

Primljen / Received: 26.4.2024.

Ispravljen / Corrected: 9.10.2024.

Prihvaćen / Accepted: 11.9.2024.

Dostupno online / Available online: 10.2.2025.

# Novi pristup modeliranju metaansambla i usporedba modela strojnog učenja za procjenu cijene armature

## Autori:



Doc.dr.sc. **Sahin Tolga Guvel**, dipl.ing.građ.  
Sveučilište Osmaniye Korkut Ata, Turska  
Odjel za građevinarstvo  
[sahintolga guvel@osmaniye.edu.tr](mailto:sahintolga guvel@osmaniye.edu.tr)



Doc.dr.sc. **Abdulkadir Budak**, dipl.ing.građ.  
Sveučilište Osmaniye Korkut Ata, Turska  
Odjel za građevinarstvo  
[abudak@osmaniye.edu.tr](mailto:abudak@osmaniye.edu.tr)



Dr.sc. **Ibrahim Karataş**, dipl.ing.građ.  
Sveučilište Osmaniye Korkut Ata, Turska  
Odjel za građevinarstvo  
[ibrahimkaratas@osmaniye.edu.tr](mailto:ibrahimkaratas@osmaniye.edu.tr)  
Autor za korespondenciju

Prethodno priopćenje

**Sahin Tolga Guvel, Abdulkadir Budak, Ibrahim Karataş**

## Novi pristup modeliranju metaansambla i usporedba modela strojnog učenja za procjenu cijene armature

Rano utvrđivanje troškova u građevinskim projektima ključno je za planiranje troškova u svakoj fazi ulaganja. Realni izračun troškova učinkovit je način sprječavanja prekoračenja troškova do kojeg može doći u kasnijim fazama. Predviđanje cijene armaturnih šipki, uzimajući u obzir ekonomske pokazatelje, znatno utječe na troškove ulaganja i na odluke. Zato je u ovom istraživanju na temelju povijesnih podataka o materijalu za armiračke radove i ekonomskim pokazateljima upotrijebljeno devet algoritama strojnog učenja za dobivanje procijenjene cijene armaturne šipke za zakašnjenja od jedan, tri, šest, devet i dvanaest mjeseci. Algoritam za strojno učenje glasačkog metaansambla pokazao je najbolju izvedbu za sva istražena razdoblja zakašnjenja. Najuspješnija procjena dobivena je za zakašnjenje od tri mjeseca. Srednja apsolutna postotna pogreška (MAPE) i koeficijent determinacije (R<sup>2</sup>) za procjenu cijene armature tijekom tog razdoblja iznosile su 3,79 % odnosno 95,51 %.

### Ključne riječi:

procjena cijene armature, upravljanje izgradnjom, planiranje proizvodnje, metaansambl, strojno učenje

Research Paper

**Sahin Tolga Guvel, Abdulkadir Budak, Ibrahim Karataş**

## Novel meta-ensemble modelling approach and comparison of machine-learning models for rebar price estimation

The early determination of costs in construction projects is crucial for the planning of expenses throughout each investment stage. Making realistic cost calculations is an effective way of preventing cost overruns that may occur in later stages. Rebar price prediction by considering economic indicators significantly affects investment costs and decisions. Therefore, in this study, using historical data for rebar construction material and economic indicators, nine machine-learning algorithms were used to determine the estimated rebar price for 1-, 3-, 6-, 9-, and 12-month lags. The voting meta-ensemble machine-learning algorithm exhibited the best performance for all lag periods investigated. The most successful estimate was obtained for a 3-month lag period. The mean absolute percentage error (MAPE) and coefficient of determination (R<sup>2</sup>) values for the rebar price estimation during this period were 3.79 % and 95.51 %, respectively.

### Key words:

rebar price estimation, construction management, production planning, meta-ensemble, machine learning

## 1. Uvod

Istraživanja izvedivosti ključna su u građevinske sektoru za donošenje odluka o ulaganju, no ostaje izazov utvrditi donose li se odluke o ulaganju ispravno [1]. Zato je trošak ulaganja jedan od najistaknutijih parametara, osobito kada se uzimaju u obzir parametri koji utječu na odluke o ulaganju [2]. Međutim, troškovi ulaganja mijenjaju se s tehnološkim razvojem i rastućim očekivanjima naručitelja. Zato najpreciznija moguća procjena troškova ulaganja [3] utječe na prekoračenje troškova projekta, njegovo pravodobno završavanje i sposobnost postizanja kvalitete koju naručitelj želi [4].

Razni zadaci i postupci mogu utjecati na ukupne troškove izgradnje zgrade. Jedan od glavnih parametara jest utjecaj prvotno procijenjenih troškova građevnog materijala na troškove fluktuacije cijena tijekom projekta [4]. Među građevnim materijalima željezo [4, 5] i cement [6] najviše utječu na cijenu građevina [7]. Razlog je taj što promjena u njihovim jediničnim troškovima ima mnogo dinamičniju strukturu od ostalih građevnih materijala uz visoki ukupni trošak.

Prethodna istraživanja uglavnom su se temeljila na različitim kontaktnim točkama povezanima s izračunom približnih troškova izgradnje [8]. Budući da mnogi čimbenici utječu na troškove izgradnje [1, 2], primjena ekonomskih pokazatelja danas je česta. Zbog njihove dinamične i promjenjive strukture, osobito u zemljama u razvoju, utjecaj ekonomskih pokazatelja na sve resurse koji čine troškove izgradnje neizbježan je. Međutim, među istraživanjima na tu temu najopsežnija su ona koja razmatraju strukturalna obilježja tijekom faze izvedivosti za izračunavanje početnih troškova. Na primjer, indeks troškova [8], analiza jedinične cijene [3] vrijednosti i svojstva projekta/materijala [1, 5] mogu se upotrijebiti za izračun troškova izgradnje. Međutim, postoji malo istraživanja o procjeni budućih troškova izgradnje na temelju varijabilnosti troškova izgradnje pod utjecajem ekonomskih pokazatelja [4, 8-10]. Osim toga nekoliko je istraživanja ispitivalo promjene u cijeni građevnog materijala primjenom ekonomskih pokazatelja [4, 9, 11].

U tome kontekstu postoji nedostatak znanja u tome području, a koji je povezan s promjenom cijene građevnog materijala pod utjecajem ekonomskih pokazatelja. Vrlo je malo istraživanja provedeno na tu temu, posebno u području armature, što znatno utječe na cijenu građevnog materijala. Zato je ovo istraživanje ispitivalo promjenu cijena armature pod utjecajem ekonomskih pokazatelja.

U građevinske sektoru za procjenu troškova mogu se primijeniti različite metode analize kao što su osnovna statistička analiza, regresijska analiza i umjetna inteligencija (AI) [10, 12, 13]. Međutim, u istraživanjima koja su ispitivala učinke promjena cijena tijekom vremena primjenom ekonomskih pokazatelja uspješni rezultati dobiveni su analizom vremenskih serija, regresijskom analizom i aplikacijama umjetne inteligencije [2, 7, 14]. U posljednje vrijeme, ne samo u građevinske sektoru, već i u drugim sektorima, pokazalo se da se primjenom umjetne inteligencije postižu uspješniji rezultati [6, 15-17].

Danas gospodarski učinci poput inflacije uzrokuju neizvjesnost u vezi s troškovima građevinskog sektora. Postoji velika potreba za točnim predviđanjem troškova građevnih materijala kao što je armaturno željezo, na koje ekonomski pokazatelji utječu brzo, osobito u fazama prije ulaganja i fazama nabave. U skladu s time bi analiza promjene cijena armaturnih šipki pod utjecajem ekonomskih pokazatelja primjenom AI aplikacija i procjena najbližih troškova tijekom faze izvedivosti popunila spomenutu prazninu u znanju. Zato su ciljevi ovog istraživanja bili sljedeći:

- procjena budućih cijena armaturnog željeza na temelju prošlih cijena armaturnih šipki i podataka o ekonomskim pokazateljima odgovarajućeg datuma
- usporedba modela strojnog učenja koji se primjenjuju za predviđanje cijene armaturnog željeza i odabir modela s visokom prediktivnom vrijednosti
- predlaganje novog modela metaansambla za predviđanje cijena armaturnog željeza objedinjavanjem postojećih modela temeljenih na računalnim znanostima.

