

MLOps Junior Project

Opis zadatka

Istrenirati modele koji će da rade predikciju ocene filma. Potrebno je istrenirati 2 modela po želji. Ono što je bitno je da jedan model bude algoritam iz familije tradicionalnih ML algoritama (logistic regression, logistic SVM, Decision Tree, XGBoost...), a drugi model mora biti neka neuronska mreža. Nakon što su modeli istrenirani potrebno je uraditi deploy modela na način po želji.

Podaci

Unutar **data.csv** nalaze se podaci o svakom filmu. Potrebno je uraditi predikciju **vote_average** kolone. Potrebno je uraditi preprocessing datog fajla i učitati train/val/test dataset u SQL bazu. Potrebno je napraviti docker-compose u kome će biti pgadmin i postgresSQL koji će biti korišćeni za skladištenje train/test/val podataka.

Treniranje

Potrebno je istrenirati jedan ML model i jedan Deep Learning model koji radi predikciju ocene filma. **Potrebno je dokerizovati trening.** Napisati odvojene docker fajlove i instrukcije kako pokrenuti trening. Nakon treninga treba ispisati metrike i sačuvati model.

Takođe, potrebno je dodati dokerizovano treniranje kao poseban servis unutar docker-compose.yaml fajla iz sekcije **Podaci**.

Nije potrebno ulagati poseban trud da metrike budu iznad 80-90%, fokus je na clean code-u i OOP principima.

Deployment

Potrebno je, na način po zelji, uraditi deployment oba istrenirana modela. Ovo je takođe potrebno dokerizovati i dodati u docker-compose.yaml fajla iz sekcije **Podaci**.

Test i CICD

Potrebno je ispisati par unit i integracionih testova i korišćenjem GitHub Actions-a potrebno je na svaki commit izvršiti sve testove.

Fokus

Prilikom izrade potrebno je fokusirati se na:

- Postovanje CleanCode i OOP principa u Python-u
- Docker
- Dokumentacija
 - **Za svaku funkciju i klasu koristiti tipove i docstr**