

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 791

**PREDVIĐANJE CIJENA NEKRETNINA REGRESIJSKIM  
METODAMA STROJNOG UČENJA**

Andrea Fusek

Zagreb, srpanj 2025.

SVEUČILIŠTE U ZAGREBU  
FAKULTET ELEKTROTEHNIKE I RAČUNARSTVA

DIPLOMSKI RAD br. 791

**PREDVIĐANJE CIJENA NEKRETNINA REGRESIJSKIM  
METODAMA STROJNOG UČENJA**

Andrea Fusek

Zagreb, srpanj 2025.

## **DIPLOMSKI ZADATAK br. 791**

Pristupnica: **Andrea Fusek (1191239402)**

Studij: Računarstvo

Profil: Znanost o podacima

Mentor: izv. prof. dr. sc. Alan Jović

Zadatak: **Predviđanje cijena nekretnina regresijskim metodama strojnog učenja**

### Opis zadatka:

Predviđanje cijena nekretnina složeni je zadatak koji treba uzeti u obzir povijesne statističke podatke, aktualne trendove i stvarne cijene nekretnina koje se postižu u prodaji. U ovom diplomskom radu potrebno je razmotriti i opisati regresijske metode strojnog učenja (npr. linearna regresija, uzdizanje gradijenta za regresiju, stroj s potpornim vektorima za regresiju) te izgraditi modele za predviđanje cijena nekretnina u Hrvatskoj. Pritom je potrebno istražiti predviđanje indeksa cijena nekretnina na mjesečnoj, tromjesečnoj i godišnjoj razini kao i predviđanje cijene individualne nekretnine na temelju njezinih značajki. Da bi se to ostvarilo potrebno je uzeti u obzir podatke iz Državnog zavoda za statistiku kao i dostupne podatke iz hrvatskih internetskih portala koji se bave ovim pitanjem. Potrebno je podržati učitavanje podataka, izgradnju modela strojnog učenja te njihovu međusobnu usporedbu po pitanju mjera srednje kvadratne pogreške. Rezultate je potrebno vizualizirati na odgovarajući način te objasniti dobivene modele strojnog učenja u kontekstu najvažnijih značajki korištenih u njihovom učenju.

Rok za predaju rada: 4. srpnja 2025.

*Za moje roditelje, koji su uvijek bili tu za mene.*

## Sadržaj

|                                                                 |    |
|-----------------------------------------------------------------|----|
| Uvod .....                                                      | 1  |
| 1. Tržište nekretnina i metode predviđanja cijena .....         | 3  |
| 2. Regresijske metode strojnog učenja .....                     | 5  |
| 2.1. Linearna regresija .....                                   | 5  |
| 2.2. Gradijentno poboljšavanje modela .....                     | 6  |
| 2.3. Stroj s potpornim vektorima .....                          | 8  |
| 2.4. Slučajne šume .....                                        | 10 |
| 2.5. Stablo odluke .....                                        | 11 |
| 3. Podaci i obrada .....                                        | 13 |
| 3.1. Indeksi cijena nekretnina .....                            | 13 |
| 3.2. Tržišni podaci .....                                       | 14 |
| 4. Izgradnja i vrednovanje modela .....                         | 16 |
| 4.1. Priprema podataka za modeliranje .....                     | 16 |
| 4.2. Učenje modela i rezultati .....                            | 16 |
| 4.2.1. Izgradnja modela metodom linearne regresije.....         | 17 |
| 4.2.2. Izgradnja modela metodom gradijentnog poboljšanja.....   | 18 |
| 4.2.3. Izgradnja modela metodom stroja potpornih vektora.....   | 18 |
| 4.2.4. Izgradnja modela metodom slučajne šume.....              | 19 |
| 4.2.5. Izgradnja modela metodom stabla odluke .....             | 20 |
| 4.3. Vrednovanje modela.....                                    | 21 |
| 5. Normalizacija tržišnih cijena i povezivanje s indeksima..... | 23 |
| 5.1. Predviđanje cijena nekretnina u drugoj polovini 2025.....  | 24 |
| Zaključak .....                                                 | 25 |
| Literatura .....                                                | 26 |
| Popis tablica.....                                              | 29 |

|                  |    |
|------------------|----|
| Popis slika..... | 30 |
| Sažetak.....     | 31 |
| Summary.....     | 32 |

# Uvod

Predviđanje cijena nekretnina izazovan je analitički problem koji obuhvaća širok spektar međusobno povezanih čimbenika, od makroekonomskih pokazatelja do fizičkih karakteristika pojedinih nekretnina. Iako se intuitivno može činiti kao problem koji se može riješiti jednostavnom regresijskom analizom, dinamika tržišta i složenost odnosa među varijablama čine točnu procjenu iznimno zahtjevnom.

Precizne procjene tržišne vrijednosti nekretnina od velike su važnosti ne samo za kupce i prodavatelje, već i za investitore, financijske institucije te tijela zadužena za oblikovanje stambene i gospodarske politike. Tradicionalne metode, temeljene na linearnim statističkim modelima, često nisu dovoljne za prepoznavanje nelinearnih odnosa i skrivenih obrazaca u podacima. Zbog toga se algoritmi strojnog učenja, poput linearne regresije, gradijentnog poboljšavanja modela (engl. *Gradient Boosting*) i strojeva s potpornim vektorima (engl. *Support Vector Machines*, SVM) pokazuju znatno učinkovitijima u modeliranju složenijih predikcija [1, 2].

U ovom radu istražuju se mogućnosti primjene regresijskih modela strojnog učenja u svrhu predviđanja cijena nekretnina na području Hrvatske. Modeli su učeni na indeksima cijena nekretnina iz baze Državnog zavoda za statistiku, kako bi se prepoznali dugoročni tržišni trendovi. Potom su predviđeni indeksi primijenjeni na konkretnim podacima s portala za oglašavanje nekretnina Njuškalo.hr, čime su povezane stvarne tržišne cijene s procijenjenim indeksima.

Analiza je provedena na kvartalnoj razini, a podijeljena je po regijama Hrvatske: Ukupno, Zagreb, Obala i Ostalo.

Cilj ovog rada je razviti i usporediti različite modele strojnog učenja za predviđanje cijena nekretnina te analizirati njihove prednosti i ograničenja u odnosu na tradicionalne pristupe.

Prvi dio rada uvodi osnovne pojmove tržišta nekretnina te objašnjava izazove i specifičnosti predviđanja cijena nekretnina u odnosu na tradicionalne ekonomske pristupe. U drugom dijelu detaljno su opisani izvori podataka, način prikupljanja i obrada podataka s

web oglasnika, kao i konstrukcija vremenskih nizova potrebnih za učenje modela. Treći dio posvećen je opisivanju podataka prikupljenih s web oglasnika te objašnjavanju rukovanja nad tim podacima kako bi se dobio skup pogodan za rad i usporedbu s indeksima. U četvrtom dijelu predstavljeni su korišteni modeli strojnog učenja te način na koji su učeni i vrednovani. Peti dio objašnjava i povezuje indekse cijena nekretnina i njihove predviđene vrijednosti s tržišnim cijenama s web oglasnika.

# 1. Tržište nekretnina i metode predviđanja cijena

Tržište nekretnina jedno je od temeljnih područja svake nacionalne ekonomije jer snažno utječe na investicije, zapošljavanje i financijsku stabilnost države. Nekretnine predstavljaju imovinu specifičnu po visokoj vrijednosti, niskoj likvidnosti i snažnoj ovisnosti o lokalnim faktorima, poput lokacije, prometne povezanosti, infrastrukture i regulatornog okvira.

Cijene nekretnina određuje niz međuvisnih čimbenika: makroekonomski pokazatelji (kamate, inflacija, nezaposlenost), demografske promjene, tržišna dinamika ponude i potražnje te individualne karakteristike same nekretnine (površina, starost, stanje, broj soba itd.). Zbog takve kompleksnosti, precizno predviđanje cijena zahtijeva modele sposobne obraditi nestrukturirane, heterogene i često nelinearne podatke [1].

Tradicionalni pristupi, osobito hedonistički regresijski modeli, pretpostavljaju linearnu vezu između atributa nekretnine i njezine tržišne cijene. Hedonistička regresija je metoda u kojoj je cijena nekretnina izražena kao funkcija njezinih karakteristika, odnosno broja soba, površine, udaljenosti od centra i sl., kako bi se odredilo koliko svaka od tih karakteristika utječe na ukupnu vrijednost. Na taj način cijena se „rastavlja“ na doprinos svake karakteristike, kako bi se lakše vidjelo što je zaslužno za višu ili nižu vrijednost. Iako su jednostavni za interpretaciju, takvi modeli često zanemaruju kompleksne interakcije među varijablama i teško se nose s tržišnim nestabilnostima [1]. Višestruka linearna regresija, logistička regresija i prostorno-vremenski modeli predstavljaju pokušaj poboljšanja klasičnih pristupa, no njihova prediktivna snaga i dalje ostaje ograničena u dinamičnim uvjetima [2].

Posljednjih godina sve je izraženiji pomak prema korištenju složenijih modela strojnog učenja za predikciju cijena nekretnina. Algoritmi poput slučajne šume, XGBoosta, SVM-a, kao i neuronskih mreža, pokazali su se znatno uspješnijima u modeliranju nelinearnih odnosa i obradi većih količina podataka [2][3]. Za razliku od tradicionalnih metoda, ovi algoritmi uče obrasce iz podataka bez potrebe za unaprijed definiranom funkcionalnom formom, što povećava fleksibilnost modela i njegovu sposobnost generalizacije.

