* Podatci korišteni u razdoblju od 1996. do 2022. godine: postotci nezaposlenih, podatci o uvozu, izvozu i proizvodnji, indeksi o prosječnoj neto plaći ??(jel uopce radim ove?)
* R2 – koeficijent determinacije, mjeri koliko dobro prediktivni model objašnjava varijabilnost podataka, negativne vrijednosti znače da model ima odstupanja veća od prosječne vrijednosti ciljne varijable, pozitivne vrijednosti su moguće od 0 do 1, gdje 1 označava model koji savršeno predviđa, odnosno njegove predikcije jednake su stvarnim vrijednostima ??? (zbrisi sve za r2)

(Pozitivna koreliranost označava proporcionalnost s brojem umrlih, a negativna obrnutu proporcionalnost. Veći apsolutni iznos koeficijenta korelacije označava veću koreliranost.)?

### Linearna regresija za broj rođenih

Ulazni podatci sastoje se od sljedećih varijabli: godina, stanovništvo najvećeg grada, % ruralnog stanovništva, stanovništvo(15-64), žensko stanovništvo(20-24), BDP.

Najbolji rezultat postignut je sa varijablama godina i BDP na razdoblju od 1961. do 2021. godine. Model ima sljedeće performanse: ??(promijeniti eng. nazive u hrvatske)??

* R2: 0.9710
* MAE: 1859.80
* MSE: 5232963.26
* RMSE: 2287.57
* MAPE: 3.32

Model koji je postigao najbolje rezultate bez varijable godine, koristio je sljedeće varijable: stan\_grad, Dob\_15\_64, DobŽ\_20\_24, Rural% u razdoblju od 1961. do 2021. godine. Performanse takvog modela su:

* R2 score: 0.8345
* Mean Absolute Error (MAE): 4334.76
* Mean Squared Error (MSE): 29878866.58
* Root Mean Squared Error (RMSE): 5466.16
* Mean Absolute Percentage Error (MAPE): 8.09

Iako su performanse dobre, model uvelike ovise o veličini skupa podataka za testiranje i podataka za treniranje. Za promjene od samo 0.05% veličine, R2 modela opada na 0.47 što je velika razlika u odnosu na početni R2 od 0.84. Na sljedećem grafu možemo vidjeti stupanj ovisnosti modela o varijablama. ??graf?

A graph with green and blue squares

Description automatically generated

Ostali podatci dostupni su u manjem vremenskom razdoblju, od 1986. odnosno 1995. godine, no na takvim manjim vremenskim razdobljima linearna regresija nije uspjela pružiti zadovoljavajuće performanse modela.

### Linearna regresija za broj umrlih

Za ostvarenje dobrih performansi, linearnoj regresiji za broj umrlih bilo je potrebno puno veći broj varijabli. Performanse najboljeg modela su: ?(zaokruzit i hrv nazivi)???

* R2 score: 0.905746080194187
* Mean Absolute Error (MAE): 1443.7798382224282
* Mean Squared Error (MSE): 3691883.6976246587
* Root Mean Squared Error (RMSE): 1921.4275155791483
* Mean Absolute Percentage Error (MAPE): 3.122035271334225

Na sljedećem grafu vide se ulazne varijable na y-osi (u razdoblju od 1961. do 2021. godine) te ovisnost modela o njima na x-osi. Iako je ovisnost varijabli neto migracije i BDP-a toliko malena da se ne vidi na grafu, izbacivanje jedne od tih varijabli iz modela dovelo bi do smanjenja R2 na 0.85. Uklanjanjem varijable „godina“ performanse su zanemarivo lošije.

A graph with different colored squares

Description automatically generated

Linearna regresija za broj umrlih ostvarila je dobre performanse i na manjem vremenskom razdoblju podataka, no uz iznimno velike razlike u performansi kod male promjene veličine skupa podataka za treniranje i veličine skupa podataka za testiranje. To ukazuje na loš model s mogućim problemom prenaučenosti ili nedovoljne generalizacije.

Sličan problem se pokazivao i kod linearne regresije za broj umrlih na razdoblju od 1961. do 2021. godine, ali u puno manjoj mjeri.

To može ukazivati na veliki broj odstupanja i nelinearnosti u podatcima.

NOVI REZULTATIIII: