

**UNIVERZITET U BEOGRADU
FAKULTET ORGANIZACIONIH NAUKA**

SEMINARSKI RAD

Predmet: Alati i metode veštačke inteligencije i softverskog inženjerstva

Tema: Analiza IMDB filmskog skupa podataka u programskom jeziku Clojure

Profesor:

Dr. Dragan Đurić

Student:

Božidar Mastilović 2025/3814

Beograd, 2026.

1. Ideja I Motivacija

Ideja za seminarski rad je proizašla iz želje da se praktično primene koncepti funkcionalnog programiranja i obrade podataka u programskom jeziku Clojure. Kao domen izabrana je analiza filmskih podataka jer filmski skupovi podataka sadrže raznovrsne attribute pogodne za filtriranje, grupisanje i statističku obradu.

Kao izvor podataka korišćen je javno dostupan IMDB skup podataka u CSV formatu, koji sadrži informacije o filmovima kao što su:

- Naziv
- Godina izlaska
- Žanrovi
- Ocena
- Režiser
- Broj glasova
- Prihod
- ...

Ovakav skup podataka omogućava jedinstvenu, ali ilustrativnu primenu funkcionalnih koncepta, kao i demonstraciju osnovnih metoda analize i obrade podataka.

Cilj projekta je demonstracija razumevanja funkcionalnog pristupa programiranju, organizacije koda u Clojure projektu i primene osnovnih metoda analize podataka nad realnim skupom podataka.

2. Korišćene tehnologije i alati

Projekat je realizovan u programskom jeziku Clojure, koji se izvršava na JVM platformi i zasnovan je na funkcionalnom programiranju. Za upravljanje projektom i zavisnostima korišćen je Leiningen, standardni alat za rad sa Clojure projektima.

Razvojno okruženje je IntelliJ IDEA sa Cursive pluginom, koji omogućava podršku za Clojure sintaksu, REPL i rad sa Leiningen projektima. Podaci su učitavani iz CSV fajla korišćenjem standardnih alata Clojure programskog jezika, bez oslanjanja na spoljne biblioteke za obradu CSV fajlova.

Za testiranje funkcionalne logike korišćen je ugrađeni Clojure framework clojure.test, čime je obezbeđena ispravnost i stabilnost ključnih funkcija.

3. Struktura projekta i organizacija koda

Projekat je organizovan u više namespace-ova, u skladu sa preporukama za struktuiranje Clojure projekta. Organizacijom koda po namespace-ovima postignuto je jasno razdvajanje odgovornosti i bolja čitljivost koda.

- **Parser** - sadrži logiku za učitavanje i obradu CSV podataka, uključujući parsiranje redova, čišćenje vrednosti i mapiranje kolona u odgovarajuće strukture podataka.
- **Analysis** – obuhvata funkcionalne operacije nad filmovima, kao što su filtriranje, grupisanje i agregacija podataka.
- **Core** – predstavlja ulaznu tačku aplikacije i služi za demonstraciju rada implementiranih funkcionalnosti.
- **ML** – sadrži jednostavnu demonstraciju primene metode mašinskog učenja nad filmskim podacima.

U okviru projekta definisana su tri osnovna domenska entiteta: film, režiser i žanr. Entitet film predstavlja osnovnu jedinicu podataka, dok su entiteti režiser i žanr izvedeni iz osnovnog entiteta primenom funkcionalnih transformacija nad kolekcijama podataka.

4. Funkcionalna analiza podataka

Analiza filmskih podataka realizovana je korišćenjem osnovnih funkcionalnih koncepta kao što su funkcije višeg reda: map, filter, reduce. Implementirane su funkcije za filtriranje filmova po žanru, izdvajanje naziva filmova, izračunavanje prosečne ocene, kao i grupisanje filmova po režiseru i žanru.

Posebna pažnja posvećena je grupisanju filmova po žanru, pri čemu jedan film može pripadati više žanrova. Ova funkcionalnost je realizovana korišćenjem ugnježdenih funkcionalnih operacija, čime je demonstrirana fleksibilnost funkcionalnog pristupa u obradi složenijih struktura podataka.

Sve funkcije u ovom delu projekta su čiste funkcije, bez sporednih efekata, što omogućava jednostavno testiranje i ponovnu upotrebu koda.

5. Testiranje

Za proveru ispravnosti funkcionalne logike napisani su unit testovi korišćenjem biblioteke clojure.test. Testirane su funkcije koje sadrže netrivijalnu logiku, kao što su grupisanje filmova po žanru i režiseru, kao i izračunavanje prosečnih ocena.

Testovi su pisani nad malim, ručno definisanim skupom podataka, čime je obezbeđeno determinističko ponašanje i lakša provera očekivanih rezultata. Ulazna tačka aplikacije i I/O operacije nisu testirane, jer ne sadrže poslovnu logiku.

6. Demonstracija primene mašinskog učenja

U projektu je takođe implementirana jednostavna demonstracija primene metode mašinskog učenja u vidu linearne regresije. Cilj ovog dela projekta je upoznavanje sa osnovnim koracima treniranja i evaluacije modela.

Model predviđa IMDB ocenu filma na osnovu godine izlaska i broja glasova. Pre treniranja, podaci su filtrirani kako bi se zadržali samo filmovi sa potpunim numeričkim vrednostima, a zatim normalizovani radi obezbeđivanja numeričke stabilnosti. Skup podataka podeljen je na trening i test skup u odnosu 80:20.

Model je treniran iterativno, prilagođavanjem težina linearne funkcije u cilju smanjenja greške predikcije. Kvalitet modela ocenjen je korišćenjem srednje apsolutne greške (MAE). Dobijeni rezultati pokazuju da je greška relativno niska, što je posledica inicijalizacije modela približnom prosečnom ocenom i jednostavne strukture modela. Cilj ovog dela rada je demonstracija procesa treniranja i evaluacije, a ne postizanje optimalnih rezultata.

Rezultati:

```
**** ML demo ****
Train size: 799
Test size: 200
Mean Absolute Error: 0.20172424600526093
```

7. Zaključak

U okviru seminarskog rada prikazana je praktična primena funkcionalnog programiranja i osnovnih metoda analize podataka u programskom jeziku Clojure. Projekat obuhvata učitavanje i obradu realnog skupa podataka, funkcionalnu analizu i grupisanje podataka, testiranje funkcionalne logike, kao i jednostavnu demonstraciju primene metoda mašinskog učenja.

8. Literatura i resursi

- [1] [Clojure for the Brave and True](#)
- [2] [Zvanična clojure dokumentacija](#)
- [3] [IMDB top 1000 csv](#)
- [4] [Stackoverflow - diskusije na temu csv parsiranja u clojure](#)