Primjena ovih metoda dokazana je i u više konkretnih istraživanja. Baldominos i sur. (2018) analizirali su mogućnosti strojnog učenja u detekciji investicijskih prilika na tržištu

nekretnina koristeći skupove stvarnih podataka i kombinaciju različitih modela [1]. Jha i sur. (2020) usporedili su performance više regresijskih algoritama te zaključili kako modeli poput slučajne šume i XGBoosta pokazuju najbolje rezultate u pogledu točnosti [2]. Slično tome, istraživanja provedena 2024. godine pokazuju da suvremeni pristupi umjetne inteligencije u kombinaciji s bogatim skupovima značajki (npr. broj katova, udaljenost od gradskog centra, tip objekta) postižu značajna poboljšanja u odnosu na klasične metode [3][4].

Zbog velike financijske važnosti tržišta nekretnina i sve veće dostupnosti digitalnih izvora podataka (*online* oglasnici, geoprostorni podaci, cjenovni indeksi), predikcija cijena postaje ključna ne samo za pojedince i tvrtke, već i za državne institucije u planiranju i regulaciji stambene politike [2][4].

## 2. Regresijske metode strojnog učenja

### 2.1. Linearna regresija

Linearna regresija jedna je od osnovnih i najraširenijih metoda za modeliranje odnosa između ciljane varijable i jednog ili više prediktora. Može se koristiti kako bi se kvantificirao utjecaj karakteristika nekretnine, poput površine, lokacije ili broja soba, na njezinu tržišnu cijenu [5].

U najjednostavnijem slučaju, kada postoji samo jedan ulazni parametar  $x$ , linearna regresija modelira odnos s ciljnim izlazom  $y$  pomoću sljedeće funkcije:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (1)$$

Primjer takvog modela s prikazom skupa podataka i prilagođene regresijske linije dan je na slici 1. Za više varijabli, koristi se višestruka linearna regresija:

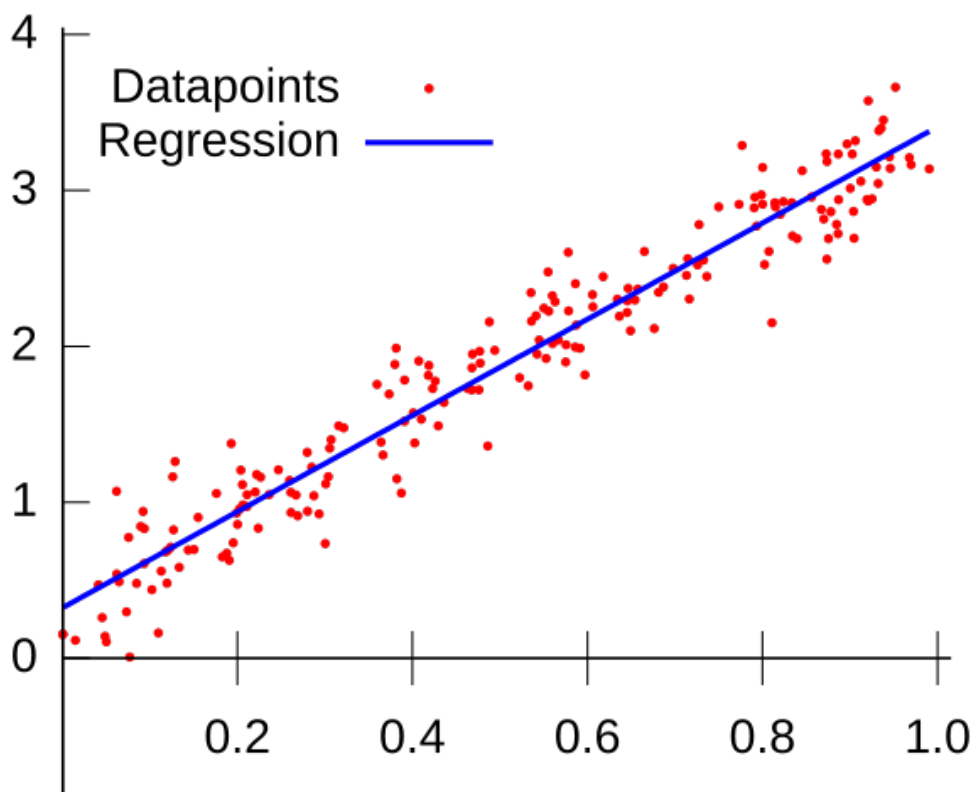
$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon \quad (2)$$

gdje je  $y$  procijenjena cijena,  $x_i$  su ulazne značajke (npr. kvadratura, godina izgradnje),  $\beta_i$  regresijski koeficijenti, a  $\varepsilon$  je pogreška modela. Koeficijenti se procjenjuju minimiziranjem funkcije pogreške, najčešće metodom najmanjih kvadrata (engl. *Ordinary Least Squares*) [5].

Zbog svoje jednostavnosti, linearna regresija često se koristi kao početna točka u izgradnji prediktivnih modela. Njena interpretabilnost je velika prednost jer je moguće jasno vidjeti kako i u kojoj mjeri svaka značajka utječe na konačnu cijenu [1][2]. Budući da je brza za implementirati i ima nisku računalnu složenost, vrlo je praktična, posebno kad se radi s manjim skupovima podataka [5].

Međutim, taj model ima niz ograničenja. Ključna pretpostavka je da su odnosi među varijablama linearni, što često nije slučaj na stvarnim tržištima. Osim toga, prisutnost višestruke kolinearnosti među značajkama može dovesti do nestabilnih koeficijenata i smanjene pouzdanosti modela [2]. Model je također osjetljiv na stršeće vrijednosti (engl. *outlier*) i ne uspijeva dobro opisati kompleksne nelinearne obrasce koji su česti u modeliranju cijena nekretnina [4].

Slika 1. Linearna regresija [Izvor: Berland, Wikimedia Commons, 2007.]



## 2.2. Gradijentno poboljšavanje modela

Gradijentno poboljšavanje modela (engl. *Gradient Boosting*) ubraja se među najpreciznije i najsnažnije algoritme za regresijske probleme u strojnom učenju. Ova metoda pripada skupini algoritama ansambala i temelji se na ideju da se više jednostavnih modela, najčešće stabala odluke, kombinira kako bi se izgradio prediktivni model. Za razliku od paralelnih metoda poput slučajnih šuma, modeli koji koriste poboljšavanje gradijenta uče se sekvencijalno pri čemu svaki novi model pokušava ispraviti pogreške prethodnog [6].

U ovom radu korišten je *GradientBoostingRegressor* iz Python biblioteke *scikit-learn*. Ova implementacija koristi stohastično gradijentno spuštanje kako bi iterativno minimizirala zadanu funkciju gubitka, najčešće srednju kvadratnu pogrešku (MSE) [9]. U svakoj iteraciji, model uči na rezidualima, odnosno razlikama između stvarnih i predviđenih vrijednosti. Time se postupno smanjuje ukupna pogreška modela [8].

Model se ažurira sljedećom formulom:

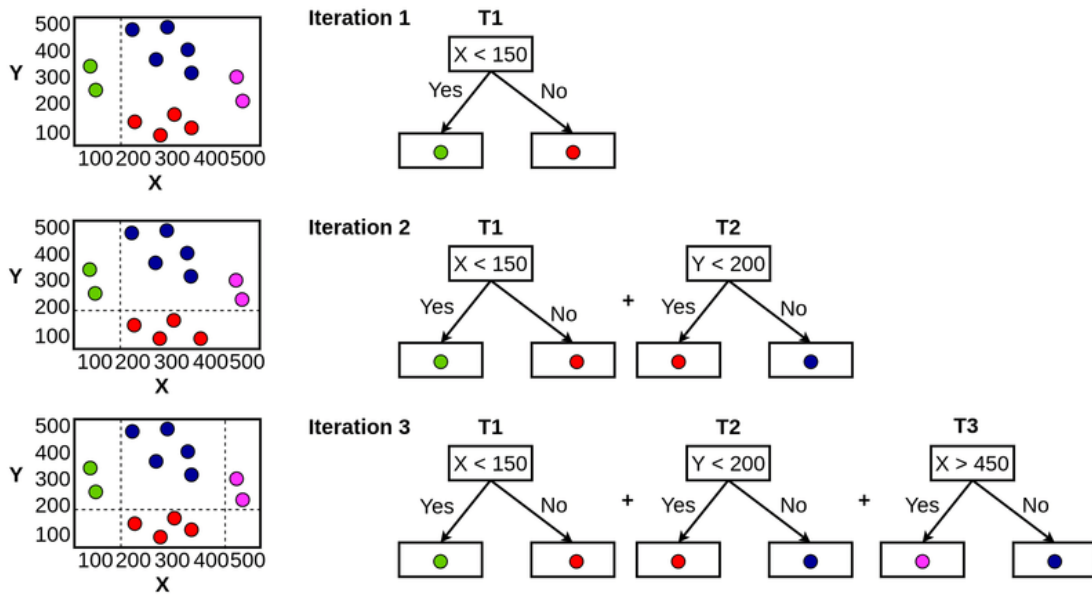
$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \quad (3)$$

gdje  $F_m(x)$  predstavlja novu, poboljšanu predikciju,  $h_m(x)$  je novo stablo koje uči na pogreškama, a  $\gamma_m$  je stopa učenja (engl. *learning rate*) koja određuje doprinos svakog stabla [6]. Slika 2 prikazuje stvaranje novih stabala iz početnog skupa podataka.

Algoritam gradijentnog poboljšavanja modela posebno se ističe u predviđanju cijena nekretnina zbog svoje sposobnosti modeliranja nelinearnih odnosa među varijablama. Za razliku od klasičnih modela poput linearne regresije ili stroja s potpornim vektorima, koji zahtijevaju eksplicitno definiranje odnosa među značajkama, ovaj model automatski uvažava međusobnu povezanost i interakciju ulaznih varijabli bez potrebe za ručnim definiranjem istih. Ova karakteristika daje algoritmu gradijentnog poboljšavanja modela značajnu prednost pri modeliranju cijena nekretnina.

