



Da li ste spremni za kolačiće?

U skladu sa praznicima, zadatak ovog domaćeg je da se napravi klasifikator koji će na osnovu prisustva/odsustva određenih sastojaka odrediti da li je određen recept za pravljenje kolačića, peciva ili pice.

Zadatak: Napraviti 2 različita tipa klasifikatora koji će na osnovu prisustva/odsustva određenih sastojaka odrediti da li je u pitanju recept za kolačić (*cookie*), za pecivo (*pastries*) ili picu (*pizzas*). Prikupljeno je 1931 recept i za svaki je naznačeno prisustvo (1) ili odsustvo (0) svakog od 133 sastojka. Dat je skup za obuku (*cookies_train.csv*) i test skup (*cookies_test.csv*) na kom treba testirati klasifikatore i uporediti ih.

1. Za svaku klasu napraviti karakterističan obrazac, kao **histogram pojavljivanja određenih sastojaka**. Iskomentarisati i uporediti ih za različite klase.
2. Broj svog indeksa podeliti po modulu 6 - dobijeni broj označava par klasifikatora koje treba da koristiš:
 0. kNN i logistička regresija
 1. kNN i stablo odluke
 2. kNN i SVM
 3. SVM i logistička regresija
 4. SVM i stablo odluke
 5. Stablo odluke i logistička regresija
3. Za svaki od dodeljenih klasifikatora uraditi sledeće:
 - a. Koristeći metodu unakrsne validacije sa 10 podskupova odrediti optimalne parametre oslanjajući se na željenu meru uspešnosti. Obratiti pažnju da u svakom od podskupova za unakrsnu validaciju bude dovoljan broj uzoraka svake klase.
 - b. Za konačno odabrane parametre prikazati i analizirati matricu konfuzije dobijenu akumulacijom matrica iz svake od 10 iteracija unakrsne validacije. Na osnovu matrice konfuzije izračunati prosečnu tačnost klasifikatora, kao **i tačnost za svaku klasu**.

- c. Klasifikator sa konačno odabranim parametrima obući na celokupnom trening skupu, a potom testirati na datom test skupu.
 - d. Izračunati matricu konfuzije na osnovu rezultata na test skupu, te izračunati prosečnu tačnost klasifikatora i **tačnost za svaku klasu**.
 - e. Uporediti rezultate unakrsne validacije sa rezultatima dobijenim na test skupu.
4. Uporediti konačna dva klasifikatora po odzivu, specifičnosti i preciznosti, za svaku klasu posebno, kao i prosečnu tačnost klasifikatora, mikro i makro preciznost, kao i mikro i makro F-meru. Iskomentarisati prednosti i mane svakog od klasifikatora.
 5. Sve rezultate prikazati i diskutovati u izveštaju (2-4 strane).

Za sva eventualna pitanja, nejasnoće ili ako smatrate da je traženo nešto što se ne može uraditi ili deluje preterano zahtevno, obratiti se mailom na tijana.delic92@gmail.com. Pri pisanju izveštaja pratiti uputstva koja su data. **U izveštajima ne treba objašnjavati kod niti ga prepisivati, akcenat je na interpretaciji rezultata.** Ako je ostalo bilo šta nejasno povodom pisanja izveštaja, stojim na raspolaganju. Izveštaj se ne može napisati za sat-dva, tako da ostavite sebi dovoljno vremena da ga uradite kvalitetno. **Putem moodla najkasnije do 23:59 15.01.2021. treba predati 2 fajla: skriptu** koja sadrži kod (.py ili .ipynb, ako imate više skripti, smestite sve u jednu) **i izveštaj** (u .pdf formatu). Možete pisati kod u PyCharmu, Jupyteru, Colabu, Spyderu... Domaći radite samostalno – **dva ista koda ili dva ista izveštaja dobijaju 0 bodova bez daljeg istraživanja kako su nastali.**