

Duboko Učenje - Drugi domaći

Datum: 04.01.2021

Rok za odbranu: Do ispita

Domaći zadatak vredi 25 poena i radi se individualno. Odbrana će biti organizovana po dogovoru. Postoje dve opcije domaćeg zadatka, odlučite se za jednu.

Opcija 1

Prva opcija je zadatak iz oblasti NLP-a. Na sledećem [linku](#) se nalazi Kaggle takmičenje sa skupom podataka koji treba da iskoristite. Podaci su tvitovi vezani za Covid-19 virus i labelirani su u 5 klasa, *Positive*, *Extremely positive*, *Neutral*, *Negative*, *Extremely negative*. Vaš zadatak je da kreirate modele koji klasifikuju tvitove u ove kategorije. Podaci nisu dovoljno sređeni tako da se od vas očekuje da uradite svu potrebnu pripremu podataka.

Rezultat ovog zadatka treba da budu dva modela kao i prpratni izveštaj, isto kao u prvom domaćem, to podrazumeva tabele sa rezultatima i parametrima modela, confusion matrice i tensorboard grafike. Zadatak možete poslati u obliku Jupyter/Colab Notebook-a radi lakše reprodukcije ili kao python fajlove.

Prvi model je base model od koga ne očekujemo nikakve posebne rezultate već služi samo kao polazna osnova. Ovde možete koristiti veoma jednostavne tehnike kao što su Naive Bayes, XGBoost Classifier sa tekstualnim featurim-a, Glove embedding-e i jednostavnu neuralnu mrežu, itd...Rezultati za ovaj model nisu bitni, ne očekujemo ništa posebno. Drugi model treba da ima potencijal da ostvari dosta bolje rezultate, tu se misli na pretrained modele kao što su BERT, DistilBERT, GPT...koje možete fine-tunovati ili možete od nule trenirati vaš model (Embedding layer iz TF-a + LSTM slojevi na primer). Rezultati ovog modela već moraju biti dosta bolji, 70-80% ili čak više.

Opcija 2

Pronaći rad iz 2019./2020. Ili 2021. godine iz oblasti dubokog učenja, konkretno iz oblasti obrade prirodnog jezika (NLP) ili iz generativnih modela (konkretno GAN-ovi neki). Rad mora biti značajan po inovativnosti ili po rezultatima i objavljen u nekom cenjenom naučnom časopisu (ne možete obrađivati random radove koji nemaju poentu). Rad je potrebno detaljno pročitati i razumeti. Potrebno je pripremiti prezentaciju u kojoj ćete objasniti o čemu se radi u radu, šta je njegov doprinos i kako se to odnosi na trenutno stanje u oblasti. Pored ovoga poželjno je da imate demo (kod autora, ne morate kucati vi) koji demonstrira rad, ukoliko je ovo tehnički moguće.

Rad zapakovati u **ime_prezime_index.zip** i poslati na mmilunovic@raf.rs.
Subject maila mora biti **[DL DZ2] Ime Prezime**.