Unatoč brojnim prednostima, taj algoritam ima i određene nedostatke. Jedno od glavnih ograničenja je njegova osjetljivost na prenaučenosť, osobito kad se koristi velik broj stabala bez odgovarajuće regulacije (npr. ograničenja dubine stabala, manji korak učenja). Također, u usporedbi s jednostavnijim modelima, učenje ovog modela može biti računalno zahtjevnije što može predstavljati izazov kod vrlo velikih skupova podataka [7].

Slika 2. Uzdzivanje gradijenta za regresiju [Izvor: Chetri et al, A Combined Metrics Approach to Cloud Service Reliability using Artificial Intelligence, 2021.]



## 2.3. Stroj s potpornim vektorima

Stroj s potpornim vektorima (SVM) najčešće se koristi za klasifikacijske probleme, no njegova varijanta za regresiju, *Support Vector Regression* (SVR), pokazala se vrlo učinkovitom u rješavanju problema predviđanja kontinuiranih vrijednosti, poput cijena nekretnina.

Za razliku od linearne regresije, koja pokušava minimizirati ukupnu pogrešku između predviđenih i stvarnih vrijednosti, SVR koristi drugačiji pristup. Umjesto da "kažnjava" svaku pogrešku, SVR definira granicu unutar koje se pogreške toleriraju i zanemaruju. Ta se granica označava simbolom  $\varepsilon$  (epsilon), a cilj modela je pronaći funkciju koja odstupa od stvarnih vrijednosti za najviše  $\varepsilon$ , uz što manji broj iznimki (točaka koje su izvan tog toleriranog područja) [10].

U osnovnom slučaju, SVR pokušava naći funkciju oblika

$$f = w^T x + b \quad (4)$$

gdje je  $w$  vektor težina (koeficijenata),  $x$  vektor ulaznih značajki, a  $b$  slobodni član (*bias*).

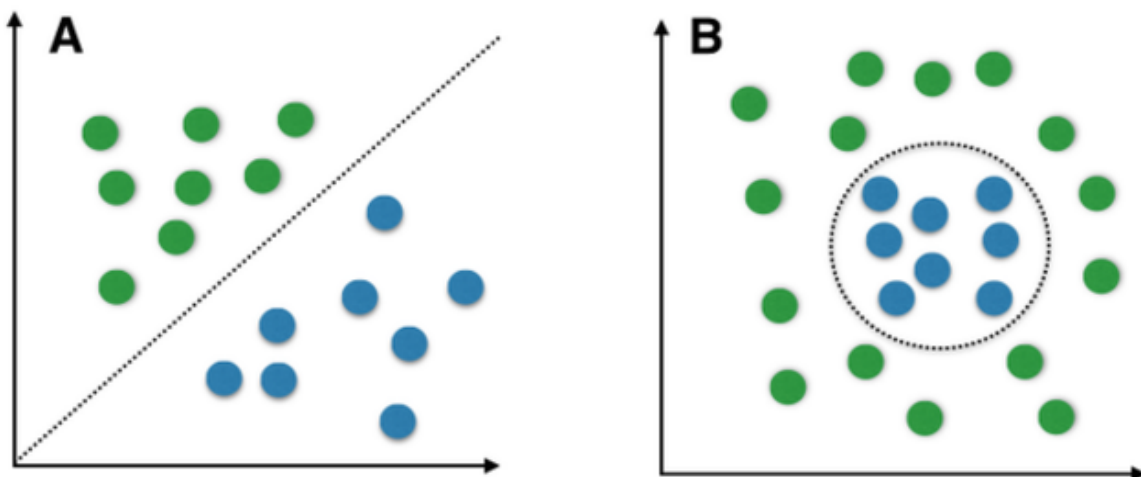
Ako odnos između značajki i ciljne vrijednosti nije linearan, što je često slučaj kod cijena nekretnina, SVR koristi funkciju jezgre (engl. *kernel function*) koja transformira podatke u višu dimenziju gdje se može lakše pronaći odgovarajuća linearna granica. Slika 3 prikazuje razdvajanje podataka odgovarajućom granicom.

Najčešće korištene funkcije jezgre su:

- Linearna jezgra – kada je odnos među varijablama već linearan
- Polinomna jezgra – za složenije, ali još uvijek donekle strukturirane odnose
- Jezgra s radijalnom jezgrom (engl. *Radial Basis Function*, RBF) – prikladna za situacije kada odnosi među podacima nisu očiti ni linearni, jer omogućuje modelu da sam prepozna obrasce bez dodatnog oblikovanja ulaznih značajki [11].

Jedna od važnijih prednosti SVR-a je to što dobro podnosi neuobičajene ili stršće vrijednosti u podacima. Za razliku od mnogih drugih modela koji mogu biti osjetljivi na takve vrijednosti, SVR koristi tzv.  $\epsilon$ -insensitive loss funkciju što znači da manje pogreške jednostavno ignorira. Model se fokusira samo na veća odstupanja, pa se tako smanjuje utjecaj šuma i nepotrebnih varijacija u podacima.

Slika 3. Stroj s potpornim vektorima [Izvor: Sebastian Raschka, Wikimedia Commons, 2017.]



## 2.4. Slučajne šume

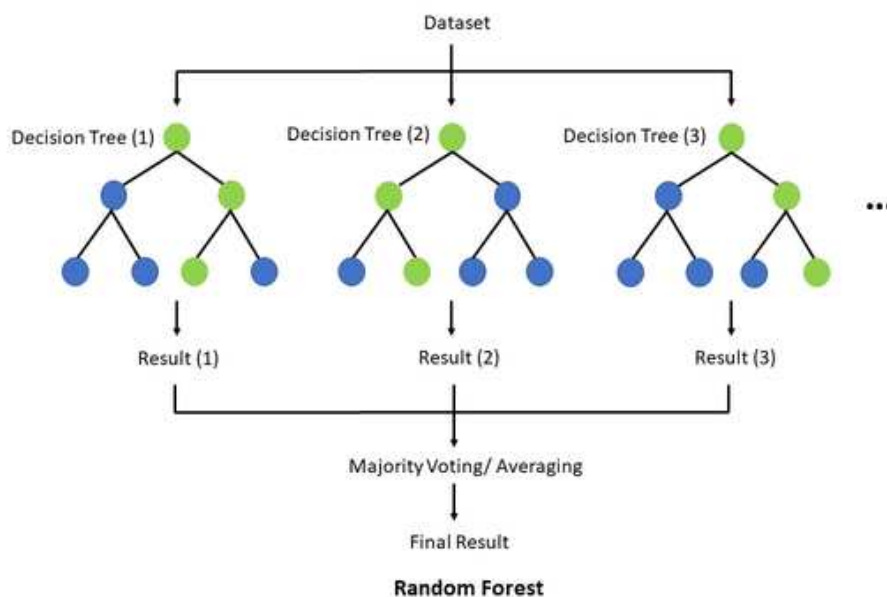
Algoritam slučajne šume (engl. *random forest*) popularan je algoritam ansambla koji kombinira više stabala odluke kako bi poboljšao točnost predikcije i smanjio rizik od prenaučenosti. Svako stablo uči se na drugačijem podskupu podataka, a prilikom svakog dijeljenja čvora koristi se slučajan podskup značajki, čime se dodatno povećava varijabilnost modela i smanjuje međusobna korelacija među stablima.

Predikcija modela dobije se agregacijom rezultata svih stabala, najčešće kao prosjek njihovih predikcija. Ovo smanjuje varijancu modela u odnosu na pojedinačno stablo i daje stabilnije i pouzdanije rezultate [12]. Slika 4 ilustrira na koji način model donosi odluku.

Algoritam slučajne šume ima mnogo prednosti u primjeni na predviđanje cijena nekretnina. Osim robusnosti i točnosti, sposoban je prepoznati nelinearne odnose i kompleksne interakcije među značajkama bez potrebe za prethodnim oblikovanjem podataka. Dodatno, on omogućava analizu važnosti značajki pa se može uočiti koji faktori (npr. kvadratura, lokacija, energetska razred) najviše utječu na cijenu.

Jedan od nedostataka ovog algoritma je povećana računalna složenost u odnosu na pojedinačna stabla te praktična nemogućnost interpretacije konačne odluke, jer se model može sastojati i od nekoliko tisuća pojedinačnih stabala, čime gubi intuitivnu preglednost koju nudi jedno stablo. [12]

Slika 4. Slučajne šume [Izvor: TseKiChun, Wikimedia Commons, 2021.]