## 2. Pregled literature

Procjene troškova vrlo su učinkovite u građevinske sektoru, osobito u pogledu odluka o ulaganju tijekom faze izvedivosti. U proizvodnji građevnog materijala važno je da buduća procjena cijena armature, koja spada među materijale koji znatno utječu na ukupnu cijenu izgradnje, bude što je moguće bliže realnoj. Analizom prethodnih istraživanja na tu temu utvrđeno je da su se čimbenici upotrijebljeni za procjenu troškova razlikovali [4]. Za predviđanja cijene građevnog materijala Ou i sur. [7] primjenjivali su analizu vremenskih serija za procjenu cijena željezne rude i koksnog ugljena proizvedenog u čeličani. U modelima koje su Faghieh i Kashani [10] predložili za predviđanje budućih cijena asfalta, čelika i cementa u Sjedinjenim Američkim Državama (SAD) primijenjeni su sljedeći podaci: broj građevinskih dozvola, indeks potrošačkih cijena, rashodi u građevinskim radovima, broj zaposlenika u građevinske projektu, stopa zaposlenosti, bruto domaći proizvod, dohodak građevinskih radnika po satu, započeta izgradnja stanova, industrijska cijena plina, cijena željezne rude, indeks cijena industrijskih proizvođača, dohodak osoblja i parametri tipa *West Texas Intermediate*. Elfahham [8] proveo je analize vremenskih nizova koristeći podatke o cijenama opeke, čelika, cementa, pijeska i šljunka kako bi procijenio indeks troškova gradnje u Egiptu. Shiha i sur. [9] također su željeli procijeniti cijene čelika i cementa u Egiptu primjenom makroekonomskih pokazatelja kao što su indeks potrošačkih cijena, devizne pričuve, bruto domaći proizvod, stopa inflacije, kreditna stopa, novac u opticaju, indeks proizvođačkih cijena, stopa nezaposlenosti i devizni tečaj američkog dolara. Osim toga proveli su procjene za zakašnjenja od jedan, tri i šest mjeseci pri odabiru vremena zakašnjenja između makroekonomskih pokazatelja i cijena izlaznog materijala. Mir i sur. [4] razvili su model za utvrđivanje procijenjenih raspona cijena asfalta

i čelika u SAD-u. Za procjenu cijena čelika primijenjeni su podaci o indeksu potrošačkih cijena i započetoj izgradnji stanova te globalni parametri cijena željezne rude.

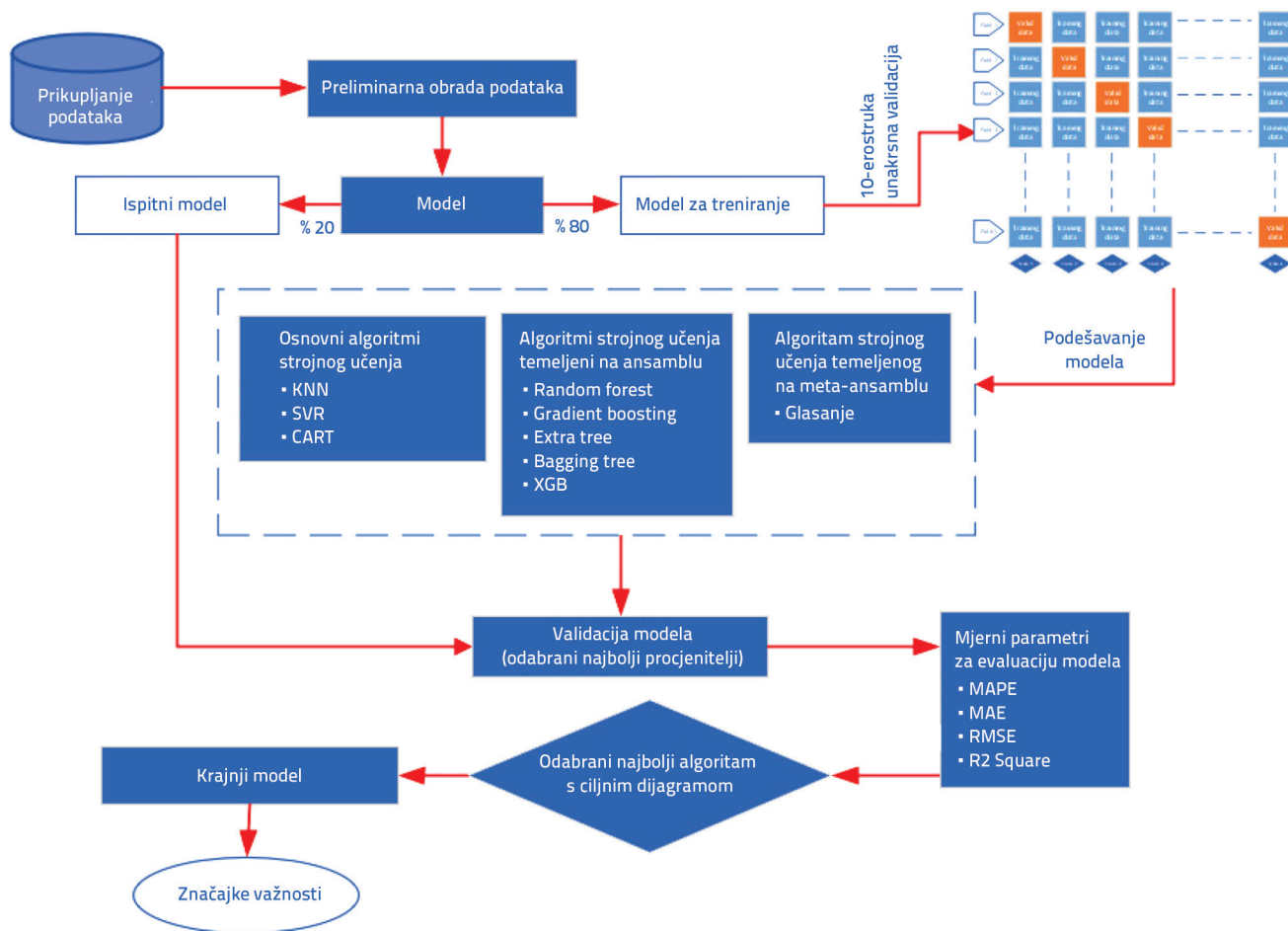
U građevinskom sektoru posljednjih se godina povećao intenzitet primjene aplikacija umjetne inteligencije u zadacima koji zahtijevaju predviđanje [4, 15, 18, 19]. Razlog leži u jednostavnoj primjeni aplikacija umjetne inteligencije i u većoj razini uspješnosti u odnosu na druge metode procjene [4, 15]. Uzimajući u obzir prethodna istraživanja o procjeni troškova u proizvodnji građevnog materijala, upotrijebljeno je mnogo različitih metoda za primjenu umjetne inteligencije.

Xu i sur. [12] primijenili su model stroja za ekstremno učenje *Gray* (GELM) za predviđanje integriran s *Gray* relacijskom analizom (GRA) i metodologijama za ekstremno učenje (engl. *extreme learning machine* - ELM). Također su usporedili predloženi model s autoregresivnim modelom s integriranim pomičnim prosjekom (engl. *autoregressive and integrated moving average* - ARIMA) i modelom generalizirane autoregresivne uvjetne heteroskedastičnosti (eng. *Generalized AutoRegressive Conditional Heteroskedasticity* - GARCH). Primjenom predloženog modela procijenjena srednja kvadratna pogreška (MSE) za cijenu armaturnog željeza iznosila je 0,0097. Predviđeni uspjeh predloženoga kombiniranog modela bio je veći od uspjeha drugih modela. Faghii i Kashani [10] predstavili su model vektorske korekcije pogreške (engl. *vector error correction* - VEC) za procjenu kratkoročnih i dugoročnih cijena građevnog materijala određivanjem odnosa između ekonomskih parametara i procijenjenih cijena materijala. Za procjenu cijene čelika jednovarijantni VEC model ima veću razinu točnosti od ostalih modela, s vrijednošću srednje apsolutne postotne pogreške (MAPE) od 56,88 %, s prediktivnim vrijednostima sa zakašnjenjem od tri mjeseca. Elfahham [8] predložio je formulu koja primjenjuje neuronske mreže, linearnu regresiju i autoregresivnu vremensku seriju za procjenu indeksa troškova izgradnje za betonske konstrukcije na temelju povijesnih zapisa o troškovima izgradnje. Kao osnova za izračun indeksa troškova gradnje korištene su povijesne cijene opeke, čelika, cementa, pijeska i šljunka. Prosjek izračunanih apsolutnih pogrešaka bio je sljedeći: 8,3 za neuronsku mrežu, 17,5 za regresijsku metodu, koja je bila najmanje točna, i 3,5 za metodu vremenskih nizova. Osnajivanje modela pomoću stope inflacije bilo bi učinkovitije. Shiha i sur. [9] razvili su tri modela predviđanja cijena armaturnih šipki. Prvi model izrađen je upotrebom proračunske tablice programa Excel, u kojoj je upotrijebljen genetski algoritam za smanjenje pogrešaka između predviđanja neuronske mreže i stvarnih cijena. Drugi je model razvijen upotrebom dodatka Excel pod nazivom *NeuralTools*, a treći model upotrebom programskog jezika *Python*, u verziji *Spider 3.6*. Za predviđanje cijena čelične armature bio je predložen treći model s MAPE vrijednostima od 7,0 % i 4,3 % za setove za obuku i testove. Taj je model odabran jer je učinkovitije zabilježio mjesečne fluktuacije u odnosu na prvi i drugi model. S druge strane,

Mir i sur. [4] predložili su metodu koja se temelji na umjetnoj neuronskoj mreži (ANN) za mjerenje nesigurnosti pomoću intervala predviđanja. Metoda optimalne procjene donje i gornje granice (optimalni LUBE) usvojena je kako bi se ANN osposobio za izravno generiranje intervala. Predloženom metodom procijenjene su cijene za asfalt i čelik u SAD-u. Na temelju rezultata zaključili su da je optimalna LUBE metoda dala realističnije rezultate od drugih metoda za procjenu cijene materijala. Chan i sur. [15] ističu da su aplikacije za strojno učenje učinkovite u procjeni i projektiranju građevnog materijala, no njihova primjena zahtijeva određenu razinu stručnosti. Xu i Zhang [12] upotrebljavaju Gaussove modele regresije procesa za prognozu dnevnih indeksa cijena proizvoda od čelika na kineskome tržištu. Primijenili su unakrižnu provjeru valjanosti i Bayesovu optimizaciju na raznim kernelima i osnovnim funkcijama. Njihov je cilj bio osigurati prognoze cijena proizvoda od čelika, koji su od znatne gospodarske važnosti za Kinu kao najvećeg potrošača i proizvođača čelika u svijetu [12]. Mi i sur. [14] prikupili su podatke o terminskim podacima za armature šipke od 2009. do 2020. i razvili model VMD-EEMD-LSTM za predviđanje terminskih cijena armature za sljedećih 14 dana trgovanja. Dai i sur. [13] prikupili su i analizirali cijene armaturnih šipki u kineskoj provinciji Guangdongu tijekom prve polovine 2023. kako bi razumjeli vremenske karakteristike sastava cijena armaturnih šipki. Prema literaturi, predviđanje cijena armature ključno je za gospodarstvo odgovarajuće zemlje. U skladu s time cilj ovog istraživanja jest primjena algoritama strojnog učenja za predviđanje cijena armature na temelju ekonomskih pokazatelja.

