

Studijski program:

Datum:

Broj indeksa:

Ime i prezime:

Prvi računarski kolokvijum iz predmeta

## Numerički algoritmi i numerički softver

### Grupa C\_2

Dat je skup podataka o zagađenosti vazduha (*AirPollution*). Cilj je napraviti model višestruke linearne regresije koji na osnovu podataka iz *data/train.csv* fajla predviđa *AirPollution*.

Kao metriku u zadacima koristiti  $r^2$  prilagođeni, osim ako nije drugačije naznačeno.

Uzeti pouzdanost od 95% za sve statističke testove, osim ako nije drugačije naznačeno.

#### Zadatak 1.

Napraviti model linearne regresije koji koristi sve atribute. Ispisati meru koju model ostvaruje.

#### Zadatak 2.

Ako se gustina saobraćaja (TrafficDensity) poveća za jedan za koliko se minimalni i maksimalno poveća zagađenje (AirPollution)? Da li su validne nadjene vrednosti?

#### Zadatak 3.

Unaprediti model koliko god je to moguće tako da su sve prepostavke zadovoljene i ostvaruje  $r^2$  prilagođeni  $\geq 0.50$ . Koristiti *train/val* podelu u odnosu: 90/10. Ispisati meru koju model ostvaruje. Napomena: koristiti `random_state = 42` (kako bi se rezultati mogli reproducovati).

#### Zadatak 4.

Objasnite čemu služi t-test.

#### Zadatak 5.

Skicirajte interpolaciju linearnim splajnom. Šta je problem kod linearog splajna?