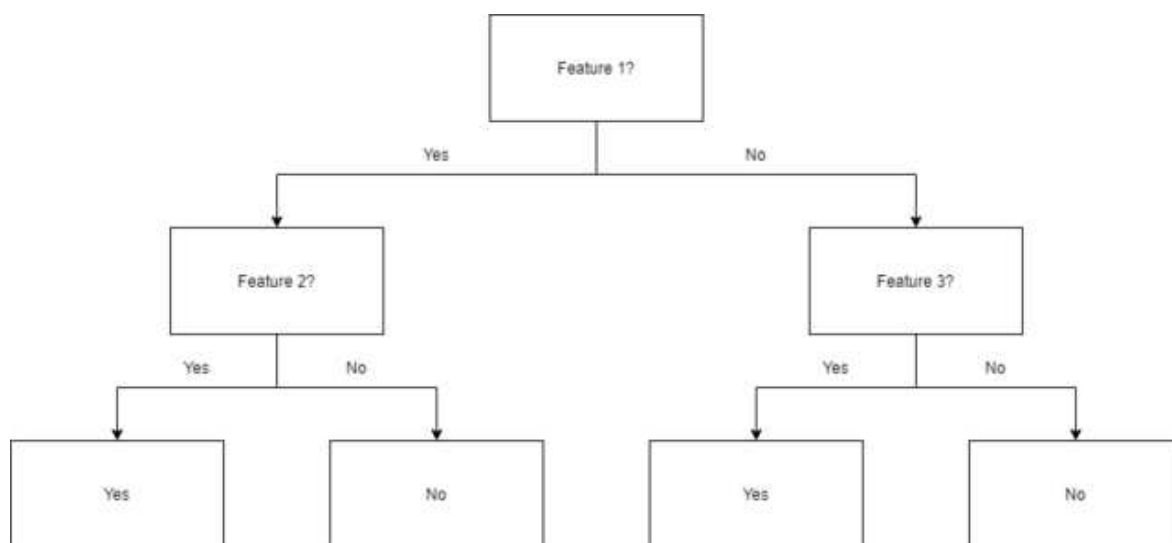
## 2.5. Stablo odluke

Stablo odluke (engl. *Decision Tree*) jedan je od najintuitivnijih algoritama strojnog učenja, a koristi se i za klasifikaciju i za regresiju. Kod regresijskih problema, poput predviđanja cijena nekretnina, stablo odluke segmentira prostor značajki u sve manje podskupove s pomoću sekvencijalnih binarnih pravila čime se formira struktura koja nalikuje stablu. Svaki unutarnji čvor predstavlja odluku temeljenu na nekoj značajki, a listovi stabla sadrže predikciju, u ovom slučaju procijenjenu cijenu nekretnine [13].

Prednost ovog algoritma leži u njegovoj interpretabilnosti jer je moguće jasno pratiti tijek odluka koje su dovele do određene procjene. Na primjer, stablo može prvo podijeliti podatke na temelju lokacije, zatim površine, a potom i starosti objekta, pružajući razumljiv niz pravila koji objašnjava cijenu pojedine nekretnine. Na slici 5 prikazano je dijeljenje podataka u manje skupove temeljem odabranih karakteristika.

Iako su stabla odluke algoritam koji je relativno jednostavan za razumijevanje, postoji mogućnost prenaučivosti, osobito kada su stabla duboka i kompleksna. Mala promjena u podacima može značajno promijeniti strukturu stabla, što naravno smanjuje njegovu robusnost. Zbog toga se algoritam stabla odluke rijetko koristi samostalno, već najčešće kao dio snažnijih metoda poput slučajne šume ili gradijentnog poboljšavanja modela za regresiju.

Slika 5. Stablo odluke [Izvor: CollaborativeGeneticist, Wikimedia Commons, 2021.]



### 3. Podaci i obrada

U ovom radu korištena su dva glavna izvora podataka: službeni podaci o indeksima cijena nekretnina preuzeti s web stranice Državnog zavoda za statistiku (DZS) te vlastito prikupljeni podaci s oglasnika Njuškalo.hr putem metode web *scrapinga* pomoću gotovog rješenja preuzetog s GitHub-a. [19].

#### 3.1. Indeksi cijena nekretnina

Indeksi cijena stambenih nekretnina preuzeti su s mrežnih stranica Državnog zavoda za statistiku (DZS) u obliku kvartalnih vremenskih serija. Služe kao službeni pokazatelji promjena u cijenama nekretnina kroz dulje razdoblje, pri čemu se izračun temelji na hedonističkoj regresijskoj metodi odnosno statističkom pristupu kojim se cijene prilagođavaju za utjecaj karakteristika nekretnina poput veličine, lokacije i tipa nekretnine.

Indeksi su dostupni za cjelokupno tržište, Grad Zagreb, Jadran (obalne županije) i ostatak Hrvatske (kontinentalni dio). Indeksna vrijednost postavljena je na 100 za referentno razdoblje, odnosno prvi kvartal 2015. godine, dok se svaka sljedeća vrijednost izražava relativno prema toj referentnoj vrijednosti.

Podaci su učitani iz Excel datoteke u kojoj su reci sadržavali kombinaciju godine i kvartala (npr. „2019 Q1“), a stupci su predstavljali različita zemljopisna područja: Ukupno, Grad Zagreb, Jadran i Ostalo. Ovaj format omogućio je pretvaranje u vremensku seriju korištenjem biblioteke Pandas.

Za potrebe vremenskog modeliranja generirani su atributi s pomakom (tzv. lagirani atributi) koji predstavljaju vrijednosti indeksa iz prethodnih kvartala. Ovaj pristup omogućio je primjenu regresijskih modela temeljenih na vremenskom slijedu, uključujući validaciju *walk-forward*.

## 3.2. Tržišni podaci

U ovom radu korišteni su podaci prikupljeni automatskim *scrapingom* oglasa s oglasnika Njuškalo.hr. Prikupljeni su oglasi isključivo za prodaju nekretnina, i to za dvije osnovne kategorije: stanovi i kuće. Nakon toga, unutar svake kategorije, zapisi su razvrstani prema županiji iz koje dolaze. Rezultat svakog ciklusa *scrapinga* bio je JSON formatirani zapis, gdje je svaki zapis predstavljao jedan oglas.

Svaka županija predstavljena je dvjema JSON datotekama koje sadržavaju informacije o svim kućama i stanovima trenutno dostupnim za prodaju. Za svaki oglas dohvaćeni su podaci o županiji, vrsti nekretnine, kvadraturi, cijeni i tekstualnom opisu nekretnine.

*Scrapeanje* je provedeno u dva dijela, krajem prvog kvartala 2025. godine te krajem drugog kvartala 2025. Ovaj pristup omogućio je prikupljanje trenutnih traženih tržišnih cijena nekretnina u ta dva razdoblja, što je bilo ključno za usporedbu s predviđenim indeksima nekretnina za te periode dobivenim regresijskim metodama. Usporedba ovih vrijednosti omogućila je procjenu usklađenosti tržišnih očekivanja s predviđenim trendovima kretanja cijena.

Kod učitavanja podataka implementirana je funkcija koja osigurava pravilno dekodiranje UTF-8 znakova, obradu grešaka i konverziju u Pandas DataFrame. Nakon učitavanja podataka kreiran je po jedan rječnik za svako razdoblje *scrapeanja* (Q1 i Q2), `dataframes_03` i `dataframes_06`, u kojem svaki ključ predstavlja kombinaciju županije i tipa nekretnine (npr. `grad-zagreb_stanovi`).

Budući da su pojedini oglasi imali iskazanu cijenu po kvadratnom metru, a pojedini ukupnu cijenu, u takvim slučajevima ukupna cijena izračunata je kao cijena po kvadratu pomnožena s kvadraturom navedene nekretnine. Tako su svi oglasi dovedeni na usporedivu osnovu s apsolutnim cijenama, što je omogućilo njihovu daljnju analizu i modeliranje.

Podaci su dodatno očišćeni, uklonjeni su oglasi s nedostajućim ili nelogičnim podacima (npr. bez cijene ili bez kvadrature), a duplikati su eliminirani.

**Tablica 1. – Broj podataka nakon čišćenja**

| Regija | Broj oglasa – Q1 2025 | Broj oglasa – Q2 2025 |
|--------|-----------------------|-----------------------|
| Ukupno | 92098                 | 89312                 |
| Zagreb | 13829                 | 10066                 |
| Obala  | 68952                 | 69875                 |
| Ostalo | 9317                  | 9371                  |

Treba naglasiti da se radi o traženim cijenama iz oglasa, koje ne moraju nužno odgovarati realiziranim tržišnim cijenama. Osim toga, struktura i broj oglasa na Njuškalu su promjenjivi i ovise o sezoni i trenutačnoj ponudi na tržištu, pa prikupljeni podaci predstavljaju stanje u trenutku *scrapinga*.

**Tablica 2. Podjela regija**

| Regija | Županije                                                                                                                                                                                                                                               |
|--------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Ukupno | Sve županije                                                                                                                                                                                                                                           |
| Zagreb | Grad Zagreb                                                                                                                                                                                                                                            |
| Obala  | Primorsko-goranska, Ličko-senjska, Zadarska, Šibensko-kninska, Splitsko-dalmatinska, Istarska, Dubrovačko-neretvanska                                                                                                                                  |
| Ostalo | Zagrebačka, Krapinsko-zagorska, Sisačko-moslavačka, Karlovačka, Varaždinska, Koprivničko-križevačka, Bjelovarsko-bilogorska, Virovitičko-podravka, Požeško-slavonska, Brodsko-posavska, Osječko-baranjska, Vukovarsko-srijemska i Međimurska županija. |

## 4. Izgradnja i vrednovanje modela

### 4.1. Priprema podataka za modeliranje

Nakon što su podaci prikupljeni i očišćeni, bilo ih je potrebno pripremiti za modeliranje. Budući da se radi o vremenskim nizovima, potrebno je osigurati da modeli u svakom trenutku koriste isključivo informacije o prošlosti, bez uvida u buduće vrijednosti.

Zbog toga je izrađen skup značajki temeljen na tzv. lagiranim vrijednostima, odnosno vrijednostima iz prethodnih kvartala. Za svaki kvartal formirano je osam takvih značajki (Lag1 do Lag8) koje predstavljaju indekse cijena nekretnina iz prethodnih osam kvartala.

Za učenje modela korištena je metoda *walk-forward*, koja je prikladna za vremenske nizove i omogućuje realističnu procjenu sposobnosti modela za predviđanje budućih podataka. Ovaj pristup podrazumijeva da se model u svakom trenutku uči na svim dostupnim podacima do tog trenutka, a zatim se koristi za predikciju vrijednosti za idući kvartal. Nakon što nova vrijednost postane poznata, uključuje se u skup za učenje i proces se ponavlja. Time se simulira način na koji bi se model koristio u stvarnom vremenu. Tako je osigurano da model ne koristi informacije iz budućnosti prilikom predikcije.

