Poštovane koleginice i kolege,

Za potrebe projektnog zadatka potrebno je primeniti metode i tehnike mašinskog učenja koje su rađene tokom kursa i sumirati urađeno u vidu izveštaja koji će imati **najviše pet strana**. U izveštaju ne treba da se prikazuju kodovi iz *Jypiter Notebook*-a ili matematičke formule, već samo tekst, slike i tabele sa rezultatima.

Poglavlja mogu biti nazivi zahteva. Primeri izveštaja:

* <http://cs229.stanford.edu/proj2012/HoRomanoWu-KickedCarPrediction.pdf>
* <http://cs229.stanford.edu/proj2013/LeeHamJiang-EcommerceTransactionAnomalyClassification.pdf>
* <http://cs229.stanford.edu/proj2014/Aditya%20Sarkar,%20Quentin%20Perrot,%20Julien%20Kawawa,%20Stay%20Alert>
* <http://cs229.stanford.edu/proj2016/report/Smolka-BeatingTheBookies-report.pdf>  
  <http://cs229.stanford.edu/proj2015/270_report.pdf>

Možete raditi na jednom izabranom problemu od ponuđenih:

* [Carvana - Predict Car Prices | Kaggle](https://fonbgacrs.sharepoint.com/:u:/s/online_nastava_oas_202223_i14132/ETExAZqqn_5GkuwRZPCA4WMB3D1vky-DXjzaIH_sXNsEgA?e=tNzu7F)
* <https://www.kaggle.com/mamtadhaker/lt-vehicle-loan-default-prediction>
* <https://www.kaggle.com/c/avito-prohibited-content/overview>
* <https://www.kaggle.com/competitions/store-sales-time-series-forecasting/data>

Formalno, zahtevi koje treba da ispunite su:

**1. Opis i razumevanje problema**

- Uvod,

- Opis problema,

- Koji su izazovi u rešavanju problema

- Zašto je bitan

- Kome je bitan

**2. Opis i razumevanje podataka**

- Opis atributa

- Disbalans klasa i opis cene različitih grešaka

- Vizualizacije podataka (nije cilj da pokažete da znate da crtate, već da se iz slika vidi neka zavisnost)

- Nedostajuće vrednosti

- Ostalo...

**3. Priprema podataka**

- Transformacija tipova podataka

- Nedostajuće vrednosti

- Izvođenje novih atributa, kao i testiranje koliko utiču na kvalitet rešenja

**4. Treniranje više algoritama i interpretacija dobijenih rezultata**

- Deljenje skupa podataka na trening i test (validacija rešenja) npr. train-test split, cross-validation ili slično - odlučiti se za način validacije rešenje i objasniti zašto ste ga izabrali

- Izbor odgovarajućih algoritama i njihovo treniranje (izabrati najmanje tri algoritma - uključujući i ansambl algoritme)

- Izbor odgovarajućih mera evaluacije (dakle, može i više od jedne) npr. tačnost, precision, recall, AUC, total\_cost i objasniti zašto ste ih izabrali

- Interpretacija i poređenje dobijenih rezultata

**5. Selekcija atributa i interpretacija dobijenih rezultata**

- Sprovesti različite metode selekcije atributa na izabranim algoritmima

- Uporediti dobijene rezultate

**6. Optimizacija parametara i interpretacija dobijenih rezultata**

- Sprovesti optimizaciju parametara na izabranim algoritmima

- Uporediti dobijene rezultate

**7. Zaključak i ideje koje biste realizovali da imate više vremena/znanja**

- Opisati kako i koji biste model primenili u praktičnoj primeni

- Opisati šta mislite da bi bilo korisno sprovesti ali ste zbog nedostatka vremena ili znanja taj deo preskočili

Na odbranu projektnog zadatka donosite izveštaj (ne mora da se štampa, pošalite ga nastavnicima i saradniku e-poštom i okačite ga u Microsoft Teams na predviđeno mesto) i Jypiter notebook-e na Vašem računaru, koji takođe šaljete e-poštom i kačite na Microsoft Teams na predviđeno mesto.

Odbrane se organizuju u junskom ispitnom roku u dva termina.

O terminima odbrana projektnog zadatka u narednim rokovima bićete blagovremeno obavešteni.

Uslov za izlazak na odbranu projektnog zadatka su odbranjena oba domaća zadatka.

Ukoliko imate bilo kakvih nedoumica u vezi sa podacima ili sa zahtevima slobodno postavite pitanje na Teams platformu.