### 3. Metodologija istraživanja

Cilj je ovog istraživanja procijeniti buduće cijene armaturnih šipki primjenom podataka o prijašnjim cijenama armaturnih šipki i ekonomskih pokazatelja relevantnih datuma kako bi se usporedili modeli strojnog učenja primijenjeni za procjenu i stvaranje novog modela metaansambla. U tu svrhu razvijen je dijagram toka prikazan na slici 1. Prvo su ispitani ekonomski pokazatelji koji se upotrebljavaju u literaturi. Uzimajući u obzir te ekonomske pokazatelje, određeni su ekonomski pokazatelji koji su/mogu biti učinkoviti u politici oblikovanja cijena u turskoj industriji željeza i čelika. Prikupljeni su podaci o cijenama armature u razdoblju od 2002. do 2022. Potom su podaci namijenjeni za modeliranje podijeljeni na skup podataka za učenje (80 %) i skup podataka za testiranje (20 %). Točnosti modela izračunane su usporedbom rezultata dobivenih treniranjem podataka za učenje pomoću navedenih algoritama strojnog učenja s ispitnim podacima. Točnost modela procijenjena je primjenom mjernih parametara MAPE, MAE, RMSE i  $R^2$ . Naknadno je korišten ciljni dijagram za određivanje modela s najboljom procjenom. U konačnici su za model koji je dao najbolju procjenu utvrđene najvažnije značajke koje utječu na procjenu cijena armaturnih šipki u građevinarstvu.



Slika 1. Dijagram toka istraživanja

### 3.1. Prikupljanje podataka

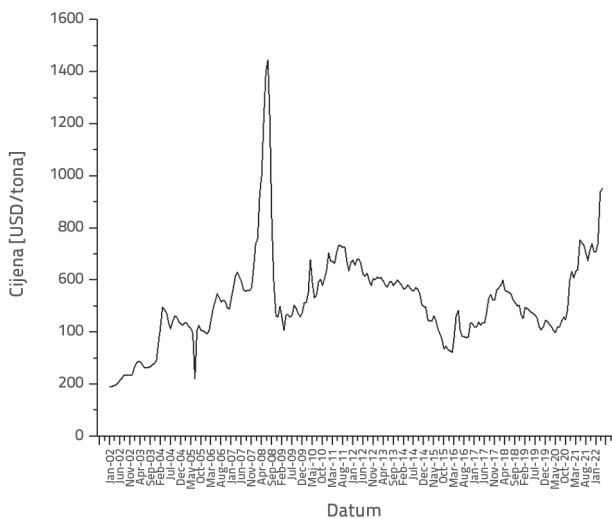
Prema literaturi, mnogi ekonomski pokazatelji utječu na obračun troškova građenja. Međutim, budući da svaka zemlja ima svoje vlastite ekonomske uvjete, ekonomski pokazatelji u izračunima troškova izgradnje razlikuju se ovisno o zemlji. S obzirom na to da se ovo istraživanje bavi procjenom troškova armature proizvedene u Turskoj, uzeti su u obzir ekonomski pokazatelji koji utječu na cijenu ugljena, plina, željeza, nafte, ferolegura, električne energije, indeksa proizvođačkih cijena, tečaja američkog dolara/turske lire i omjera kamatnih stopa u Turskoj. Cijene tih pokazatelja koje je izračunao Turski zavod za statistiku (TUIK) vrijednosti su indeksa cijena, čije se promjene mjere usporedbom kroz vrijeme. Zato su jedinice tih vrijednosti izražene kao 2003. = 100. Ekonomski pokazatelji korišteni u ovome istraživanju prikupljeni su iz baze podataka TUIK-a [20] između siječnja i ožujka 2022. Njihova opisna statistika prikazana je u tablici 1. Svi prikupljeni pokazatelji izraženi su kao čimbenici koji utječu na cijenu armature. Osim toga cijene armature izvezene iz Turske tijekom tog razdoblja dobivene su od tvrtke *SteelData* [21]. Promjene u cijenama armature između siječnja i ožujka 2022. prikazane su na slici 2. Na slici

3. prikazana je korelacijska matrica prikupljenih podataka. Ta matrica pomaže u određivanju odnosa između različitih podatkovnih točaka. Podaci upućuju na jaku korelaciju između ugljena, plina, željeza, nafte, ferolegura, električne energije, PPI-ja i tečaja američkog dolara/turske lire. Varijabla kamata pokazala je negativnu korelaciju s ostalim varijablama, dok je varijabla cijene željeza pokazala pozitivnu, iako slabu korelaciju s ostalim varijablama.

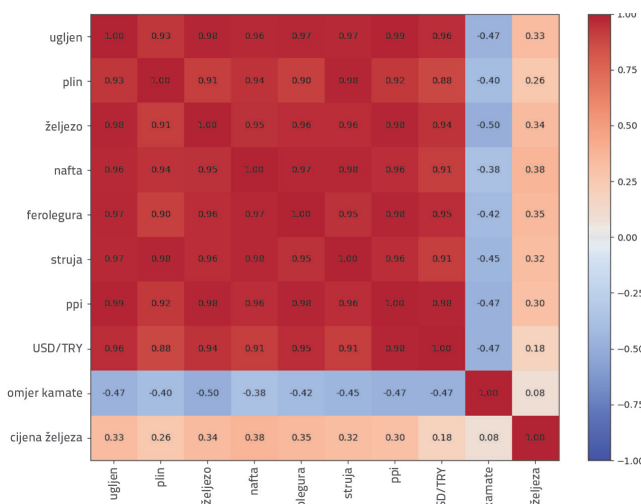
Na cijene armature u idućim mjesecima utječu ekonomski pokazatelji izračunani za određeni mjesec. Točna procjena cijene armature, koja je jedna od najvažnijih komponenti troška izgradnje, ključna je za završetak ulaganja unutar predviđenog budžeta. Buduće cijene armature treba procijeniti i razlikovati od projekta do projekta. U projektu za koji još nije donesena odluka o ulaganju važno je točno predvidjeti cijene devet i 12 mjeseci unaprijed. S druge strane, u projektima koji su sklopljeni pisanim ugovorima potrebno je točno predvidjeti cijene jedan, tri i šest mjeseci prije. Zato su u ovom istraživanju ispitani učinci ekonomskih pokazatelja na cijene armature nakon jedan, tri, šest, devet i dvanaest mjeseci. U tu svrhu izrađeno je pet skupova podataka. Ukupno su utvrđena 243, 241, 238, 235 i 232 boda za zakašnjenja od jedan, tri, šest, devet i 12 mjeseci.

Tablica 1. Opisna statistika skupa podataka sastavljenog za potrebe ovog istraživanja

Ekonomski pokazatelj	Maks.	Min.	Srednja vrijednost	Medijan	Standardno odstupanje	Asimetričnost	Spljoštenost
Ugljen	2319,26	95,33	421,49	377,79	347,60	2,71	10,22
Plin	2060,51	56,28	251,67	215,98	257,21	4,81	29,75
Željezo	3622,76	78,35	509,26	483,63	446,03	2,85	12,66
Nafta	4257,26	77,43	476,91	342,42	530,26	4,10	22,39
Ferolegura	3005,04	58,95	379,99	247,18	449,27	3,44	13,68
Struja	1610,03	85,46	230,87	202,29	183,87	4,29	24,80
Indeks cijena proizvođača (ppi)	1423,27	71,11	259,95	202,08	205,17	2,75	9,82
Tečaj za USD/TRY	14,67	1,17	2,99	1,78	2,57	2,27	5,72
Omjer kamate	0,38	0,04	0,18	0,17	0,08	0,85	0,50
Cijena armaturene šipke (USD/tona)	1445,00	190,00	518,95	498,25	176,93	1,58	6,32



Slika 2. Promjena cijena armature između siječnja i ožujka 2022.



Slika 3. Korelacijska matrica između različitih varijabli

### 3.2. Preliminarna obrada podataka

Preliminarna obrada provedena je pomoću dobivenih podataka kako bi se dodatno povećala sposobnost predviđanja modela strojnog učenja. Najprije su provedeni kontrola podataka koji nedostaju i procesi normalizacije. U prikupljenim podacima nisu pronađeni podaci koji nedostaju. Međutim, čini se da se ulazne vrijednosti razlikuju. Neke se vrijednosti kreću od nule do jedinice, dok se druge kreću od 100 do 500. Te su vrijednosti zahtijevale normalizaciju. Normalizacija je provedena na sljedeći način [22]:

$$Z_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{1}$$

### 3.3. Deseterostruka unakrsna validacija

Nakon što su podaci pripremljeni za modeliranje podijeljeni u omjeru 80 : 20, provedena je deseterostruka unakrsna provjera kako bi se smanjila pristranost podataka namijenjenih za treniranje i odredila slučajnost izvedbe modela [23]. Taj je proces analizirao 10 % podataka namijenjenih za treniranje kao validacijske podatke i 90 % podataka namijenjenih za treniranje. Taj proces ponovljen je deset puta, a konačni rezultat određen je prosjekom svih rezultata.