Također, podaci su standardizirani funkcijom `StandardScaler` kako bi se osiguralo da sve značajke budu na istoj skali. Time se izbjegava dominacija značajki s većim numeričkim vrijednostima.

### 4.2. Učenje modela i rezultati

Za svaki od regresijskih modela proveden je zaseban postupak učenja i vrednovanja, a primijenjeni su identični skupovi podataka i identična struktura značajki za sve modele.

Podaci su prethodno standardizirani, a svaki model dodatno je optimiran funkcijom `GridSearchCV` za odabir hiperparametara. Posebna pažnja posvećena je sprječavanju curenja informacija iz budućnosti u skup za učenje.

Rezultati ovakvog učenja modela pokazali su jasne razlike u točnosti predikcije, ovisnosti o složenosti metode i tržišnoj kategoriji. Napredniji algoritmi poput gradijentnog poboljšanja modela i slučajnih šuma pokazali su visoku točnost, dok su jednostavniji modeli, poput linearne regresije i stabla odluke očekivano pokazali slabiju točnost.

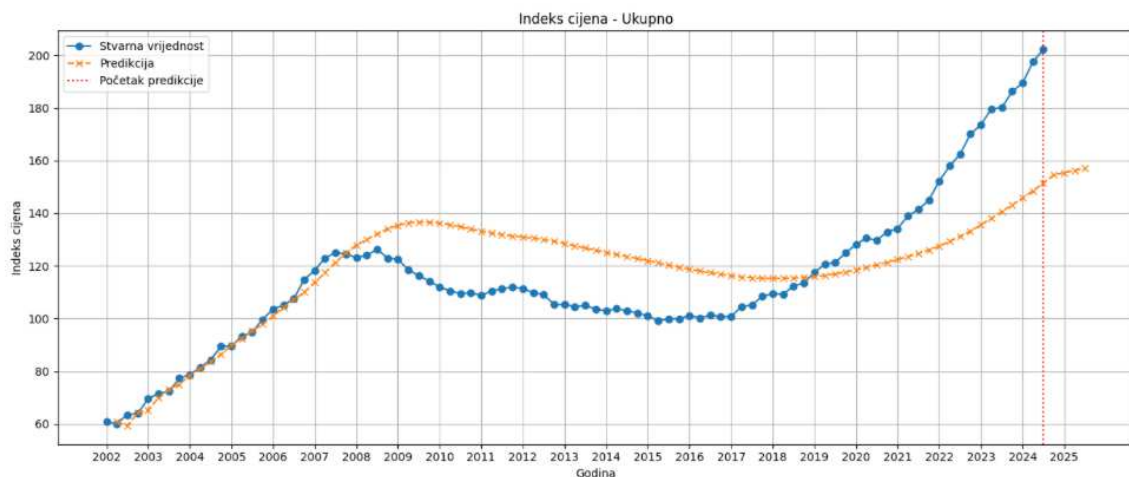
### 4.2.1. Izgradnja modela metodom linearne regresije

Linearni regresijski model je najjednostavniji oblik regresije i služi kao početna točka za usporedbu s naprednijim tehnikama. U ovom radu korištena je obična višestruka linearna regresija bez regularizacije, kako bi se dobio jasan uvid u osnovne linearnosti među kvartalnim vrijednostima.

Model je učen na podacima koji su standardizirani StandardScalerom, a cilj mu je pronaći najbolje koeficijente koji povezuju vrijednosti kvartalnih cijena iz prošlosti s budućom vrijednošću. Iako je linearna regresija vrlo ograničena u svojoj mogućnosti modeliranja nelinearnih odnosa, ona ipak omogućuje lako tumačenje rezultata i daje uvid u doprinos svake od ulaznih značajki.

Rezultati pokazuju solidne performanse, s  $R^2$  vrijednostima oko 0.55 što implicira da čak i najjednostavniji model linearne regresije može objasniti oko 55% varijance u podacima, dok su se MAE vrijednosti kretale oko 15. Slika 6 prikazuje da, iako linearna regresija uspješno prati trendove kretanja cijena nekretnina (pad ili rast), ona i dalje nije dovoljno precizna kako bi se korisno predvidjeli podaci za budućnost.

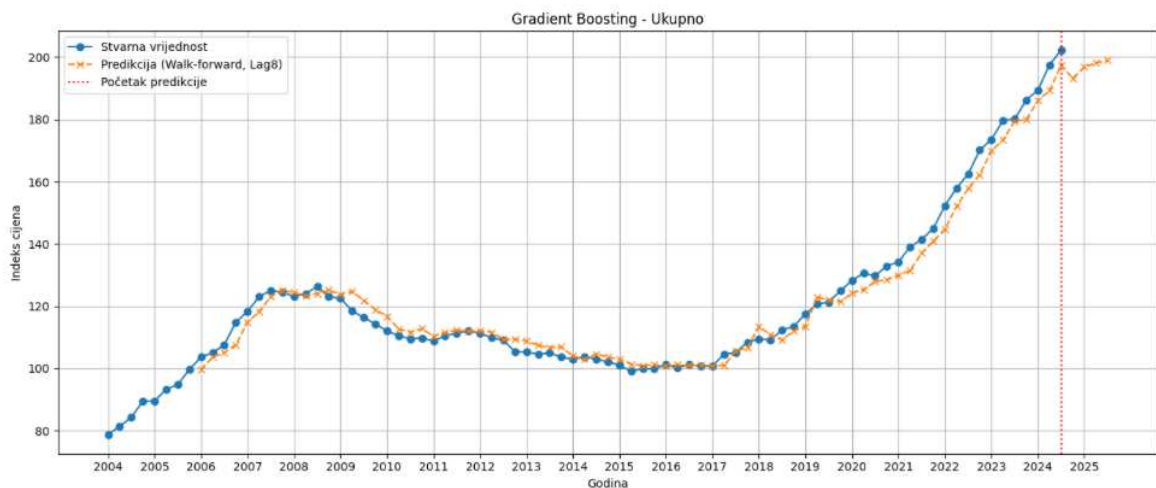
Slika 6. Procjena metodom linearne regresije



### 4.2.2. Izgradnja modela metodom gradijentnog poboljšanja

Gradijentno poboljšanje vrlo je sofisticirana metoda koja koristi više slabijih modela kako bi izgradila model s većom mogućnošću predviđanja. U ovom radu korišten je GradientBoostingRegressor s optimiranim parametrima broja estimatora, dubine stabla i koraka učenja. Model je učen na istom, standardiziranom, skupu podataka kao i ostali modeli, a prilikom validacije pokazao je najbolje performance.  $R^2$  vrijednosti bile su stabilno iznad 0.9 što upućuje na vrlo visoku sposobnost objašnjavanja varijance. MAE se prosječno kretao oko 3 što je među najmanjim vrijednostima od svih izgrađenih modela. Važno je istaknuti da model vrlo dobro reagira na sezonske promjene i oscilacije na tržištu, prikazano na slici 7, što je konzistentno s fleksibilnošću ove metode pri modeliranju složenih nelinearnih odnosa.

Slika 7. Procjena metodom gradijentnog poboljšanja modela



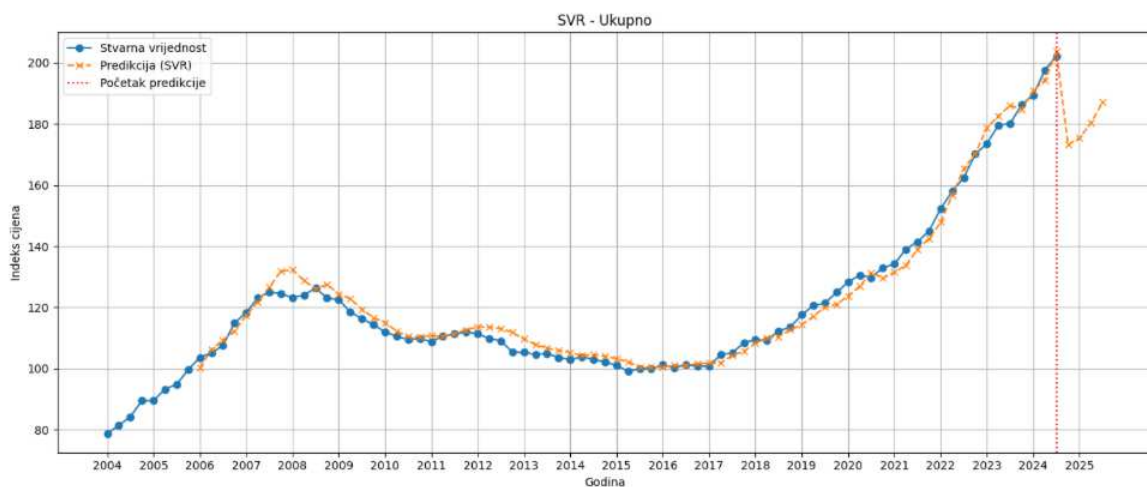
### 4.2.3. Izgradnja modela metodom stroja potpornih vektora

Pri izgradnji SVR modela ispitano je više različitih funkcija jezgre (npr. linearna, polinomijalna, RBF), a konačan odabir hiperparametara, poput  $C$ ,  $\epsilon$  i  $\gamma$ , za svaku kategoriju napravljen je na temelju rezultata GridSearchCV optimizacije.

Ova optimizacija provedena je zasebno za svaku kategoriju podataka (Ukupno, Obala, Zagreb i Ostalo), čime se model dodatno prilagodio specifičnostima pojedinog tržišta.

