

1. Zadaća, Inteligentni sustavi 2, 0-8 bodova,

Zadavanje zadaće:

03.04.2020. rok za predaju: 16.04.2020.

Kao što je rečeno u DINP-u, studenti mogu diskutirati o zadacima koje rješavaju, ali svatko treba napisati svoju zadaću sam. Identične zadaće se ocjenjuju sa 0 bodova.

Zadatak:

Izvršiti istraživačku analizu podataka (EDA) za zadani skup podataka sa ciljem otkrivanja utjecaja varijable x na atribut klase grade.

- a) (2 boda) U analizi uzeti u obzir utjecaj same varijable x (vizualizacija) uz kratku interpretaciju grafičkih prikaza. Analizu napraviti za skup „edukacija90.csv“ (tri vrijednosti atributa klase).
- b) (2 boda) Provjeriti i utjecaje varijable x u kombinaciji sa drugim varijablama: primjerice dijagram raspršenja za dvije varijable uz označavanje vrijednosti atributa klase različitom bojom (interpretirati ukratko grafičke prikaze). Analizu napraviti za skup „edukacija90.csv“ (tri vrijednosti atributa klase).
- c) (2 boda) Izvesti logističku regresiju na skupu „edukacija88.csv“ (dvije vrijednosti atributa klase) za predviđanje vjerojatnosti vrijednosti klase za razne kombinacije prediktora - varijabli)
- d) (2 boda) Sažeti najvažnije rezultate analize na najviše jednoj stranici (grafički prikazi i interpretacije – simulirati eksperta u smislu zaključivanja o utjecaju varijable-aktivnosti x na uspjeh studenta). Iako je naglasak istraživanja na varijabli x opišite i značajna opažanja o drugim prediktorima-varijablama. **Priložiti skriptu s kodom uz komentare koda da bi se moglo ponoviti izvođenje postupaka za dobivanje najznačajnijih rezultata analize.**

Svaki student analizira podatke s naglaskom na jednoj varijabli x . Koja varijabla x vam je dodijeljena provjerite u priloženoj tablici (IS2 popis.pdf). Pri analizi se koristite materijalima sa predavanja i vježbi, zadanom literaturom, radovima studenata, izvorima sa Interneta. Referirajte se na izvore koje koristite pri izradi zadaće, te sve korištene izvore navedite numerirane na kraju zadaće. Slobodno primijenite i postupke koje niste upoznali na nastavi uz navođenje izvora. Ukoliko razmjenjujete i koristite informacije odnosno rezultate analize sa kolegama pri izradi zadaće, možete se i referirati na njihove zadaće (osim ako vam je naglasak na istoj varijabli).

Opis skupa podataka možete naći u diplomskom radu studenta Josipa Čanića na Merlinu [1]. Skup podataka edukacija88.csv dobiven je na temelju skupa edukacija90.csv zamjenom vrijednosti atributa klase GOOD u PASS.

[1] Josip Čanić: Primjena logističke regresije u znanosti o podacima, diplomski rad, srpanj 2017.