

Projekti na predmetu:

Strojno učenje – PMF (2010)

Tomislav Šmuc; Matko Bošnjak

Kao projekte u okviru ovog predmeta predlažemo učešće na tzv. "challenge-e" u području data mining-a. Prijedlog je da se organizirate u dvije do tri ekipe od po 2-3 člana. Možete izabrati bilo koji od dolje predložena 2 challenge-a (ili oba?!), ali pazite na vremenska ograničenja. Dobro proučite uvjete natjecanja. Predviđamo da će vam za dolaženje do relativno kvalitetnih rješenja trebati 2-3 tjedna.

Datumi:

- a) formiranje ekipa i odabir zadatka – prijava predavanju 19.04. ili najkasnije na e-mail (21.04.2010.)
- b) predaja izvještaja 25.05.2010.
- c) prezentacija 28.05.2010.

Za one koji baš ne žele sudjelovati u ovim projektima – ostavljamo mogućnost da naprave predavanje o posebnim područjima primjene algoritama strojnog učenja (vidi pod II).

I) Međunarodna natjecanja vezana uz data mining

DATA-MINING-CUP

Izazov: Modelirati maksimizaciju profita temeljenu na inteligentnoj dodjeli kupona.

Koristeći dostupne podatke o narudžbama pojedinih korisnika određene online trgovine (količina artikala, neotkrivajući podaci o korisniku, podaci o grupama naručenih artikala...itd) donijeti odluku da li poslati korisniku kupon u vrijednosti od 5€ ili ne. Cilj je poslati kupone onim osobama koje ne bi ponovno naručivale s iste trgovine u svrhu poticanja istih da ponovno troše novce.

Rok za slanje: 31.05.2010.

Nagradni fond: 1. mjesto: 2500€
2. mjesto: 1500€
3. mjesto: 1000€

Ako osvojite jedno od prva tri mjesta, također imate osiguranu ocjenu iz predmeta. Dodatni bodovi će se dodjeljivati u ovisnosti o vašoj poziciji na ukupnoj rang listi.

Uvjeti: Sudionici moraju biti isključivo studenti.

Dodatni uvjet: Morate sastaviti kratki izvještaj o tome što ste napravili za rezultat koji ste postigli.

Detalji na: <http://www.data-mining-cup.de/en/dmc-competition/>

PAKDD-2010 Data Mining Competition

Izazov: Rekalibracija sustava za procjenu kreditnog rizika na pristranim podacima

Skup podataka za učenje sastoji se od 50000 klijenata koji su dobili određeni kredit s određenim (neotkrivajućim) informacijama o samim klijentima kao i o konačnoj informaciji da li su uspjeli vratiti kredit na vrijeme ili nisu. Uz navedeni skup podataka možete skinuti i „leaderboard“ skup od 20000 klijenata čiju klasifikaciju možete poslati te uživo vidjeti kako ste pozicionirani prema drugim timovima na „leaderboardu“. U konačnici je potrebno klasificirati novi skup podataka od 20000 klijenata koji se sastoje od opće populacije (ljudi kojima je sustavom za procjenu kreditnog rizika odobren kredit, ali i onima kojima nije, u svrhu kontrole rada samog sustava).

Rok za slanje: 03.05.2010.

Nagradni fond: Nije objavljen, no vrijedi isto kao i za prethodni što se propozicija predmeta tiče: ako osvojite jedno od prva tri mjesta, imate osiguranu ocjenu iz predmeta. Dodatni bodovi će se dodjeljivati u ovisnosti o vašoj poziciji na ukupnoj rang listi.

Uvjeti: Uz finalne rezultate morate poslati i kratki izvještaj o tome što ste radili s podacima. Taj izvještaj je nužan da biste bili prikazani na konačnoj rang listi (i uvjet je za ocjenu vašeg projekta).

Detalji na: <http://sede.neurotech.com.br/PAKDD2010/>

NAPOMENE: U oba izazova se natječete sa svim profilom ljudi, od početnika poput vas do iskusnih znalaca u svom području. Neka vas loš plasman ne zabrinjava, a odličan plasman neka vam svakako laska.

Potrebno je pridržavati se pravila propisanih odgovarajućim izazovom, sastavi timova su proizvoljni. Dozvoljeno je sve, svaka nova ideja za rješavanje zadatka. Kreativnost se dodatno nagrađuje.

Dodatne informacije i konzultacije:

- e-mail (matko.bosnjak@irb.hr; tomislav.smuc@irb.hr)
- u sklopu termina vježbi - petkom

II) **Predavanja o posebnim temama**

Oni koji odaberu ovu opciju, trebaju napraviti predavanje o temi (mi ćemo ponuditi literaturu) o jednoj temi, prema izboru. Predviđeno je da ta predavanja budu u sklopu termina predviđenog za predavanja – krajem mjeseca svibnja. Ovo su naši prijedlozi:

- a) Obrada teksta (text mining)
- b) Sustavi za predlaganje/preporučivanje (Recommender systems & collaborative filtering)
- c) Strojno učenje u bioinformatici i računalnoj biologiji
- d) ... po vašem izboru – uz dogovor