### 3.4. Algoritmi strojnog učenja

U ovome su istraživanju uspoređeni osnovni algoritmi i algoritmi strojnog učenja u ansamblu za procjenu cijene građevne armature.

#### 3.4.1. Regresija K-najbližeg susjeda

Algoritam K-najbližeg susjeda (KNN) procjenjuje izlaznu vrijednost treniranjem ulaznih značajki, njihovim skladištenjem



u prostoru značajki i usporedbom novih ulaznih podataka s tim prostorom značajki na temelju najbliže udaljenosti [24]. Glavni parametar koji treba odrediti za taj model jest cjelobrojna vrijednost  $k$  [25]. U algoritmu KNN-a najbolje predviđanje određuje se optimiranjem vrijednosti  $k$  koja označava susjeda [25]. Optimalna udaljenost ( $D$ ) mjeri se izračunavanjem udaljenosti od  $k$  najbližeg opažanja za svako opažanje kako slijedi:

$$D = \sqrt{\sum_{j=1}^k (x_i - y_j)^2} \quad (2)$$

gdje su  $x_i$  i  $y_j$  koordinatne vrijednosti svakog opažanja, a  $D$  udaljenost između vektora.

### 3.4.2. Regresija potpornim vektorima

Cilj modela regresije potpornim vektorima (engl. *Support Vector Regression - SVR*) jest pronaći funkciju koja je što ravnija i koja je najbliža značajkama prema najvećemu odstupanju  $\epsilon$  određene između predviđanja dobivenih iz uvježbanih podataka i stvarnih vrijednosti. Drugim riječima, zanemaruje pogreške sve dok je odstupanje malo i procjenjuje model prema pogreškama većima od vrijednosti odstupanja  $\epsilon$  [11, 26, 27]. Teorijski, funkcija SVR može se formulirati na sljedeći način [28]:

$$f(x) = w^t \times \varphi(x) + b \quad (3)$$

gdje je  $f(x)$  procijenjena vrijednost dobivena pomoću funkcije SVR, a  $w$  i  $b$  su koeficijenti određeni umanjivanjem prilagođene funkcije rizika, koja se izračunava na sljedeći način:

$$R = \frac{1}{2} w^2 + C \cdot \frac{1}{n} \sum_{i=1}^n |y_i - f(x_i)|_{\epsilon} \quad (4)$$

gdje je  $R$  regulirana funkcija rizika,  $C$  označava euklidsku normu i troškovni parametar koji određuje empirijski rizik, a  $|y_i - f(x_i)|_{\epsilon}$   $\epsilon$ -neosjetljivu funkciju gubitka, koja kontrolira pristranost i čini procjenu robusnom. Funkcija jezgre upotrebljava se za pretvorbu podataka u višedimenzionalni prostor, što doprinosi linearizaciji nelinearnih podataka i uspostavljanju snažnijeg odnosa između modela i podataka. U ovom je istraživanju kernel radialne bazne funkcije (RBF) korištena kao zadani model jezgre u primijenjenome SVR modelu. RBF kernel izračunava sličnost između dviju podatkovnih točaka na temelju udaljenosti između njih primjenom Gaussove distribucije [29].

### 3.4.3. Klasifikacijsko i regresijsko stablo

Za regresijske probleme u ovome istraživanju odabran je algoritam klasifikacijskog i regresijskog stabla (engl. *Classification And Regression Tree - CART*), koji je primijenjen na probleme klasifikacije i regresije. Drugim riječima, algoritam stabla

odlučivanja sastoji se od odluka i lisnih čvorova [30]. Svrha je tog algoritma povećati uspjeh predviđanja odvajanjem složenih heterogenih struktura u podacima u jednostavnije homogene podgrane. Počevši od gornjeg čvora, postupak razdjeljivanja ponavlja se sve dok lišće ne bude čisto. Kako bi se čvorove podijelilo na najinformativniji način, mora se definirati objektnu funkciju koju treba optimirati pomoću algoritma strojnog učenja temeljenog na stablima. Ovdje bi funkcija cilja trebala u najvećoj mogućoj mjeri povećati dobivene informacije (engl. *information gain - IG*) u svakoj podjeli, i to kako slijedi [31]:

$$IG(D_p, F) = I(D_p) - \left( \frac{N_{left}}{N_p} I(D_{left}) + \frac{N_{right}}{N_p} I(D_{right}) \right) \quad (5)$$

gdje  $F$  označava svojstvo izvođenja podjele,  $D_p$ ,  $D_{left}$  i  $D_{right}$  skupove podataka korijenskih i podređenih čvorova,  $I$  mjeru nečistoće,  $N_p$  ukupan broj uzoraka na nadređenome čvoru, a  $N_{left}$  i  $N_{right}$  broj uzoraka u podređenim čvorovima.

### 3.4.4 Regresija na temelju algoritma slučajnih šuma

Algoritam slučajnih šuma, koji procjenjuje i kombinira predviđanja dobivena primjenom višestrukih algoritama stabla odlučivanja, kombinira metode odvajanja (*bagging*) i metode slučajnog podskupa [32]. Drugim riječima, vrijednosti opažanja za stabla koja tvore slučajnu šumu odabiru se odvajanjem slučajnog uzorka, a varijable se odabiru slučajnim podskupom. Predviđene vrijednosti izračunavaju se ponderiranjem vrijednosti pogreške iz svakog stabla. Osnovni parametri te metode određeni su kao  $m$  (broj parametara) i  $k$  (broj stabala) [32, 33].

### 3.4.5. Stroj za povećanje gradijenta

Stroj za povećanje gradijenta (engl. *Gradient Boosting Machine - GBM*) moćan je algoritam strojnog učenja koji se primjenjuje i u zadacima regresije i klasifikacije. Gradi modele postupno i generalizira ih, omogućavajući optimiranje diferencijabilne funkcije gubitka. Drugim riječima, to je fleksibilan i moćan alat za prediktivno modeliranje koji iskorištava principe gradijentnog spuštanja i pojačavanja za iterativno poboljšanje točnosti modela. Posebno je učinkovit u kombinaciji sa stablima odlučivanja i osigurava ravnotežu između složenosti modela i mogućnosti tumačenja [34]. Niz setova modela s GBM-om stvoren je uklapanjem pogrešaka prethodnog modela. Taj je postupak bio ponavljan sve dok nije postignut maksimalan broj ponavljanja [35].

### 3.4.6. Regresija dodatnim stablima

Algoritam dodatnog stabla, koji se pojavio kao proširenje algoritma slučajnih šuma, trenira svaki osnovni procjenitelj primjenom slučajnog podskupa značajki [36]. Svi podaci o treningu korišteni su za trening svakog stabla odlučivanja [37].

### 3.4.7. Regresija odvajajućim (bagging) stablom

Algoritam *bagginga* predložen u [38] dobiven je kombiniranjem više algoritama stabla odlučivanja. Novo predviđanje postignuto je po prosječnoj vrijednosti predviđanja algoritama stabla odlučivanja. To pomaže smanjiti velika odstupanja u algoritmu stabla odlučivanja.

### 3.4.8. Regresija pomoću eXtreme gradient boostinga (XGBoost)

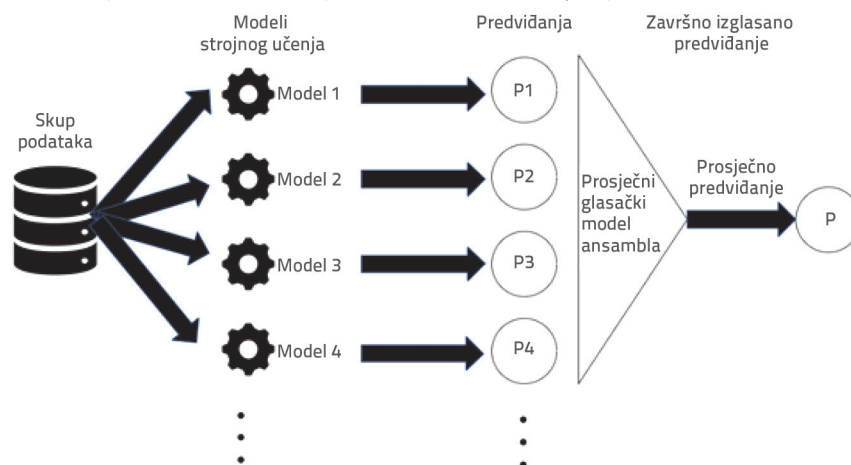
XGBoost razvijen je za poboljšanje brzine i performansi predviđanja GBM algoritma [39]. Uz visoke prediktivne performanse tog algoritma odlikuje ga i značajka koja sprječava pretreniranje [40]. XGBoost sastoji se od funkcije gubitka i normalizacijskog člana u procesu učenja kao u jednadžbi 6. Normalizacijski član kontrolira složenost modela, izbjegavajući pretreniranje, dok funkcija gubitka izračunava razliku između svake predviđene i prave vrijednosti [41].

$$Aim^p = \sum_{i=1}^n l(\bar{y}_i, y_i) + \sum_{i=1}^p \sigma(f_i) \quad (6)$$

gdje je  $l$  funkcija gubitka,  $n$  broj korištenih promatranja,  $\sigma$  normalizacijski član, a  $f_i$  procjena na koraku  $i$ .