Nakon što su za svaku kategoriju pronađene najbolje kombinacije parametara, one su se koristile za učenje konačnih modela nad standardiziranim podacima. Rezultati pokazuju da ovaj algoritam daje konzistentne, iako malo lošije rezultate u odnosu na algoritam gradijentnog poboljšanja modela. Slika 8 prikazuje da model predviđa značajan pad cijena nekretnina počevši od prvog kvartala 2025. godine, što drugi modeli nisu predvidjeli. Također, iako su  $R^2$  vrijednosti su visoke, preko 0.95, MAE je nešto niži u odnosu na ostale algoritme.

Slika 8. Procjena metodom potpornih vektora



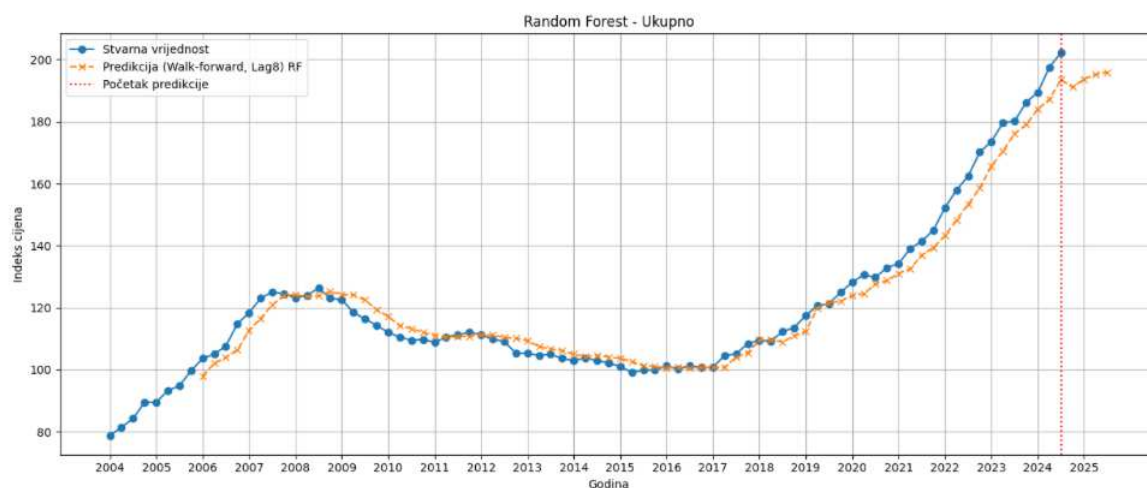
#### 4.2.4. Izgradnja modela metodom slučajne šume

Za učenje modela slučajnih šuma korišten je RandomForestRegressor, a optimizacija parametara kao što su broj stabala ( $n\_estimators$ ), maksimalna dubina i minimalni broj uzoraka po listu provedena je korištenjem GridSearchCV. Optimizacija je, kao i u ostalim modelima, provedena zasebno za svaku od četiri kategorije podataka (Ukupno, Zagreb, Obala i Ostalo), što omogućava najbolju prilagodbu modela tržištu.

Iako s nešto većim MAE i RMSE vrijednostima, i ovaj model ima vrlo visoke  $R^2$  vrijednosti, preko 0.95, što je izuzetno dobar rezultat i blizu modela s najboljim performansama. Na slici 9 vidljivo je da, iako model predvidi mali pad cijena nekretnina u drugom kvartalu 2025., on generalno dobro predviđa cijene uz male greške. U većini slučajeva najveću važnost imale su najrecentnije kvartalne vrijednosti, što je i očekivano. Model je nešto manje precizan u

predviđanju naglih promjena u cijenama nekretnina, no generalno se pokazao stabilnim i pouzdanim.

Slika 9. Procjena metodom slučajne šume



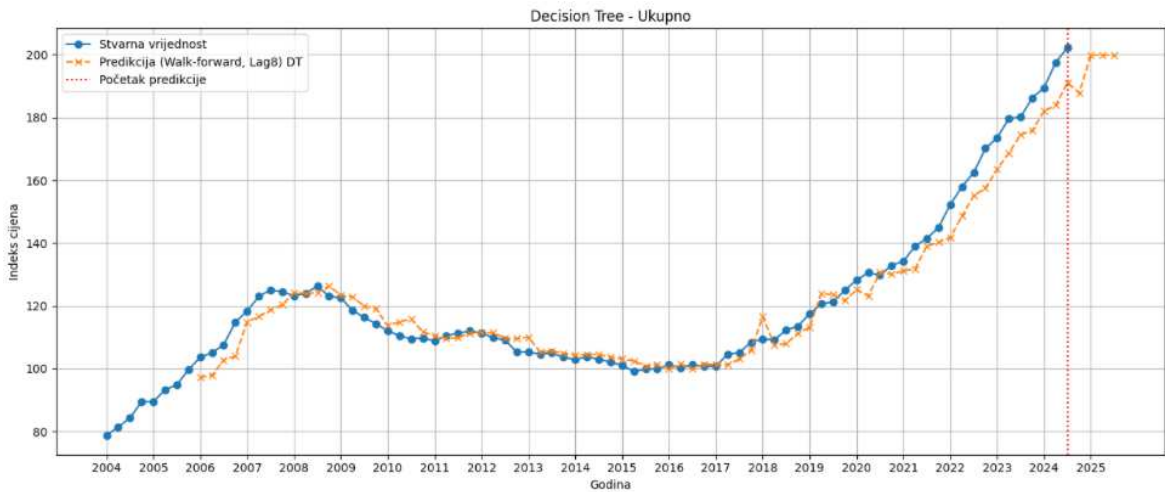
#### 4.2.5. Izgradnja modela metodom stabla odluke

Model stabla odluke učen je korištenjem `DecisionTreeRegressor`, a kao i kod svih prethodnih modela, optimizacija hiperparametara provedena je korištenjem funkcije `GridSearchCV`. Najvažniji parametri bili su maksimalna dubina stabla i minimalni broj uzoraka potrebnih za podjelu. Rezultati su pokazali da, iako stablo može jako dobro pratiti podatke na skupu za učenje, na testnim podacima dolazi do značajnog pada performanci.

Iako su  $R^2$  i MAE ukazivali na relativno dobar rezultat, prilikom predviđanja indeksa cijena nekretnina za iduća četiri kvartala uočeno je da su sve predviđene vrijednosti identične, prikazano na slici 10. Budući da stablo odlučuje na temelju podjela unutar vrijednosti značajki, očito se dogodilo da je u završnim čvorovima stabla završilo više uzastopnih ulaza koji imaju vrlo sličnu ili istu vrijednost ciljne varijable, pa je model svim novim ulazima dodijelio istu vrijednost.

Stablo odluke pokazalo se korisno za interpretaciju jer omogućuje jednostavno praćenje logike predikcije, ali zbog lošijih kvantitativnih pokazatelja nije prikladno za krajnju upotrebu u ovom kontekstu.

Slika 10. Procjena metodom stabla odluke



### 4.3. Vrednovanje modela

Nakon izgradnje svih odabranih modela i njihove promjene na vremenskim indeksima cijena nekretnina, provedeno je vrednovanje performanci kako bi se usporedila njihova točnost, robusnost i sposobnost generalizacije. Vrednovanje je temeljeno na trima osnovnim mjerama: srednjoj apsolutnoj pogrešci (MAE), srednjoj kvadratnoj pogrešci (MSE) i koeficijentu determinacije ( $R^2$ ).

Definicije korištenih mjera pogreške su sljedeće:

- Srednja apsolutna pogreška:  $MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$  (5)

- Srednja kvadratna pogreška:  $MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$  (6)

- Koeficijent determinacije:  $R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$  (7)

Gdje su  $y_i$  stvarne vrijednosti,  $\hat{y}_i$  predviđene vrijednosti,  $\bar{y}$  srednja vrijednost stvarnih podataka, a  $n$  broj uzoraka.

Rezultati vrednovanja modela prikazani su u tablici 3. Iz nje je vidljivo da su najbolji rezultati postignuti metodom stroja s potpornim vektorima, koji je ostvario najviši koeficijent determinacije ( $R^2 = 0.99$ ) te najniže pogreške MAE i RMSE na ukupnom skupu podataka. Vrlo dobre performance postigao je i model uzdizanja gradijenta za regresiju s koeficijentom determinacije  $R^2 = 0.98$ , no nešto višim pogreškama.

Promatrajući rezultate po regijama vidljivo je da modeli najbolje predviđaju cijene u regijama Ukupno i Obala, gdje su i pogreške najmanje, dok je za Grad Zagreb pogreška najveća kod svih modela. To može značiti da je u Zagrebu heterogenost cijena nekretnina veća nego na obali.

