

Studijski program:

Datum:

Broj indeksa:

Ime i prezime:

Prvi računarski kolokvijum iz predmeta

Numerički algoritmi i numerički softver

Grupa E_1

Dat je skup podataka o broju bodova na testu (*test_score*). Cilj je napraviti model višestruke linearne regresije koji na osnovu podataka iz *data/train.csv* fajla predviđa *test_score*.

Kao metriku u zadacima koristiti *RMSE*, 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 produži učenje sa jedan sat (*study_hours*) za koliko se očekuje minimalno i maksimalno povećanje u test skoru? Obrazloži da li su dobijene vrednosti validne.

Zadatak 3.

Unaprediti model koliko god je to moguće tako da su sve prepostavke zadovoljene i ostvaruje $RMSE < 254$. Koristiti *train/val* podelu u odnosu: 80/20. Ispisati meru koju model ostvaruje.

Napomena: koristiti random state = 50 (kako bi se rezultati mogli reprodukovati).

Zadatak 4.

Objasnite čemu služi koeficijent determinacije (R^2).

Zadatak 5.

Opišite osnovnu ideju interpolacije splajnom.