### 3.4.9. Model prosječnoga glasačkog ansambla

Metoda prosječnoga glasačkog ansambla temelji se na shemi glasovanja koja kombinira prethodno opisane algoritme strojnog učenja za postizanje vrhunske izvedbe. Svrha prosječnoga glasovanja jest prosjek predviđanja više modela, čime se dobivaju točnija ukupna predviđanja. Kombinira predviđanja uz pretpostavku da je model s najviše glasova pobjednik kao što je to prikazano na slici 4 [42]. U tome pristupu predviđanjima svakoga sastavnog modela dodijeljene su jednake težine. Predviđanja modela kombiniraju se i izračunava se prosjek.



Slika 4. Struktura modela glasačkog ansambla [34]

Ta metoda sprječava algoritme strojnog učenja u tome da generiraju razne pogreške i pretreniranje. Također poboljšava ukupnu izvedbu kombinirajući prednosti različitih modela.

### 3.5. Mjerni parametri evaluacije modela

Za izračun i usporedbu točnosti procjene i stope pogreške modela analiziranih za procjenu cijene armature, mjerni parametri MAPE-a, srednje apsolutne pogreške (MAE), korijena srednje kvadratne pogreške (RMSE) i koeficijenta determinacije ( $R^2$ ) korišteni su za procjenu modela. MAPE jest metrika procjene koja daje stopu pogreške modela kao postotak i izračunava se onako kako je to opisano u jednadžbi 7. MAE je izračunan uzimanjem apsolutne vrijednosti razlike između svake predviđene vrijednosti i stvarne vrijednosti kao u jednadžbi 8. Što je izračunana vrijednost MAE bliža nuli, to je bolje predviđanje modela. RMSE, koji je široko korištena metrika za regresijske probleme, izračunan je pomoću jednadžbe 9. U slučaju RMSE-a minimalne izračunane vrijednosti kao što je MAE upućuju na povoljno predviđanje. Osim toga, budući da ima iste jedinice kao ovisna varijabla, RMSE se upotrebljava češće od MSE-a i MAE-a za procjenu izvedbe regresijskih modela u usporedbi s drugim slučajnim modelima. Međutim, mjerni parametar regresijske procjene, koja se razlikuje između nule i jedinice (omogućujući izvrsno uklapanje modela dok se približava jedinici), jest vrijednost  $R^2$ , izračunana pomoću jednadžbe 10. Rezultat za  $R^2$  također se može ocijeniti kao omjer varijance objašnjen modelom.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100 \quad (7)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \quad (8)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (10)$$

pri čemu vrijedi:  
 $n$  - broj opažanja u skupu podataka  
 $y_i$  - stvarna vrijednost  
 $\hat{y}_i$  - predviđena vrijednost  
 $\bar{y}_i$  - srednja vrijednost stvarnih vrijednosti.

### 3.6. Ciljni dijagram

U ovome istraživanju ciljni dijagram [43] korišten je za usporedbu izvedbe predviđanja algoritama osnovnoga

strojnog učenja, strojnog učenja u ansamblu i metaansamblu koji se upotrebljavaju za predviđanje cijene konstrukcijske armature, jer je teško usporediti ukupno devet algoritama. Ti se dijagrami mogu upotrijebiti za određivanje toga koji algoritam ima najbolju izvedbu predviđanja pomoću grafičkog prikaza. Ciljni dijagram proizlazi iz odnosa između statističke metrike pristranosti (B) i nepristranoga kvadratnog korijena razlike (uRMSD). Odnos između B-a i uRMSD-a vraća kvadratni korijen razlike (RMSD) kako slijedi:

$$\text{RMSD}^2 = B^2 + \text{uRMSD}^2 \quad (11)$$

Očekuje se da algoritmi s boljim predvidivim performansama imaju najniže vrijednosti RMSD-a.

### 3.7. Važnost značajki

Važnost značajke jest koncept koji mjeri utjecaj ulaznih značajki na rezultate predviđanja modela, čime se poboljšava sposobnost objašnjavanja modela [44]. Korišten je u brojnim prethodnim istraživanjima [16, 45]. Vrijednosti važnosti značajki modela metaansambla izračunane su pomoću prosječne vrijednosti modela strojnog učenja primijenjenog u ovome istraživanju. Drugim riječima, srednje vrijednosti važnosti značajki dobivene su na temelju prosječnih koeficijenata svake značajke dobivene iz rezultata algoritma u ovome istraživanju. Te vrijednosti pokazuju stupanj utjecaja tih elemenata na prognoze cijena željeza.

### 4. Rezultati i rasprava

U ovome su radu procijenjene cijene armaturnog čelika, jednog od građevnih materijala s najvećim utjecajem na troškove izgradnje, tijekom sljedećih jedan, tri, šest, devet i 12 mjeseci, ispitane u dvije faze pomoću modela strojnog učenja. Prvo, za svaki mjesec vrijednosti MAPE, MAE, RMSE i  $R^2$  za svaki model određene su pomoću osnovnih modela strojnog učenja (KNN, SVR i CART), a usporedno su provedeni modeli strojnog učenja temeljeni na ansamblu (slučajna šuma, regresija gradijentnim povećanjem, dodatno stablo, Odvajajuće (bagging) stablo, XGB

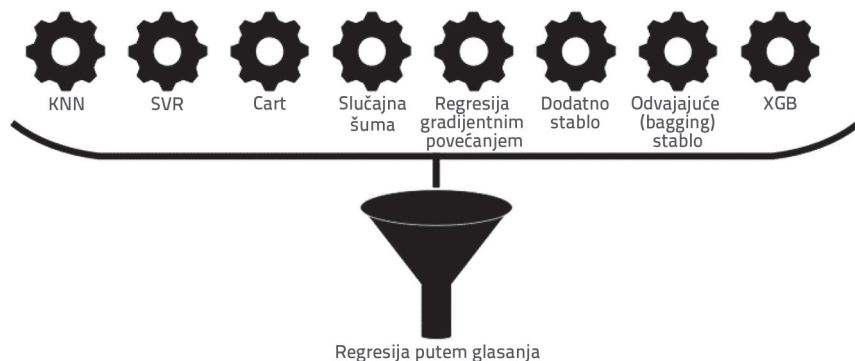
stablom, regresija XGB-om). Pri analizi svaki od tih modela zahtijeva specifične parametre koji se moraju prilagođavati tijekom treninga. Hiperparametri korišteni u modelima navedeni su u tablici 2.

Nakon toga, a kako bi se povećala točnost predviđanja, testirane su različite varijacije osnovnog modela i modela ansambla te je njihovim kombiniranjem izrađen regresijski model metaansambla glasanja kao što je to prikazano na slici 5. Taj je model zatim uspoređen s drugim modelima za procjenu. Potom je grafički prikazan i ispitani odnos između procjena cijena za zakašnjenja od jedan, tri, šest, devet i dvanaest mjeseci i stvarnih vrijednosti te učinak svakog ulaznog čimbenika na izvedbu modela.

Mjerila evaluacije dobivena analizom provedenom prema učincima ekonomskih pokazatelja za slučaj procjene cijene armature s kašnjenjem od jednog mjeseca prikazana su u tablici 3. Kao što se može vidjeti, usporedbom osnovnog i skupnog modela kao model s najvećom točnosti predviđanja prema mjernim parametrima MAPE i MAE određen je model slučajne šume, dok najveću točnost predviđanja prema mjernim parametrima RMSE i  $R^2$  ima model *bagging* stabla. Na temelju usporedbe najpreciznijim je utvrđen model regresije glasačkim ansamblom. Vrijednosti MAPE, MAE, RMSE i  $R^2$  za taj model iznosile su 3,90 %, 19,2721, 28,59717 odnosno 0,953522.

Mjerila evaluacije dobivena analizom provedenom prema učincima ekonomskih pokazatelja za slučaj procjene cijene armature s kašnjenjem od tri mjeseca prikazana su u tablici 4. Kao što se može vidjeti, usporedbom osnovnog modela i modela temeljenog na ansamblu kao model s najvećom točnosti predviđanja prema svim metrikama procjene utvrđen je model *bagging* stabla. Usporedbom svih modela model regresije glasačkim metaansamblom utvrđen je kao najprecizniji. Vrijednosti MAPE, MAE, RMSE i  $R^2$  za taj model iznosile su 3,80 %, 19,06886, 28,10399 odnosno 0,955111.

