

# Mašinsko učenje - Praktičan deo - Domaći zadaci

## 1. Organizacija

**Praktičan deo kursa je organizovan sa ciljem da se kroz timski rad i rešavanje realnih problema savladaju koncepti, elementi i tehnike mašinskog učenja.** Domaći zadaci pokrivaju zasebne (pod)oblasti mašinskog učenja obrađene na predavanjima.

**Svaki od 6 zadataka dolazi sa svojim skupom podataka i kriterijumom koji je potrebno ispuniti da bi se zadatak smatrao uspešno urađenim. Gradivo obrađeno na predavanjima je dovoljno kako bi se zadatak uspešno uradio. Dodatnim istraživanjem (pod)oblasti i problema moguće je ostvariti bolje rezultate.** Rešenja (modeli) će se evaluirati na skrivenom testnom skupu podataka na online platformi. Opis same platforme dat je u sekciji 2.

**Određeni zadaci će zahtevati samostalno implementiranje algoritama mašinskog učenja.** Pomenuti algoritmi će biti detaljno obrađeni na predavanjima i materijali sa predavanja će sadržati njihove pseudokodove. Studentima se savetuje da redovno prate predavanja, pošto će se na vežbama podrazumevati da se gradivo sa predavanja zna. **Za izradu preostalih zadataka studentima će na raspolaganju biti Python biblioteke za mašinsko učenje.**

Uz svaki (urađeni) zadatak, timovi će predavati i propratni izveštaj na srpskom ili engleskom jeziku. Opširnije o samom izveštaju u sekciji 3.

**Najuspešniji timovi će biti oslobođeni izrade projekta.** Kriterijumi na osnovu kojih će se određivati najuspešniji timovi su:

- **Ostvareni rezultati i kako se do njih došlo:**
  - Pristup problemima
  - Korišćeni algoritmi
  - Određivanje (hiper)parametara algoritama
  - Rad sa trening skupom podataka
- **Propratni izveštaji:**
  - Sadržaj propratnih izveštaja
  - Usklađenost izveštaja i izvornih kodova rešenja
- **Diskusija:**
  - Presentovanje rešenja i odgovori na pitanja prilikom prezentovanja.

Takođe, **postoji mogućnost oslobađanja polaganja teorijskog dela ispita za najuspešnije timove koji pokažu izuzetno znanje i razumevanje prilikom prezentovanja i odbrane svojih rešenja.**

## 2. Platforma

Za izradu zadataka koristiti **Python 3.5.x**. Sama platforma “trči” na **Python 3.5.5**.

Timovima je dostupan **trening** i **deo test skupa podataka**. Rešenja se evaluiraju na **kompletnom test** skupu podataka. **Deo test** skupa je dostupan kako biste videli zaglavlje **kompletnog test** skupa i kako bi proverili da li će vaše rešenje moći da se izvrši na platformi.

Prilikom *submit*-a Python skripte sa rešenjem, platforma izvršava skriptu pomoću sledeće komande:

```
python <ime_skripte>.py <train_set_path> <test_set_path>
```

Lokaciji test skupa podataka (i trening skupa - ako je neophodno) se pristupa pomoću argumenata komandne linije (**sys.argv**).

Neophodno je da vaša skripta kao rezultat ispiše samo broj koji predstavlja rezultat, bez ikakvog pratećeg teksta:

- **Nepavilno: RMSE: 10.23**
- **Nepavilno: Dobijena vrednost je 10.23 (RMSE)**
- **Pravilno: 10.23**

Takođe, potrebno je zakomentarisati sva iscrtavanja (*plot*-ove) i one ispise (*print*-ove) koji ne predstavljaju ispis konačnog rešenja (rezultata).

Nakon što se skripta izvrši, ukoliko je sve prošlo bez problema, u desnom delu prozora (*My Submissions*) će biti prikazan ostvareni rezultat sa zelenom bojom kao pozadinom. Ukoliko nema rezultata, već samo crvena boja kao pozadina - pogledajte sekciju **2.3**.

Ukoliko *submit*-ujete skriptu koja radi samo sa test skupom (npr.: na trening skupu ste izračunali slope i intercept, “zakucali” ste ih u kodu i hoćete samo da evaluirate vaš regresioni model na kompletnom testnom skupu podataka) u skripti kao komentar ostaviti kompletan izvorni kod koji ste implementirali da dođete do tog rešenja.

