

Istraživanje podataka 1

O projektima

Prijava teme	2
Preduslovi za prijavljivanje projekata	2
Prijava teme za projekat	2
Otvaranje repozitorijuma i rad na projektu	2
Odbrane projekata	3
Očekivanja od implementacije	3
Očekivanja od izveštaja	4
Očekivanja od git repozitorijuma	4
Način bodovanja (moguće su manje izmene)	5

Opšte informacije

Projekat na kursu "Istraživanje podataka 1" ima za cilj primenu različitih modela u cilju rešavanja osnovnih problema nadgledanog i nenadgledanog učenja.

Projekti se zvanično vode kao predispitna obaveza pa se zato moraju završiti i odbraniti pre izlaska na pismeni deo ispita. Jednom odbranjen projekat važi do kraja školske 2022/2023 godine.

Prijava teme

Preduslovi za prijavljivanje projekata

Svaki student mora da otvori svoj GitHub nalog, ukoliko ga već ne poseduje. Obratite pažnju na sledeće elemente prilikom otvaranja novog naloga:

- Adresa elektronske pošte - na ovu adresu će stići pozivnica za priključivanje projektu
- Korisničko ime - ovo ime se koristi u prijavi projekta

Dodatno, potrebno je da se upoznate sa *git* sistemom za kontrolu verzija. Dobar materijal za upoznavanje kako *git* funkcioniše kroz niz primera se može pronaći [na veb prezentaciji Git How To](#).

Prijava teme za projekat

Spisak tema za školsku 2022/2023 se može pronaći na narednom [linku](#). Pored izabrane teme u tabeli upišite tražene informacije. Projekat nije moguće raditi u paru/timu.

Temu je moguće prijaviti do kraja školske godine.

Temu možete i sami da predložite. Linkovi gde možete pronaći skupove: [UCI](#) i [Kaggle](#). Uslov je da skup ima minimum 10 atributa (kolona) i 1000 instanci (redova). Za dogovor poslati mejl asistentima.

Otvaranje repozitorijuma i rad na projektu

Nakon prijave teme, prvo će biti poslat poziv za pridruživanje organizaciji na githubu, koji morate prihvatiti u roku od 7 dana od slanja

Nakon prijave teme, repozitorijumi će biti otvoreni od strane asistenata. Nije ni potrebno ni dozvoljeno da studenti samostalno kreiraju repozitorijume za projekte na GitHub servisu.

Takođe, isključivo koristite dodeljeni repozitorijum za izradu projekta.

Ukoliko želite da se tokom rada oko nečega konsultujete sa asistentima, pored diskusije online ili putem mejla, možete otvoriti novu primedbu (*issue*) u okviru vašeg GitHub repozitorijuma i dodeliti (*assign*) tu primedbu asistentima. U tekstu primedbe je neophodno da detaljno opišete za šta vam je potrebna konsultacija. Poželjno je da primedbu označite labelom za potraživanje pomoći.

Predaja projekata pred odbranu

Za predaju projekta, potrebno je pripremiti i postaviti u repozitorijum projekta naredne elemente:

- Izveštaj - LaTeX ili word.
- Izveštaj u pdf-u.
- Prezentacija za odbranu (max 15 slajdova).
- Svi istrenirani modeli sačuvani u odgovarajućem formatu.
- Svi kodovi.

Ukoliko ste projekat razvijali na drugim granama, molimo vas da integrišete sve izmene na main granu.

Odbrane projekata

Odbrane projekata će biti održavane u svim ispitnim rokovima (nekoliko dana pre pismenog ispita čiji termin održavanja je naznačen u rasporedu ispita). Odbrana će se sastojati od:

- Prezentacija projekta (ukratko opis problema, podataka, modela, rezultata) - maksimum 10 minuta.
- Pitanja vezana za način izrade projekta, motivaciju iza korišćenih modela, implementaciju, kao i uopštena pitanja o gradivu sa vežbi.

Očekivanja od projekta

U nastavku navodimo nekoliko najvažnijih elemenata kojima se treba voditi prilikom izrade projekta.

Očekivanja od implementacije

Obavezni elementi:

- Eksplorativna analiza skupa podataka.
- Pretprocesiranje (rad sa nedostajućim vrednostima, autlajerima, feature selection, feature extraction, balansiranje...).

- **Klasifikacija** - 2+ algoritma (optimizacija hiper parametara, potencijalno korišćenje ansambala, nije dovoljno samo iskoristiti fit/predict). Obavezno poređenje modela.
- **Klasterovanje** - 2+ algoritma (posebna priprema podataka za primenu algoritama). Poređenje modela.
- **Pravila pridruživanja** - 1+ alg u SPSS-u.
- Obavezna vizuelizacija svih rezultata.

Ukoliko preuzimate tuđi deo koda - obavezno obeležite i naznačite odakle je kod preuzet!

Studenti se podstiču i na samostalno istraživanje, primenu modela i tehnika za analizu podataka koji nisu pokriveni gradivom vežbi, mada to nije neophodno za uspešnu odbranu projekta.

Očekivanja od izveštaja

Uz implementaciju je neophodno priložiti i izveštaj o toku izrade projekta. Preporučuje se korišćenje LaTeX-a. Izveštaj treba da se sastoji od:

- Uvod - opis skupa podataka i problema.
- Opis modela - motivacija za odabir modela, opis problema prilikom implementacije i način rešavanja...
- Rezultati - tabele/slike, diskusija
- Zaključak

Ukoliko niste upoznati sa LaTeX-om preporučujemo narednu literaturu:
[LaTeX za autore](#), kao i [Overleaf - LaTeX editor](#).

Očekivanja od git repozitorijuma

- Nazivi datoteka moraju biti smisleni.
- Skup podataka se mora nalaziti na repozitorijumu.
- Preporučuje se naredna organizacija repozitorijuma:
 - models/
 - classification/
 - Knn.ipnyb
 - ...
 - clustering/
 - Kmeans.ipnyb
 - ...
 - association_rules/
 - Apriori.ipnyb

● ...

- README treba da sadrži:
 - informacije o autoru
 - Ukoliko koristite dodatne biblioteke za rad - informacije o njima i način instaliranja.

Način bodovanja (moguće su manje izmene)

- Projekat nosi ukupno 30 poena:
 - Klasifikacija - 8 poena
 - Klasterovanje - 8 poena
 - Pravila pridruživanja - 4 poena
 - Analiza i pretprocesiranje - 5 poena
 - Opšti utisak - 5 poena