Mjerni parametri za evaluaciju dobiveni analizom provedenom prema učincima ekonomskih pokazatelja za slučaj procjene cijene armature sa zakašnjenjem od šest mjeseci prikazani su u tablici 5. Kao što se može vidjeti, usporedbom osnovnog modela i modela temeljenog na ansamblu kao model s najvećom točnosti predviđanja prema svim mjernim parametrima za procjenu utvrđen je model dodatnog stabla. Usporedbom svih modela model glasačkog metaansambla utvrđen je kao najtočniji model procjene, osim za mjerne parametre  $R^2$ . Vrijednosti MAPE, MAE i RMSE za taj model iznosile su 3,92 %, 19,62414 odnosno 28,95394. Što se tiče mjernih parametara za  $R^2$ , vrijednosti modela dodatnog stabla i modela regresije glasanjem vrlo su bliski, a iznose 0,952496 odnosno 0,952355. Mjerni parametri za evaluaciju dobiveni analizom provedenom prema učincima ekonomskih pokazatelja za procjenu



Slika 5. Glasanje za model metaansambla



Tablica 2. Hiperparametri koji se upotrebljavaju za svaki model strojnog učenja

	Zakašnjenje od 1 mjeseca	Zakašnjenje od 3 mjeseca	Zakašnjenje od 6 mjeseci	Zakašnjenje od 9 mjeseci	Zakašnjenje od 12 mjeseci
<b>KNN</b>	k:3, težine: ujednačene	k:3, težine: ujednačene	k:2, težine: ujednačene	k:4, težine: ujednačene	k:3, težine: ujednačene
<b>SVR</b>	C:7200, kernel: RBF	C:7400, kernel: RBF	C:9500, kernel: RBF	C:51000, kernel: RBF	C:54400, kernel: RBF
<b>CART</b>	max_leaf_nodes:30, min_samples_split: 3	max_leaf_nodes: 38, min_samples_split: 5	max_leaf_nodes: 18, min_samples_split: 2	max_leaf_nodes: 5, min_samples_split: 8	max_leaf_nodes: 33, min_samples_split: 2
<b>Slučajna šuma</b>	bootstrap: False, max_depth: 200, max_features: sqrt, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 600	bootstrap: False, max_depth: 750, max_features: sqrt, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 300	bootstrap: False, max_depth: 200, max_features: sqrt, min_samples_leaf: 2, min_samples_split: 3, n_estimators: 200	bootstrap: False, max_depth: 300, max_features: sqrt, min_samples_leaf: 2, min_samples_split: 4, n_estimators: 20	bootstrap: False, max_depth: 1100, max_features: sqrt, min_samples_leaf: 4, min_samples_split: 5, n_estimators: 9
<b>Pojačavanje gradijenta</b>	learning_rate: 0,05, max_depth: 50, max_features: 0,1, min_samples_leaf: 2, n_estimators: 2000, subsample: 1,0	learning_rate: 0,05, max_depth: 10, max_features: 0,01, min_samples_leaf: 2, n_estimators: 400, subsample: 0,5	learning_rate: 0,05, max_depth: 20, max_features: 0,3, min_samples_leaf: 2, n_estimators: 500, subsample: 0,5	learning_rate: 0,05, max_depth: 90, max_features: 0,5, min_samples_leaf: 5, n_estimators: 700, subsample: 0,5	learning_rate: 0,01, max_depth: 30, max_features: 0,5, min_samples_leaf: 4, n_estimators: 800, subsample: 1,0
<b>Dodatno stablo</b>	bootstrap: False, max_depth: 100, max_features: 0,75, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100	bootstrap: False, max_depth: 50, max_features: 0,25, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 400	bootstrap: False, max_depth: 30, max_features: 0,8, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 800	bootstrap: False, max_depth: 30, max_features: 0,6, min_samples_leaf: 2, min_samples_split: 2, n_estimators: 10	bootstrap: False, max_depth: 40, max_features: 1,0, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 8
<b>Bagging stablo</b>	base_estimator: None, bootstrap: False, bootstrap_features: True, max_features: 0,5, max_samples: 1,0, n_estimators: 800	base_estimator: None, bootstrap: False, bootstrap_features: netočno, max_features: 0,6, max_samples: 1,0, n_estimators: 800	base_estimator: None, bootstrap: False, bootstrap_features: False, max_features: 0,5, max_samples: 1,0, n_estimators: 800	base_estimator: None, bootstrap: Istina, bootstrap_features: True, max_features: 0,5, max_samples: 1,0, n_estimators: 60	base_estimator: None, bootstrap: False, bootstrap_features: False, max_features: 5, max_samples: 100, n_estimators: 200
<b>XGB</b>	colsample_bytree: 0,5, learning_rate: 0,1, max_depth: 5, min_child_weight: 2, n_estimators: 1500, subsample: 1,0	colsample_bytree: 0,3, learning_rate: 0,05, max_depth: 20, min_child_weight: 1, n_estimators: 1000, subsample: 0,1	colsample_bytree: 0,5, learning_rate: 0,05, max_depth: 10, min_child_weight: 3, n_estimators: 3000, subsample: 1,0	colsample_bytree: 0,7, learning_rate: 0,1, max_depth: 12, min_child_weight: 11, n_estimators: 200, subsample: 1,0	colsample_bytree: 0,7, learning_rate: 0,1, max_depth: 20, min_child_weight: 1, n_estimators: 50, subsample: 1,0

Tablica 3. Rezultati modela strojnog učenja za procjenu cijene armature sa zakašnjenjem od jednog mjeseca

Model	MAPE [%]	MAE	RMSE	R <sup>2</sup>
KNN	5,97	28,44813	39,48362	0,9114
SVR	8,51	39,51059	52,64472	0,842489
Cart	7,83	37,09775	57,3994	0,812753
Slučajna šuma	<b>4,37</b>	<b>21,40749</b>	31,50025	0,943607
Regresija gradijentnim povećanjem	4,79	23,0923	30,915	0,945683
Regresija dodatnim stablom	4,78	23,44043	32,84048	0,938706
Regresija odvajajućim (bagging) stablom	4,49	21,64282	<b>30,83886</b>	<b>0,94595</b>
Regresija XGB-om	4,69	23,3149	34,86586	0,930912
Regresija glasanjem	<b>3,90</b>	<b>19,2721</b>	<b>28,59717</b>	<b>0,953522</b>

Tablica 4. Rezultati modela strojnog učenja za procjenu cijene građevnih armaturnih šipki sa zakašnjenjem od tri mjeseca

Model	MAPE [%]	MAE	RMSE	R <sup>2</sup>
KNN	10,90	52,51305	74,39734	0,805022
SVR	12,87	58,86498	89,51642	0,717722
Cart	9,24	41,96255	69,92181	0,827775
Slučajna šuma	8,45	37,0673	52,3793	0,903352
Regresija gradijentnim povećanjem	9,82	44,61394	61,82911	0,865334
Regresija dodatnim stablom	8,63	36,52421	52,80427	0,901778
Regresija odvajajućim ( <i>bagging</i> ) stablom	<b>8,12</b>	<b>35,40851</b>	<b>51,50008</b>	<b>0,90657</b>
Regresija XGB-om	11,38	51,66069	67,52914	0,83936
Regresija glasanjem	<b>3,80</b>	<b>19,06886</b>	<b>28,10399</b>	<b>0,955111</b>

Tablica 5. Rezultati modela strojnog učenja za procjenu cijene građevnih armaturnih šipki sa zakašnjenjem od šest mjeseci

Model	MAPE [%]	MAE	RMSE	R <sup>2</sup>
KNN	7,82	51,57516	100,8423	0,758166
SVR	13,58	80,882	155,4422	0,425395
Cart	10,99	52,47877	72,66854	0,874419
Slučajna šuma	8,33	45,35044	66,37932	0,895215
Regresija gradijentnim povećanjem	7,37	40,54662	57,1967	0,922201
Regresija dodatnim stablom	<b>6,62</b>	<b>34,20766</b>	<b>44,69392</b>	<b>0,952496</b>
Regresija odvajajućim ( <i>bagging</i> ) stablom	8,37	41,76773	53,85572	0,931024
Regresija XGB-om	8,17	42,14981	54,58297	0,929149
Regresija glasanjem	<b>3,92</b>	<b>19,62414</b>	<b>28,95394</b>	<b>0,952355</b>

Tablica 6. Rezultati modela strojnog učenja za procjenu cijene građevnih armaturnih šipki sa zakašnjenjem od devet mjeseci

Model	MAPE [%]	MAE	RMSE	R <sup>2</sup>
KNN	9,28	56,20286	120,9789	0,719113
SVR	13,60	83,25784	184,2311	0,348614
Cart	17,57	104,1335	193,7997	0,279194
Slučajna šuma	9,23	55,1738	119,7518	0,724782
Regresija gradijentnim povećanjem	<b>8,52</b>	<b>44,34587</b>	<b>75,30403</b>	<b>0,89117</b>
Regresija dodatnim stablom	8,95	49,73613	95,94524	0,823332
Regresija odvajajućim ( <i>bagging</i> ) stablom	10,11	56,2993	115,2702	0,744996
Regresija XGB-om	9,93	50,05226	81,11086	0,873739
Regresija glasanjem	<b>4,42</b>	<b>22,19465</b>	<b>31,59304</b>	<b>0,943274</b>

cijene armature sa zakašnjenjem od devet mjeseci prikazani su u tablici 6. Kao što se može vidjeti, usporedbom osnovnog modela i modela temeljenog na ansamblu kao model s najvećom točnosti predviđanja prema svim metrikama procjene utvrđen je model s gradijentnim povećanjem. Usporedbom svih modela model regresije glasanjem utvrđen je kao najtočniji. Vrijednosti MAPE, MAE, RMSE i R<sup>2</sup> za taj model iznosile su 4,42 %, 22,19465, 31,59304 odnosno 0,943274.

Mjerni parametri za evaluaciju dobiveni analizom provedenom prema učincima ekonomskih pokazatelja za procjenu cijene armature sa zakašnjenjem od 12 mjeseci prikazani su u tablici 7. Kao što se može vidjeti, usporedbom osnovnog modela i modela temeljenog na ansamblu kao model s najvećom točnosti predviđanja prema svim mjernim parametrima procjene utvrđen je model *XGBoost*. Usporedbom svih modela model regresije glasanjem utvrđen je kao najtočniji. Vrijednosti MAPE, MAE, RMSE i R<sup>2</sup> za taj model iznosile su 4,11 %, 20,39988, 30,71019 odnosno 0,9464.