**Tablica 3. Pogreške modela**

|                                   |               | $R^2$       | MAE          | RMSE         |
|-----------------------------------|---------------|-------------|--------------|--------------|
| Linearna Regresija                | Ukupno        | 0,55        | 15,02        | 19,43        |
|                                   | Grad Zagreb   | 0,49        | 18,74        | 24,29        |
|                                   | <b>Obala</b>  | <b>0,64</b> | <b>13,38</b> | <b>17,17</b> |
|                                   | Ostalo        | 0,41        | 13,23        | 17,36        |
| Uzdizanje gradijenta za regresiju | <b>Ukupno</b> | <b>0,98</b> | <b>2,92</b>  | <b>3,59</b>  |
|                                   | Grad Zagreb   | 0,98        | 3,71         | 4,88         |
|                                   | Obala         | 0,97        | 3,11         | 3,79         |
|                                   | Ostalo        | 0,95        | 3,36         | 4,52         |
| Stroj s potpornim vektorima       | <b>Ukupno</b> | <b>0,99</b> | <b>2,44</b>  | <b>3,01</b>  |
|                                   | Grad Zagreb   | 0,98        | 3,88         | 4,97         |
|                                   | Obala         | 0,98        | 2,64         | 3,38         |
|                                   | Ostalo        | 0,96        | 3,15         | 4,29         |
| Slučajne šume                     | <b>Ukupno</b> | <b>0,97</b> | <b>3,58</b>  | <b>4,57</b>  |
|                                   | Grad Zagreb   | 0,97        | 4,31         | 5,62         |
|                                   | Obala         | 0,96        | 3,63         | 4,56         |
|                                   | Ostalo        | 0,94        | 3,69         | 5,17         |
| Stablo odluke                     | Ukupno        | 0,96        | 3,97         | 5,17         |
|                                   | Grad Zagreb   | 0,97        | 4,56         | 5,97         |
|                                   | <b>Obala</b>  | <b>0,97</b> | <b>3,39</b>  | <b>4,21</b>  |
|                                   | Ostalo        | 0,94        | 3,80         | 5,10         |

## 5. Normalizacija tržišnih cijena i povezivanje s indeksima

Nakon što su podaci s oglasnika Njuškalo.hr očišćeni i pripremljeni za daljnje rukovanje, agregirani su po kvartalima (Q1 2025 i Q2 2025) te po županijama.

Nakon ovakve obrade, pristupilo se uspoređivanju i stavljanju stvarnih cijena nekretnina u ispravan odnos s indeksima cijena nekretnina koje objavljuje Državni zavod za statistiku (DZS). Cilj ove manipulacije bio je omogućiti normalizaciju i međusobnu usporedivost tržišnih cijena oglašanih nekretnina s kretanjem službenih indeksa, kao i pretvorbu relativnih vrijednosti u apsolutne promjene.

S obzirom na to da su službeni indeksi izraženi u odnosu na referentno razdoblje prvog kvartala 2015. godine, kojem je pridružena vrijednost 100, svi kvartalni medijani ukupnih cijena iz podataka s Njuškala podijeljeni su s vrijednošću indeksa (za pripadajući kvartal i regiju), koja je prethodno normirana dijeljenjem sa 100. Tako je dobivena modificirana cijena nekretnine koja bi odgovarala cijeni takve ili slične nekretnine u referentnom razdoblju, odnosno u prvom kvartalu 2015. Drugim riječima, računanjem:

$$\text{Normirani medijan} = \frac{\text{medijan ukupnih cijena nekretnina}}{\frac{\text{indeks cijena nekretnina}}{100}} \quad (8)$$

dobiven je standardizirani medijan koji je moguće koristiti za rekonstrukciju tržišnih cijena nekretnina u bilo kojem razdoblju, povijesnom ili budućem. Pomnoži li se dobivena vrijednost normiranog medijana s nekim drugim kvartalnim indeksom, dobije se procjena cijene te nekretnine u željenom kvartalu.

Primjerice, u prvom kvartalu 2025. godine medijan ukupnih traženih cijena nekretnina na Njuškalu iznosi 314.753,00 eura. Istodobno, procjena indeksa cijena nekretnina za taj period u regiji Ukupno iznosi 193,19, uzimajući u obzir predviđanje metodom uzdizanja gradijenta. Računamo:

$$\text{Normirani medijan} = \frac{314.753,00}{\frac{193,19}{100}} \approx 162.924,00$$

Ovaj normirani medijan tumači se kao „osnovna“ vrijednost takve nekretnine u prvom kvartalu 2015. Ako želimo procijeniti koja bi bila cijena iste nekretnine u npr. u četvrtom kvartalu 2007. godine kada je indeks cijena nekretnina iznosio 134,48, računamo:

$$\text{Procijenjena cijena u Q4 2007} = 162.924,00 * \frac{134,48}{100} \approx 219.100,00$$

Na ovaj način može se dvosmjerno interpretirati cijene nekretnina: kako bi se kretale u prošlosti pod današnjim uvjetima te kako će se kretati u budućnosti.

## 5.1. Predviđanje cijena nekretnina u drugoj polovini 2025.

U ovom dijelu rada pokušava se doći do procjene kako bi se cijene nekretnina mogle kretati u drugoj polovini 2025. godine. Koristit će se predikcije modela gradijentnog poboljšanja koji se pokazao vrlo pouzdan.

U tablici 4. prikazane su medijalne cijene nekretnina po regijama i kvartalima za ovu godinu. Prva dva kvartala temelje se na stvarno prikupljenim podacima, dok su za treći i četvrti kvartal navedene vrijednosti koje predviđa model.

**Tablica 4. Medijalne cijene nekretnina (u eur) po regijama i kvartalima**

| Regija | Q1 2025    | Q2 2025    | Q3 2025    | Q4 2025    |
|--------|------------|------------|------------|------------|
| Ukupno | 314.753,25 | 325.000,00 | 327.097,00 | 328.467,47 |
| Zagreb | 271.176,30 | 290.400,00 | 293.288,10 | 295.541,85 |
| Obala  | 355.000,00 | 360.000,00 | 361.619,87 | 366.104,90 |
| Ostalo | 157.600,00 | 160.000,00 | 160.892,10 | 164.174,85 |

# Zaključak

U ovom radu istražena je primjena regresijskih modela strojnog učenja u svrhu predviđanja cijena nekretnina na području Hrvatske, s ciljem postizanja što veće preciznosti i robusnosti modela. Izgrađeno je i vrednovano više modela, a to su linearna regresija, gradijentno poboljšanje modela, stroj s potpornim vektorima, slučajna šuma i stablo odluke. Svaki model naučen je i testiran na indeksima cijena nekretnina preuzetih od Državnog zavoda za statistiku (DZS).

Rezultati pokazuju da jednostavni modeli poput linearne regresije i stabla odluke za regresiju pružaju očekivano ograničene mogućnosti predviđanja, posebno zbog nelinearnih odnosa značajki i sezonskih oscilacija tržišta. S druge strane, naprednije metode poput gradijentnog poboljšanja modela i stroja s potpornim vektorima za regresiju pokazale su visoku i konzistentnu točnost, s  $R^2$  vrijednostima većim od 0.95 i niskim srednjim apsolutnim pogreškama.

Jedan od ključnih trenutaka u radu je povezivanje povijesnih i budućih, predviđenih, indeksa cijena nekretnina sa stvarnim tržišnim podacima, što je omogućilo normalizaciju cijena i njihovu usporedbu kroz različite vremenske periode. Nakon normalizacije moguće je rekonstruirati cijenu nekretnine u bilo kojem ranijem periodu, kao i predvidjeti njezine oscilacije kroz narednu godinu. Ovakav pristup mogao bi biti dobra podloga za izgradnju složenijih prediktivnih sustava koji bi služili kao alat stručnjacima za nekretnine te institucijama zaduženim za stambenu politiku.

Konačno, rezultati potvrđuju kako su složeniji regresijski modeli strojnog učenja puno točniji i robusniji od tradicionalnih regresijskih pristupa kada se radi o procjeni cijena nekretnina. Iako implementacija takvih modela iziskuje više truda i resursa u vidu obrade podataka, optimizacije i brzine izvođenja, velika prednost u točnosti opravdava njihovu primjenu.

Za buduća istraživanja preporučuje se proširenje modela uključivanjem dodatnih značajki poput detaljnijih geoprostornih podataka, podataka o prometnoj povezanosti, energetske certifikatima i slično. Uvođenje stvarnih transakcijskih cijena, umjesto oglašanih, moglo bi dodatno povećati preciznost predikcija.

