

# Predviđanje cene nekretnina

Upoređivanje rezultata linearne regresije, decision tree, random forest i neuronske mreže

Autor: Marko Šuljak

## Skup podataka:

U ovom projektu bavio sam se problemom određivanja cene nekretnina pomoću informacija koje su nam o njoj dostupne, kao i upoređivanjem efikasnosti različitih tehnika za rešavanje ovog problema. Za izradu projekta korišćen je *Python*.

## Opis problema:

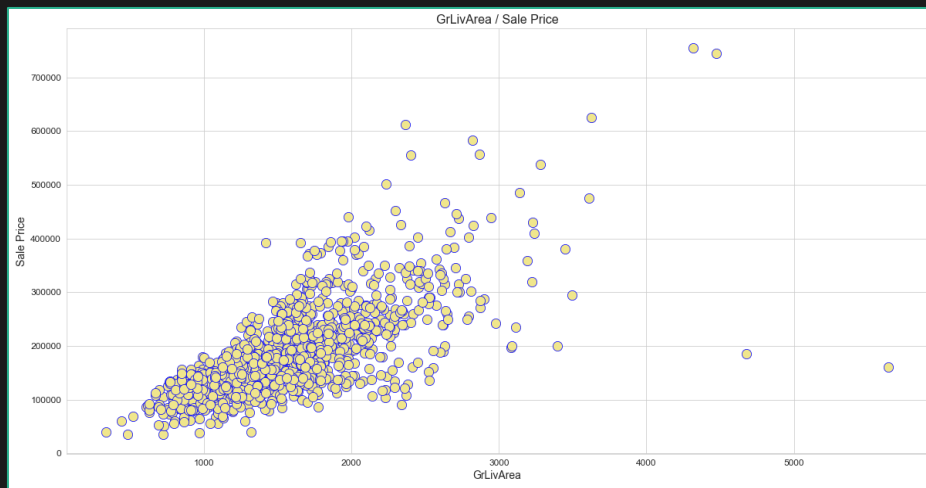
Projektni zadatak koristi bazu podataka koja sadrži 1500 različitih nekretnina sa preko 70 podataka o svakoj. Odmah pri učitavanju ovih podataka, dataset delimo u 2 grupe i to:

- Trening skup
- Test skup

Pošto skup podataka sadrži veoma mnogo nedefinisanih i nula polja, ona se odmah po učitavanju odbacuju kako bi aplikacija mogla da koristi bazu.

Od svih podataka koji se nalaze u bazi, najviše su nam značajni oni koji govore o površini nekretnine i placa pa smo njih izdvojili. Zatim se pomoću ovih podataka vrši predviđanje cene nekretnine sledećim tehnikama:

- **Linearna regresija**
- **Decision tree**: algoritam za klasifikaciju i predviđanje koji ima strukturu stabla gde svaki čvor predstavlja neku „odluku“ i povezan je grana koje su ishodi odluka.
- **Random forest**: predstavlja skup različitih decision tree-jeva koji se treniraju „bagging“ metodom. Ovaj algoritam zapravo povezuje stabla odluka kako bi povećao uspešnost.
- **Extra tree**: veoma sličan random forest algoritmu, ali razlika je u tome da koristi ceo skup podataka dok random forest koristi delove kako bi povećao diverzitet i time smanjio promenljivost rezultata u iteracijama.
- **Neuronska mreža**: serija različitih algoritama koji otkrivaju povezanosti podataka u bazi. Oponaša način funkcionisanja ljudskog mozga.



## Rezultati:

- **Linearna regresija**:
  - Koeficijent determinacije: 0.772
  - Odstupanje od cene: 2475.7
- **Decision tree**:
  - Koeficijent determinacije: 0.624
  - Odstupanje od cene: 3951
- **Random forest**:
  - Koeficijent determinacije: 0.806
  - Odstupanje od cene: 1877.9
- **Extra tree**:
  - Koeficijent determinacije: 0.813
  - Odstupanje od cene: 1951.2
- **Neuronska mreža**:
  - Koeficijent determinacije: 0.518
  - Odstupanje od cene: 6094.2