Tablica 7. Rezultati modela strojnog učenja za procjenu cijene građevnih armaturnih šipki sa zakašnjenjem od 12 mjeseci

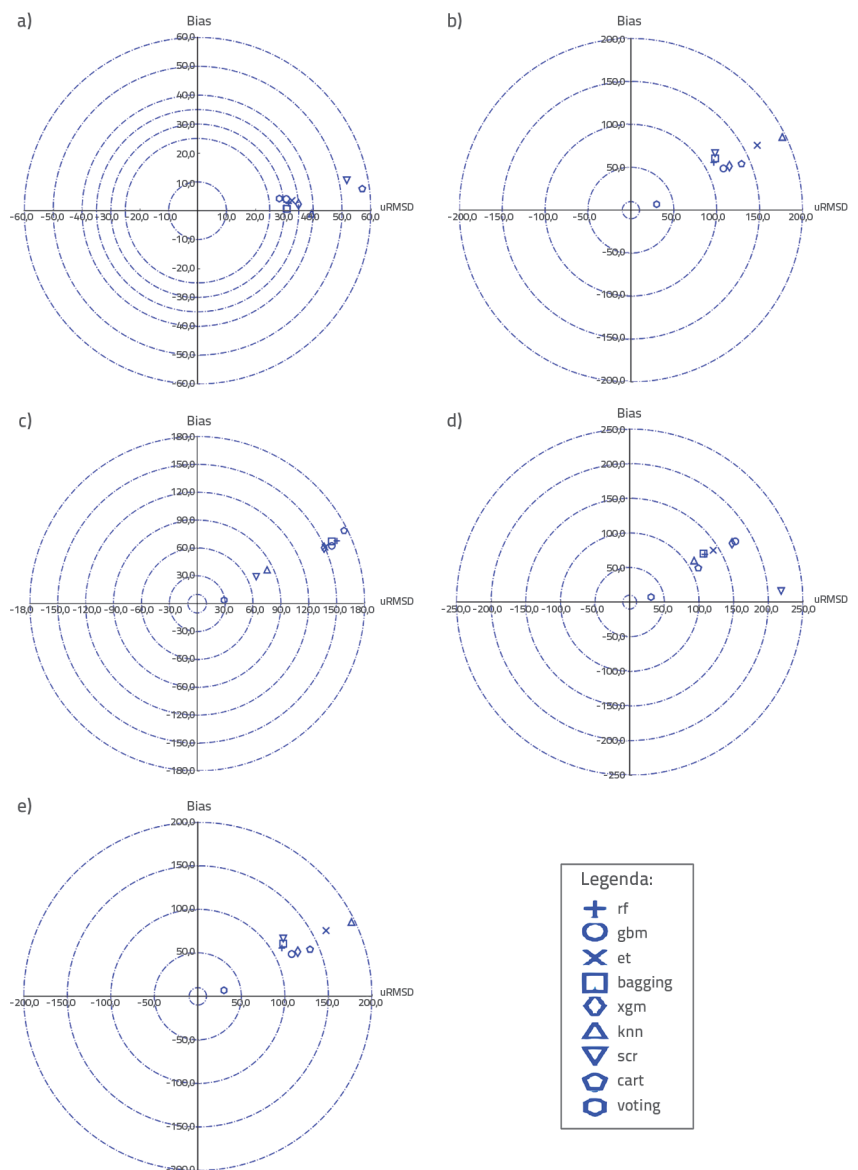
Model	MAPE [%]	MAE	RMSE	R <sup>2</sup>
KNN	12,12	70,28362	149,2291	-0,09984
SVR	8,86	50,52747	86,82169	0,627712
Cart	08,65	47,66665	77,24122	0,70534
Slučajna šuma	7,98	45,74283	65,88371	0,785623
Regresija gradijentnim povećanjem	6,31	35,44273	53,35277	0,859416
Regresija dodatnim stablom	7,20	41,23574	76,28565	0,712586
Regresija odvajajućim ( <i>bagging</i> ) stablom	7,63	40,80243	58,88146	0,82877
Regresija XGB-om	<b>5,81</b>	<b>30,42101</b>	<b>43,77327</b>	<b>0,905367</b>
Regresija glasanjem	<b>4,11</b>	<b>20,39988</b>	<b>30,71019</b>	<b>0,9464</b>

Prema tim rezultatima, modeli glasačkog ansambla pokazali su najtočnije vrijednosti procjene u svim vremenskim odmacima. Također, osim modela glasanja modeli ansambla pokazali su bolju procjenu pri svakome mjesečnom kašnjenju. Nasuprot tome, osnovni modeli strojnog učenja pokazali su niži uspjeh u procjeni u odnosu na model ansambla. Kako bi se usporedili modeli strojnog učenja koji se upotrebljavaju za procjenu cijene armature, izrađeni su ciljni dijagrami za zakašnjenja od jedan, tri, šest, devet i 12 mjeseci kao što je to prikazano na slici 6. Na temelju tih rezultata mogu se izvesti sljedeći zaključci:

U slučaju zakašnjenja od jednog mjeseca pogreške u procjeni ansambla i metaansambla među analiziranim modelima bile su bliske jedna drugoj, a model s najnižom vrijednosti pogreške u procjeni korišten je kao model glasačkog metaansambla (slika 6.a).

Postojala je znatna razlika u pogrešci predviđanja između metaansambla i ansambla te osnovnih modela strojnog učenja za kašnjenja od tri, devet i 12 mjeseci. Model s najnižom pogreškom procjene korišten je kao model glasačkog ansambla (slike 6.b, d, e).

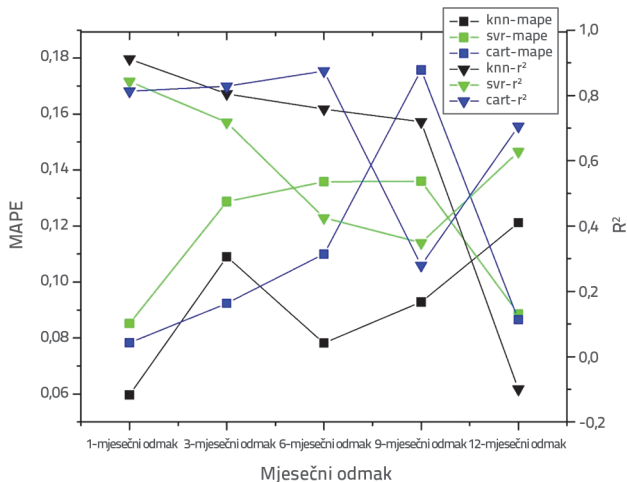
Postojala je znatna razlika u pogrešci predviđanja između metaansambla i ansambla te osnovnih modela strojnog učenja za zakašnjenja od šest mjeseci. Iako je model s najnižom procijenjenom vrijednosti pogreške bio glasački metaansambl, predviđanja najbliža tome modelu određena su osnovnim



Slika 6. Ciljni dijagrami za sve modele strojnog učenja analizirane u ovome istraživanju: a) zakašnjenje od 1 mjeseca; d) zakašnjenje od 3 mjeseca; c) zakašnjenje od 6 mjeseci; d) zakašnjenje od 9 mjeseci; e) zakašnjenje od 12 mjeseci

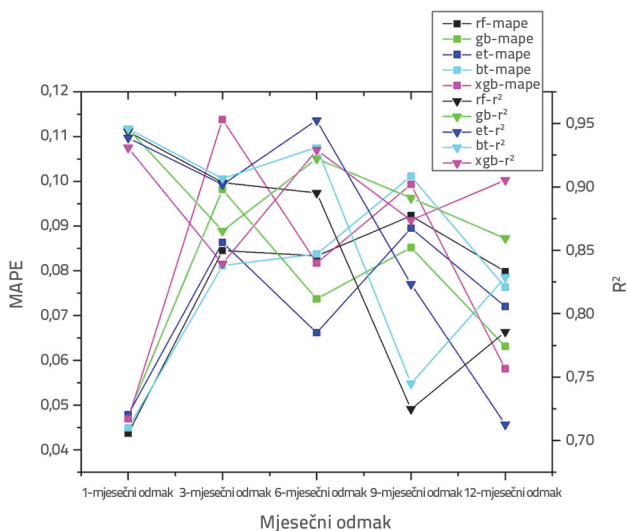
modelima strojnog učenja pomoću KNN-a i SVR-a (slika 6.c). Osnovni model, ansambl i metaansambl modeli strojnog učenja primijenjeni u procjeni cijene armature uspoređeni su na temelju mjernih parametara MAPE-a i R<sup>2</sup> evaluacije za razdoblja kašnjenja od jedan, tri, šest, devet i 12 mjeseci. Na temelju tih rezultata mogu se izvesti sljedeći zaključci:

Utvrđeno je da razdoblje zakašnjenja od jednog mjeseca pokazuje najtočniju prediktivnu vrijednost među osnovnim modelima strojnog učenja. Osnovni model strojnog učenja s najpreciznijom prediktivnom vrijednosti tijekom tog razdoblja bio je model KNN-a (slika 7.).



