Stavranje prediktivnog modela preporuke filma pomoću MovieLens dataseta

Projektna dokumentacija

Verzija 0.2

Sadržaj

1. Puni naziv projekta 4

2. Skraćeni naziv projekta 4

3. Opis problema/teme projekta 4

4. Cilj projekta 4

5. Voditelj studentskog tima 4

6. Rezultat(i) 4

7. Slični projekti 4

8. Resursi 4

9. Glavni rizici 5

10. Smanjivanje rizika 5

11. Glavne faze projekta 5

12. Struktura raspodijeljenog posla (engl. *Work Breakdown Structure* - WBS) 5

13. Kontrolne točke projekta 5

14. Gantogram 5

15. Zapisnici sastanaka 5

Prijedlog i plan projekta

Na koji način koristiti predložak?

Tekst pisan u italic formi opisuje koje informacije je potrebno uključiti u pojedino poglavlje Prijedloga.

Za upis vlastitog teksta, potrebno je pritisnuti <ENTER> nakon italic teksta.

Tekst upisan u <trokutastim zagradama> treba zamijeniti s onim što se navodi.

Svi članovi tima trebaju pročitati plan i suglasiti se s njime, a to potvrđuju svojim potpisom na kraju dokumenta.

# Puni naziv projekta

[Navesti puni naziv projekta iz kojeg će biti vidljiva problematika/tematika kojom se projekt bavi.]

Stavranje prediktivnog modela preporuke filma pomoću MovieLens dataseta.

# Skraćeni naziv projekta

[Navesti skraćeni naziv projekta, ako je definiran.]

Prediktivni model preporuke filma

# Opis problema/teme projekta

[Objasniti problem, odnosno temu projekta, objasniti ukratko tip, uvjete i kontekst projekta.]

Eksploratorna analiza te razvijanje prediktivnog modela preporuke filma algoritmima strojnog učenja pomoću MovieLens dataseta. Model bi preporučivao nove filmove korisniku s obzirom na podatke koje ima o korisniku kao što su koje žanrove preferira, koje filmove je dobro ocijenio te koje loše te s obzirom na podatke vrlo sličnih korisnika.

# Cilj projekta

[Navesti predviđeni cilj ili ciljeve projekta. Definiranje ciljeva omogućuje određivanje pravca u kojem će se kretati izvođenje projekta. Navesti predviđeno trajanje projekta.]

Izrada prediktivnog modele koji će preporučiti film korisniku na osnovi podataka koje ima o korisniku i korisnicima sličnog ukusa.

Predviđa se da projekt traje 3 mjeseca.

# Voditelj studentskog tima

<Ime i prezime studenta>

Matej Čubek

# Rezultat(i)

[Navesti što će se isporučiti na kraju projekta, voditi računa da osim rezultata u vidu nekog proizvoda ovdje treba navesti i svu dokumentaciju.]

Rezultat ovoga projekta je prediktivni model koji za bilo kojeg korisnika iz baze podataka može preporučiti novi film koji bi mu se s velikom vjerojatnošću svidio.

# Slični projekti

[Navesti projekte koji su povezani s dotičnim projektom.]

<https://github.com/AshwathSalimath/Movie-ratings-prediction-using-MovieLens-Dataset>

<https://github.com/AIGamer28100/Movie-Lens-Dataset-Visualisation-and-Prediction>

<https://datascience-enthusiast.com/Python/cs110_lab2_als_prediction.html>

# Resursi

[Navesti ljudske i ostale resurse potrebne za uspješno dovršenje projekta. Popuniti tablicu raspoloživih članova tima s podacima važnim za projekt. Mogu se navesti znanja i vještine člana koje mogu biti od koristi za projekt, na primjer znanja Java-e, XML-a, iskustvo u radu s MS Projectom, sudjelovanje u sličnim projektima ili bilo kakva korisna informacija. Ako projekt koristi i druge resurse napraviti posebnu tablicu za njih. U kolonu Napomene treba upisati sve termine kad dotični član tima neće biti raspoloživ za rad na projektu (putovanja, odmori, odsustva).]

**Tablica ljudskih resursa**

|  |  |  |  |
| --- | --- | --- | --- |
| **Ime i prezime** | **E-mail adresa** | **GSM broj** | **Napomene** |
| Iva Bokšić | Iva.boksic@fer.hr | +385976969143 | Poznavanje R |
| Tomislav Grbeša | Tomislav.grbesa@fer.hr | +385977259354 | Predznanje datascienca |
| Matej Čubek | Matej.cubek@fer.hr | +385992066001 | Znanje Jave i pythona |

# Glavni rizici

[Navesti glavne zapreke za ostvarenje uspjeha projekta, te posljedice ukoliko projekt ne uspije.]

* Nedostatak vremena za upoznavanje s znanosti o podacima i izradom prediktivnog modela.
  + Posljedica rizika je nedovoljno dorađen prediktivni model.
* Premalo prijašnjeg znanja za izradu prediktivnog modela
  + Posljedica rizika je nedovoljno dorađen prediktivni model.
* Nedostatak raspoloživih materijala za učenje
  + Nemoguće odraditi projektni zadatak

# Smanjivanje rizika

[Navesti korake koji će se poduzeti kako bi se što je moguće više umanjio svaki od prethodno navedenih rizika.]

