+5 jer se poklapa keyword scary

## Literatura

1. Bennett, James, and Stan Lanning. "The netflix prize." Proceedings of KDD cup and workshop. Vol. 2007. 2007. Retrieved from: <https://www.cs.uic.edu/~liub/KDD-cup-2007/NetflixPrize-description.pdf>
2. Sharma, Mugdha, Laxmi Ahuja, and Vinay Kumar. "A novel rule based data mining approach towards movie recommender system." Journal of Information and Organizational Sciences 44.1 (2020): 157-170. Retrieved from: <https://hrcak.srce.hr/file/348076>