# Literatura

- [1] Baldominos, A., Moreno, A. J., Iturrarte, R., Bernardez, O., Afonso, C. *Identifying Real Estate Opportunities Using Machine Learning*, 2018., ResearchGate, [https://www.researchgate.net/profile/Alejandro-Baldominos/publication/327644271\\_Identifying\\_Real\\_Estate\\_Opportunities\\_Using\\_Machine\\_Learning/links/5be1507aa6fdcc3a8dc18662/Identifying-Real-Estate-Opportunities-Using-Machine-Learning.pdf?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0](https://www.researchgate.net/profile/Alejandro-Baldominos/publication/327644271_Identifying_Real_Estate_Opportunities_Using_Machine_Learning/links/5be1507aa6fdcc3a8dc18662/Identifying-Real-Estate-Opportunities-Using-Machine-Learning.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0), (pristupljeno 10. travnja 2025.)
- [2] Jha, S. B., Pandey, V., Jha, R. K., Babiceanu, R. F. *Machine Learning Approaches to Real Estate Market Prediction Problem: A Case Study*, 2020., ResearchGate, [https://www.researchgate.net/publication/343849620\\_Machine\\_Learning\\_Approaches\\_to\\_Real\\_Estate\\_Market\\_Prediction\\_Problem\\_A\\_Case\\_Study/fulltext/5f44864ba6fdcccc43fc6dc8/Machine-Learning-Approaches-to-Real-Estate-Market-Prediction-Problem-A-Case-Study.pdf?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0](https://www.researchgate.net/publication/343849620_Machine_Learning_Approaches_to_Real_Estate_Market_Prediction_Problem_A_Case_Study/fulltext/5f44864ba6fdcccc43fc6dc8/Machine-Learning-Approaches-to-Real-Estate-Market-Prediction-Problem-A-Case-Study.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0), (pristupljeno 16. travnja 2025.)
- [3] Reddy, P. C., Sanjai, I., Abishek, V., Tripathy, B. K. *Explainable AI and Machine Learning in Prediction of Housing Prices*, 2024., ResearchGate, [https://www.researchgate.net/profile/Bk-Tripathy/publication/380621139\\_Explainable\\_AI\\_and\\_Machine\\_Learning\\_in\\_Prediction\\_of\\_Housing\\_Prices/links/66694374de777205a325aa58/Explainable-AI-and-Machine-Learning-in-Prediction-of-Housing-Prices.pdf?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0](https://www.researchgate.net/profile/Bk-Tripathy/publication/380621139_Explainable_AI_and_Machine_Learning_in_Prediction_of_Housing_Prices/links/66694374de777205a325aa58/Explainable-AI-and-Machine-Learning-in-Prediction-of-Housing-Prices.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0), (pristupljeno 23. travnja 2025.)
- [4] Maloku, F., Maloku, B. *House Price Prediction Using Machine Learning and Artificial Intelligence*, 2024., Journal of Artificial Intelligence & Cloud Computing, [https://www.researchgate.net/profile/Fatbardha-Maloku/publication/383057414\\_House\\_Price\\_Prediction\\_Using\\_Machine\\_Learning\\_and\\_Artificial\\_Intelligence/links/66ba4a69299c327096c3e221/House-Price-Prediction-Using-Machine-Learning-and-Artificial-Intelligence.pdf?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0](https://www.researchgate.net/profile/Fatbardha-Maloku/publication/383057414_House_Price_Prediction_Using_Machine_Learning_and_Artificial_Intelligence/links/66ba4a69299c327096c3e221/House-Price-Prediction-Using-Machine-Learning-and-Artificial-Intelligence.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6Ii9kaXJlY3QiLCJwYWdlIjoicHVibGljYXRpb24ifX0), (pristupljeno 30. travnja 2025.)
- [5] GeeksforGeeks. *Linear Regression i Machine learning*, 2025., <https://www.geeksforgeeks.org/machine-learning/ml-linear-regression/>, (pristupljeno 8. svibnja 2025.)
- [6] Baladram, S. *Gradient Boosting Regressor Explained: A Visual Guide with Code Examples*, 2024., Towards Data Science, <https://towardsdatascience.com/gradient-boosting-regressor-explained-a-visual-guide-with-code-examples-c098d1ae425c/>, (pristupljeno 15. svibnja 2025.)
- [7] Masui, T. *All You Need to Know About Gradient Boosting Algorithm – Part 1: Regression*, 2022., Medium, <https://medium.com/data-science/all-you-need-to-know->

[about-gradient-boosting-algorithm-part-1-regression-2520a34a502](#),  
(pristupljeno 15. svibnja 2025.)

[8] Shah, J. Gradient Boosting, 2020., ResearchGate,  
[https://www.researchgate.net/profile/Jaineet-Shah/publication/354401342\\_Gradient\\_Boosting/links/6136e96a0360302a008430da/Gradient-Boosting.pdf?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19](https://www.researchgate.net/profile/Jaineet-Shah/publication/354401342_Gradient_Boosting/links/6136e96a0360302a008430da/Gradient-Boosting.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19), (pristupljeno 18. svibnja 2025.)

[9] Tsuchiev, B. *A Guide to The Gradient Boosting Algorithm*, 2023., Datacamp,  
<https://www.datacamp.com/tutorial/guide-to-the-gradient-boosting-algorithm>,  
(pristupljeno 20. svibnja 2025.)

[10] Sethi, A. *Support Vector Regression Tutorial for Machine Learning*, 2025., Analytics Vidhya, <https://www.analyticsvidhya.com/blog/2020/03/support-vector-regression-tutorial-for-machine-learning/>, (pristupljeno 25. svibnja 2025.)

[11] GeeksforGeeks. *Support Vector Regression (SVR) using Linear and Non-Linear Kernels in Scikit Learn*, 2025., <https://www.geeksforgeeks.org/machine-learning/support-vector-regression-svr-using-linear-and-non-linear-kernels-in-scikit-learn/>,  
(pristupljeno 1. lipnja 2025.)

[12] IBM. *What is random forest*, 2025., <https://www.ibm.com/think/topics/random-forest#:~:text=Random%20forest%20is%20a%20commonly,both%20classification%20and%20regression%20problems>, (pristupljeno 3. lipnja 2025.)

[13] GeeksforGeeks. *Decision Tree*, 2025., <https://www.geeksforgeeks.org/machine-learning/decision-tree/>, (pristupljeno 5. lipnja 2025.)

[14] Liberto D., Potters C., Rathburn P. *Understanding the House Price Index (HPI) and How It Is Used*, 2024., Investopedia, <https://www.investopedia.com/terms/h/house-price-index-hpi.asp>, (pristupljeno 5. lipnja 2025.)

[15] Lincoln International. *The eventual Normalization of the Housing Market will Create a Busy Period for Building Products M&A*, 2023.,  
<https://www.lincolninternational.com/perspectives/articles/the-eventual-normalization-of-the-housing-market-will-create-a-busy-period-for-building-products-ma/>,  
(pristupljeno 7. lipnja 2025.)

[16] Eurostat. *Housing price statistics – house price index*, 2025.,  
[https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Housing\\_price\\_statistics\\_-\\_house\\_price\\_index](https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Housing_price_statistics_-_house_price_index),  
(pristupljeno 8. lipnja 2025.)

[17] Wikipedija. *Linearna regresija*, 2024.,  
[https://hr.wikipedia.org/wiki/Linearna\\_regresija](https://hr.wikipedia.org/wiki/Linearna_regresija), (pristupljeno 9. lipnja 2025.)

[18] Chhetri, T.R., Dehury, C. K., Lind, A., Srirama, S. N., Fensel, A. *A combined metrics approach to cloud service reliability using artificial intelligence*, 2021., Research Gate,

[https://www.researchgate.net/profile/Tek-Raj-Chhetri/publication/356636837\\_A\\_Combined\\_Metrics\\_Approach\\_to\\_Cloud\\_Service\\_Reliability\\_using\\_Artificial\\_Intelligence/links/621e824d19ab0c3b4d2c9bbc/A-Combined-Metrics-Approach-to-Cloud-Service-Reliability-using-Artificial-Intelligence.pdf?\\_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19](https://www.researchgate.net/profile/Tek-Raj-Chhetri/publication/356636837_A_Combined_Metrics_Approach_to_Cloud_Service_Reliability_using_Artificial_Intelligence/links/621e824d19ab0c3b4d2c9bbc/A-Combined-Metrics-Approach-to-Cloud-Service-Reliability-using-Artificial-Intelligence.pdf?_tp=eyJjb250ZXh0Ijp7ImZpcnN0UGFnZSI6InB1YmxpY2F0aW9uIiwicGFnZSI6InB1YmxpY2F0aW9uIn19), (pristupljeno 9. lipnja 2025.)

[19] xxzoltanxx, Njuškalo scraper, <https://github.com/xxzoltanxx/NjuskaloScraper>, (pristupljeno 15. lipnja 2025.)

# Popis tablica

| <b>Tablica</b>                                                  | <b>Stranica</b> |
|-----------------------------------------------------------------|-----------------|
| Tablica 1. Broj podataka nakon čišćenja                         | 15.             |
| Tablica 2. Podjela regija                                       | 15.             |
| Tablica 3. Pogreške modela                                      | 22              |
| Tablica 4. Medijalne cijene nekretnina po regijama i kvartalima | 24              |

# Popis slika

| Slika                                                     | Stranica |
|-----------------------------------------------------------|----------|
| Slika 1. Linearna regresija                               | 6.       |
| Slika 2. Uzdizanje gradijenta za regresiju                | 8.       |
| Slika 3. Stroj s potpornim vektorima                      | 9.       |
| Slika 4. Slučajna šuma                                    | 11       |
| Slika 5. Stablo odluke                                    | 12.      |
| Slika 6. Procjena metodom linearne regresije              | 16.      |
| Slika 7. Procjena metodom gradijentnog poboljšanja modela | 18.      |
| Slika 8. Procjena metodom potpornih vektora               | 19.      |
| Slika 9. Procjena metodom slučajne šume                   | 20.      |
| Slika 10. Procjena metodom stabla odluke                  | 21.      |

# Sažetak

## Naslov:

Predviđanje cijena nekretnina regresijskim metodama strojnog učenja

## Sažetak:

Ovaj rad bavi se predviđanjem cijena nekretnina korištenjem regresijskih metoda strojnog učenja. Korišteni su službeni indeksi cijena nekretnina iz baze Državnog zavoda za statistiku i tražene cijene prikupljene s oglasnika Njuškalo.hr. Primijenjeni su modeli linearne regresije, gradijentnog poboljšanja modela, stroja s potpornim vektorima, slučajne šume te stabla odluke. Uz to, provedena je normalizacija tržišnih cijena u odnosu na indekse, što je omogućilo njihovu usporedbu kroz različite vremenske periode. Rezultati potvrđuju da složeniji modeli strojnog učenja nadmašuju tradicionalne linearne pristupe u točnosti i prilagodbi nelinearnim odnosima među značajkama, čime pružaju korisnu podlogu za analizu trendova i donošenje odluka u sektoru nekretnina.

## Ključne riječi:

nekretnine, regresija, strojno učenje, predikcija cijena, normalizacija

# Summary

**Title:**

Forecasting real estate prices using machine learning regression methods

**Summary:**

This paper addresses the prediction of real estate prices using regression-based machine learning methods. Official house prices indices from the Croatian Bureau of Statistics and listing prices collected from the Njuškalo.hr portal were used. Applied models include linear regression, gradient boosting, support vector regression, random forest and decision tree. Additionally, market price normalization relative to the indices was performed, enabling comparison across different time periods. The results confirm that more complex machine learning models outperform traditional linear approaches in accuracy and in capturing non-linear relationship between features, offering a valuable foundation for trend analysis and informed decision-making in the real estate sector.

**Keywords:**

real estate, regression, machine learning, price prediction, normalization