**Svaki tim može najviše dva puta *submit*-ovati svoje rešenje za svaki zadatak. Skripta se na platformi izvršava najviše 180 sekundi, i ako se taj period prekorači, pokušaj će se računati kao neuspešan.**

Na stranici sa zadatkom, ispod dela za *submit*, nalazi se rang lista (*Leaderboard*). Moguće je videti rešenja drugih timova nakon isteka roka za izradu zadatka. **Ako nakon isteka roka za izradu zadatka element rang liste ima crvenu boju za pozadinu, to je indikator da je rešenje nakon provere od strane asistenta odbijeno.**

**Sav saobraćaj na platformi se prati (*log*-uje). Svaka neadekvatna upotreba platforme (pokušaj hakovanja, ulazak na druge naloge itd.) će biti sankcionisana.**

## 2.1. Biblioteke

Za prvi zadatak, na platformi će biti instalirane sledeće biblioteke (u zagradi su date tačne verzije biblioteka):

- **numpy (1.16.2)**
- **pandas (0.24.2)**.

Za drugi zadatak, na platformi će biti instalirane sledeće biblioteke:

- **numpy (1.16.2)**
- **pandas (0.24.2)**
- **scipy (1.2.1)**.

Počevši od trećeg zadatka, na platformi će biti instalirane sledeće biblioteke:

- **numpy (1.16.2)**
- **pandas (0.24.2)**
- **scipy (1.2.1)**
- **scikit-learn (0.20.3)**.

**Prilikom *submit*-a rešenja na platformu, studenti mogu da koriste sve *core* pakete Python-a 3.5.x i gorenavedene biblioteke.** Prilikom analize problema, analize skupa podataka, vizualizacije podataka i izrade propratnog izveštaja studenti mogu da koriste biblioteke po izboru, ali je neophodno da njihovi *import*-i i upotreba u *submit* skripti budu zakomentarisani, kako bi se ona bez problema izvršila na platformi.

## 2.2. Analiza plagijata

**Platforma nakon isteka roka za izradu zadatka vrši analizu plagijata. Izvorni kodovi će se analizirati zajedno sa ostalim kodovima iz generacije, kao i sa izvornim kodovima starijih generacija. Plagijat znači automatsku diskvalifikaciju i sankcije za plagijatore.**

## 2.3. Najčešće greške

Ukoliko nakon *submit*-a skripte u desnom delu prozora ne dobijete rešenje, već samo crvenu boju pozadine, proverite sledeće stvari:

- Da li koristite odgovarajuću verziju Python-a;
- Da li se skripta može izvršiti putem komande:  
**python <ime\_skripte>.py <train\_set\_path> <test\_set\_path>**  
npr.: **python skripta.py podaci/treniranje.csv podaci/testiranje.csv**
- Da li se kao rezultat izvršavanja vraća samo broj koji predstavlja rešenje, bez pratećeg teksta i suvišnih iscrtavanja i ispisa;
- Da li su zakomentarisani *import*-i i upotreba biblioteka koje nisu instalirane na platformi za taj zadatak;
- Da li izvršavanje skripte traje kraće od 180 sekundi.

**Ukoliko i pored provere ne uspete da utvrdite problem, pošaljite asistentu e-mail.**

### 3. Izveštaj

Uz svaki zadatak, a pre isteka roka za izradu istog, neophodno je poslati izveštaj u kome ćete opisati:

- Vaš pristup problemu
- Isprobane algoritme
- Ostvarene rezultate
- Konačno (odabrano) rešenje.

Grafici i opisi preprocesiranja podataka su dobrodošli.

Izveštaj slati u **PDF** formatu na **vdragan@uns.ac.rs** na sledeći način:

- *Email Subject*: **ML22-Z<redni\_broj\_zadatka>-<username\_sa\_platforme>**
- *Email Body*: prazan ili sa porukom po izboru
- *Attachment*: Izveštaj u PDF formatu
- PDF nazvati po šablonu:

**ML22-Z<redni\_broj\_zadatka>-<username\_sa\_platforme>.pdf**

**Izveštaji poslati nakon isteka roka za izradu zadatka će biti odbačeni.**

**Rešenja bez izveštaja će biti odbačena.**