

ML Industrija 4.0 2021/22 - završni projekti

Završni projekat iz Mašinskog učenja se vodi kao predispitna obaveza i osmišljen je sa idejom da se priđe realnim primenama usvojenih tema, uz proširivanje kako teorijskih, tako i tehničkih znanja. Tema za završni projekat se može predložiti samoinicijativno ili odabrati sa spisaka tema koji je u nastavku. Samoinicijativno predložene teme nose najviše **30** poena, a teme odabrane sa spiska u nastavku najviše **25** poena. Projekti se mogu prijavljivati samostalno ili timski, a pored svake teme piše za koliko je osoba namenjena.

Teme koje se nalaze u nastavku birane su sa liste završnih projekata kursa Mašinskog učenja Stanford univerziteta. Te teme su pokrivene izveštajima koji opisuju poželjne korake, a sadrže i referencu na korišćen skup podataka ili opisan postupak za dobijanje skupa podataka. Pored svake teme navedeni su modeli koje je potrebno implementirati. Svakako ohrabrujemo dodatna istraživanja - isprobavanje drugačijeg (većeg) skupa podataka, nadgradnje modela i dopunske eksperimente.

Prijava projekta

Prilikom prijave teme, neophodno je na adresu nevena_ciric@matf.bg.ac.rs poslati:

- naslov teme
- imena članova koji će raditi na projektu i njihove GitHub naloge
- adresu [GitHub repozitorijuma](#) na kojem će se naći kod.

Prilikom prijave možete postaviti pitanja ili predloge za eventualne izmene teme za koju se prijavljujete. Projekti se mogu prijaviti do **31. decembra**.

Informacije o svim prijavljenim projektima u školskoj 2021/22. godini biće upisani [ovde](#).

Teme

1. **Classification of Cardiac Arrhythmias Patients (2 studenata)**

<https://cs229.stanford.edu/proj2014/AlGharbi%20Fatema.%20Fazel%20Azar.%20Haider%20Batool.%20Cardiac%20Arrhythmias%20Patients.pdf?fbclid=IwAR1pAjq98pMxsgH WJjAHvKJJsxSco8OrDbKI7YYPCP-8F9R49tvKNycCZE>

Potrebno je isprobati sledeće modele:

- logističku regresiju, SVM, slučajne šume, KNN, AdaBoost i XGBoost ansambl stabala odlučivanja i potpuno povezanu neuronsku mrežu

2. **Restaurant Recommendation System (2 studenta)**

https://cs229.stanford.edu/proj2014/Ashish%20Gandhe.Restaurant%20Recommendation%20System.pdf?fbclid=IwAR18J5pgPIra1_toUdD3yNoX2Psztl0Ur6W4MMNkwUyG1IrH1bNrCIGvkmY

Skup podataka: <https://www.yelp.com/dataset>

Potrebno je isprobati sledeće modele:

- logističku regresiju, linearni SVM, kernelizovani SVM, KNN, slučajne šume i potpuno povezanu neuronsku mrežu

3. **Recognition and Classification of Fast Food Images (2 studenta)**

https://cs229.stanford.edu/proj2014/ShaoYu%20Lu,%20Sina%20Lin,%20Beibei%20Wang.%20Recognition%20and%20Classification%20of%20Fast%20Food%20Images.pdf?fbclid=IwAR1y0wXuryhKUnAmvGXXzpVRBg90QEw8flid2qdg0udR0AIMLe8F8_utoNs

Skup podataka: <https://github.com/ivanDonadello/Food-Categories-Classification>

Potrebno je isprobati sledeće modele:

- kernelizovani KNN, kernelizovani SVM i konvolutivnu neuronsku mrežu

4. **Predicting quality of wine based on chemical attributes (2 studenta)**

https://cs229.stanford.edu/proj2015/245_report.pdf

Skup podataka: <https://archive.ics.uci.edu/ml/datasets/wine+quality>

Potrebno je isprobati sledeće modele:

- KNN, linearna regresija sa težinama, aditivna logistička regresija, multinomijalna logistička regresija i potpuno povezana neuronska mreža

5. **Predicting flight delays and cancellations using weather as a feature (2 studenta)**
https://cs229.stanford.edu/proj2016/report/MenonMovva-PredictingFlightDelays-report.pdf?fbclid=IwAR1PSEtgQ-fMB6_6GVi0R7fQodTilzZVsasfdNyPKKYg1OeE7gSWth9bRI8
Skup podataka: <https://data.world/data-society/airlines-delay>

Potrebno je isprobati sledeće modele:

- linearna regresija, SVM, slučajne šume, XGBoost ansambl stabala odlučivanja i potpuno povezanu neuronsku mrežu

6. **Paragraph Topic Classification (2 studenta)**
<https://cs229.stanford.edu/proj2016/report/NhoNg-ParagraphTopicClassification-report.pdf>
Uputstvo za dobijanje skupa podataka:
<https://levelup.gitconnected.com/two-simple-ways-to-scrape-text-from-wikipedia-in-python-9ce07426579b>
<https://pdf.co/blog/python-wikipedia-module-for-scraping-wikipedia-articles>

Potrebno je isprobati sledeće modele:

- One-vs-Rest SVM i konvolutivna neuronska mreža
- obe vrste modela isprobati nad TF-IDF i GloVe reprezentacijama teksta