* Bolja organizacija vremena za projekt.
* Učenje novih materijala kako bi dopunili znanje.
* Potraga za novim materijalima na internetu ili traženje pomoći od mentora
* Česti sastanci radi podjele posla i rasprave što je i kako do sada napravljeno

# Glavne faze projekta

[Navesti glavne faze projekta, te ukratko objašnjenje po kojem načelu je projekt podijeljen na te faze- vremenska organizacija, smanjenje rizika, raspoloživost resursa i/ili nešto drugo.]

Projekt je podijeljen na vremenske faze od kojih svaka traje dva do tri tjedna.

* 1. Faza: Stvaranje tima, organizacija i odabir tehnologija.
  2. Faza: Odabir dataseta i njegova eksploracija i analiza.
  3. Faza: Obrada podataka i priprema za algoritme strojnog učenja.
  4. Faza: Izrada i odabir modela.
  5. Faza: Objava rezultata.

# Struktura raspodijeljenog posla (engl. *Work Breakdown Structure* - WBS)

[Nacrtati WBS s navedenim aktivnostima projekta.]

# Kontrolne točke projekta (engl. *milestones*)

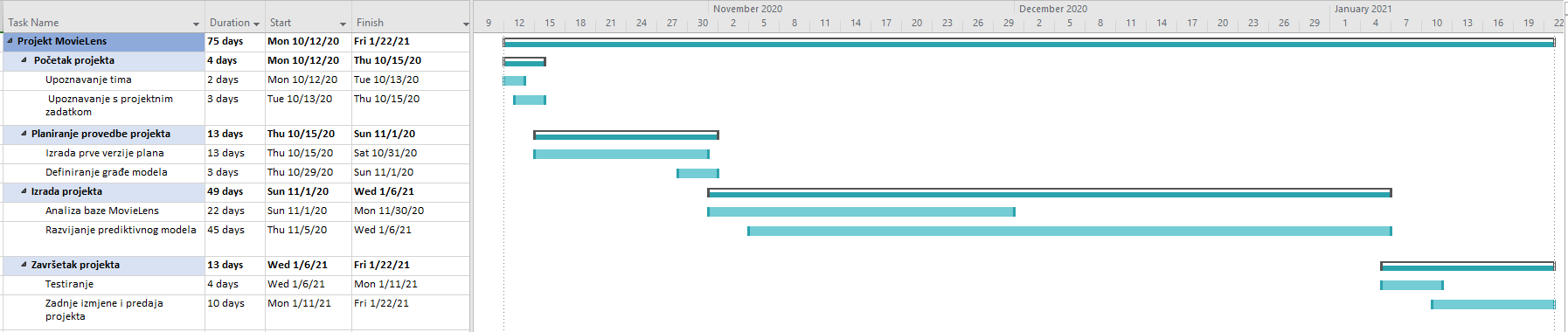
[Općenito, kontrolna točka projekta je događaj ili rezultat neke aktivnosti koji ukazuje na to je li projekt u skladu sa zadanim rokovima ili kasni. Ta informacija se upisuje u kolonu o statusu projekta. Ako projekt kasni moraju se poduzeti akcije da se rokovi dostignu. Za svaku kontrolnu točku treba odrediti točan datum. Po potrebi se mogu dodavati ili oduzimati redovi tablice.]

**Tablica kontrolnih točki projekta**

|  |  |  |  |
| --- | --- | --- | --- |
| **Kontrolne točke** | **Planirani datum** | **Realizirani datum** | **Status projekta** |
| Predaja prve verzije plana projekta | 6.11.2020. |  |  |
| Gotova analiza dataseta | 1.12.2020. |  |  |
| Razvijen prediktivni model | 6.01.2021 |  |  |
| Objava rezultata i predaja | 17.01.2021 |  |  |

# Gantogram

[Izraditi Gantogram pomoću programa MS Project, Open Workbench, Microsoft Excel - http://office.microsoft.com/hr-hr/excel/HA010346051050.aspx, i sl. Pohraniti prikaz Gantograma (screenshot) i postaviti ga unutar ovog poglavlja kao ubačenu sliku.]



# Zapisnici sastanaka

[Ovdje za svaki održani sastanak navesti: datum, vrijeme i mjesto održavanja sastanaka, popis nazočnih, glavne zaključke sastanka.]

* 1. Sastanak : 22.10.2020. 19h, Slack
     + Sudjelovali: Iva Bokšić, Tomislav Grbeša, Matej Čubek
     + Upoznavanje tima i rasprava o trenutnim znanjima pojedinog člana
     + Razmjena korisnih materijala za upoznavanje s projektom
  2. Sastanak : 29.10.2020. 18h, Slack,
     + Sudjelovali: Iva Bokšić, Tomislav Grbeša, Matej Čubek
     + Podjela rada pri izradi dokumentacije

|  |  |
| --- | --- |
| **Suglasan s dokumentom (potpisuju članovi tima):** |  |
| Iva Bokšić | Datum: \_\_\_\_\_\_\_\_\_\_\_\_ Potpis: \_\_\_\_\_\_\_\_\_\_\_\_ |
| Tomislav Grbeša | Datum: \_\_\_\_\_\_\_\_\_\_\_\_ Potpis: \_\_\_\_\_\_\_\_\_\_\_\_ |
| Matej Čubek | Datum: \_\_\_\_\_\_\_\_\_\_\_\_ Potpis: \_\_\_\_\_\_\_\_\_\_\_\_ |

|  |  |
| --- | --- |
| **Odobrio(potpisuje nastavnik):**  izv. prof. dr. sc. Damir Pintar |  |

|  |  |
| --- | --- |
| Datum: \_\_\_\_\_\_\_\_\_\_\_\_ | Potpis: \_\_\_\_\_\_\_\_\_\_\_\_ |