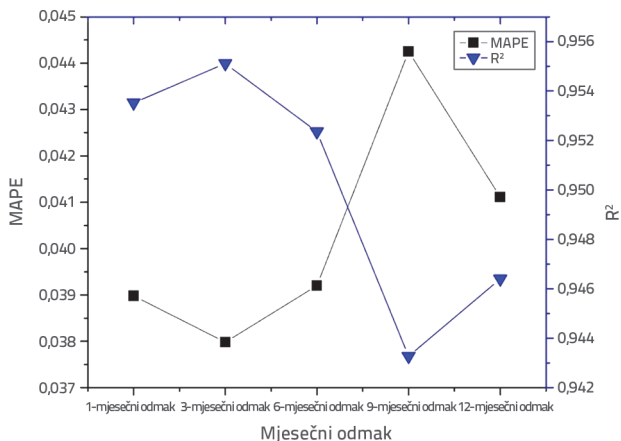
Slika 7. Usporedba osnovnih modela strojnog učenja za promatrana razdoblja zakašnjenja

Utvrđeno je da zakašnjenje od jednog mjeseca pokazuje najtočniju prediktivnu vrijednost među skupnim modelima strojnog učenja. Iako su svi modeli u tome razdoblju imali slične vrijednosti predviđanja, model s najnižom vrijednosti MAPE bio je model slučajne šume, dok je model s najvišom vrijednošću R<sup>2</sup> bio model bagging stabla (slika 8).



Slika 8. Usporedba modela strojnog učenja temeljenih na ansamblu za promatrana razdoblja zakašnjenja

Za model strojnog učenja glasačkog metaansambla razdoblje s najvećom prediktivnom vrijednosti bilo je razdoblje sa zakašnjenjem od tri mjeseca. Osim toga procijenjene vrijednosti za zakašnjenje od jednog i šest mjeseci bile su relativno visoke (slika 9.).



Slika 9. Usporedba modela strojnog učenja glasačkog metaansambla za promatrana razdoblja zakašnjenja

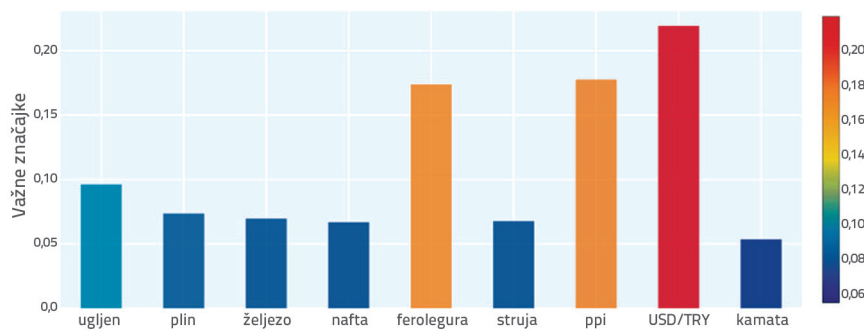
U ovom su istraživanju uspoređeni modeli strojnog učenja za procjenu cijene armature. Model glasačkog metaansambla dobiven kombinacijom tih modela dao je najtočnije rezultate predviđanja. U sljedećim potpoglavljima navedeni su glavni doprinosi ovog istraživanja procijeni cijena armature u građevinarstvu.

### 4.1. Mjerni parametri procjene

Ograničen broj istraživanja na tu temu jasno otkriva potrebu za povećanjem uspješnosti u procjeni cijena armature. Prethodna istraživanja ocjenjivala su učinkovitost predviđanja cijena primjenom stopa pogrešaka kao što su MAE, RMSE i MAPE [4, 9, 10, 23]. Uz to uzeta je u obzir jedna ili više stopa pogreške [4, 7, 9, 10]. U regresijskim analizama pomoću AI-ja kao uobičajene metode procjene upotrebljavaju se samo R<sup>2</sup> [8] ili R<sup>2</sup> sa stopama pogreške [23]. U literaturi o procjeni cijene armature nisu pronađena istraživanja koja su procjenjivala različite stope pogreške ili vrijednosti R<sup>2</sup>. Za razliku od prethodnih istraživanja na tu temu, u ovom je istraživanju vrijednost R<sup>2</sup> uzeta u obzir zajedno sa stopom pogreške. Zato su procijenjene i pogreške predviđanja i kompatibilnost procijenjenih vrijednosti sa stvarnim vrijednostima.

### 4.2. Važnost značajke

Shiha i sur. [9] navode da indeks potrošačkih cijena, indeks cijena proizvođača, stopa nezaposlenosti, BDP, devizne rezerve, tečaj američkog dolara i kreditna stopa utječu na procjenu cijene armature. Međutim, nisu naveli parametre koji utječu na iznos utjecaja. Vrijednosti važnosti značajki parametara koji utječu na cijenu armature procijenjene pomoću modela glasačkog



Slika 10. Važnost značajke za cijene armature procijenjene na temelju glasačkog metaansambla

ansambla prikazane su na slici 10. U skladu s time najutjecajnijim parametri bili su tečaj američkog dolara/turske lire, indeks cijena proizvođača i ferolegura. Globalno određivanje cijene željeza i široka upotreba američkog dolara u međunarodnoj trgovini potvrđuju učinkovitost parametra tečaja američkog dolara/turske lire. Osim toga cijene armature iz industrijskog poduzeća koje se bavi proizvodnjom željeza i čelika pod izravnim su utjecajem potrošačkih cijena i cijena ferolegura pa su ti parametri ključni za predviđanje cijena.

### 4.3. Procjena

U prethodnim istraživanjima o procjeni cijene armature dobiveni su zadovoljavajući rezultati za različita razdoblja primjenom različitih modela [9, 10]. Međutim, u ovom su istraživanju uspješni rezultati dobiveni primjenom jednog modela za razdoblja zakašnjenja od jedan, tri, šest, devet i dvanaest mjeseci. Uspješna procjena dobivena pomoću jednog modela doprinosi praktičnoj primjeni i štedi vrijeme. Drugim riječima, iako je primijenjeno osam različitih modela, primjena novoga jedinstvenog modela kombiniranjem tih različitih modela učinkovito je smanjila računsku složenost i skratila vrijeme analize.

### 4.4. Novi model predviđanja

Dok su prethodna istraživanja kombinirala jedan [4, 9, 10], dva [10, 23] ili tri [10] modela procjene, u ovom je istraživanju model predviđanja na temelju glasanja razvijen kombiniranjem osam modela strojnog učenja, od kojih su tri bila osnovna, a pet su bili ansambl. Osam modela strojnog učenja primijenjenih u ovome istraživanju često se upotrebljava u računalnoj znanosti. U ovom su istraživanju ti modeli kombinirani kako bi se stvorio novi model za predviđanje cijena građevnog materijala. Novorazvijeni model predviđanja glasanja postigao je

preciznije rezultate od ostalih modela. Parametri u tome modelu procjene mogu se razlikovati na temelju ažuriranja ili dodavanja/uklanjanja. Također, može se primijeniti i na druge predmete. Usporedba stvarnih i predviđenih vrijednosti za model strojnog učenja na temelju glasačkog metaansambla s tromjesečnim razdobljem zakašnjenja prikazana je na slici 11.

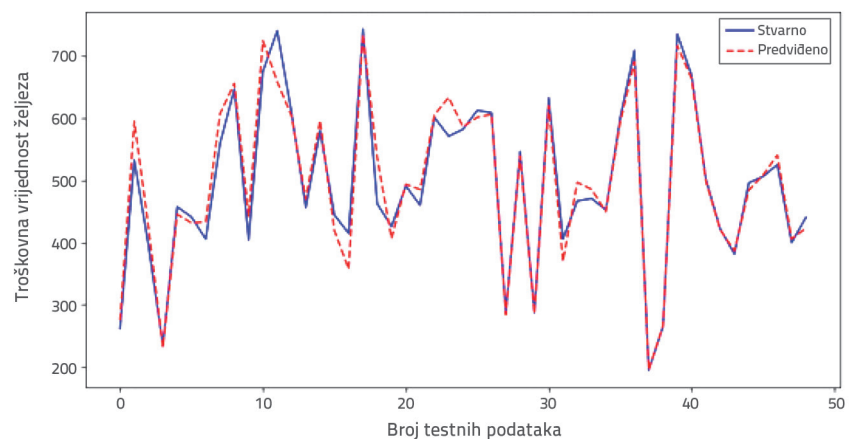
U ovome istraživanju, budući da je model glasačkog metaansambla uspješniji od ostalih modela, mogu ga primijeniti sve

strane u građevinskom sektoru za procjenu cijene armature. Očekuje se da će taj model biti koristan za određivanje vremena s najnižim troškovima za investitore pri određivanju datuma početka ulaganja. Osim toga očekujemo da će pružiti prednost u pogledu troškova pri određivanju vremena nabave armature u građevinskim projektima gdje su ugovori već potpisani i izgradnja je u tijeku. Novorazvijeni model nudi točna predviđanja cijena ne samo za određena razdoblja, već i za razdoblja koja pokrivaju cijelu godinu. U budućnosti se predloženi model može razvijati i upotrebljavati za procjenu cijena drugih građevnih materijala.

## 5. Zaključak

Točne procjene ključne su za određivanje budućih cijena armature u građevinskom sektoru, no istraživanja o toj temi ograničena su, što dodatno povećava njezinu važnost. U skladu s time cilj je ovog istraživanja bio procijeniti buduće cijene armaturnih šipki u građevinskom sektoru primjenom metode utemeljene na umjetnoj inteligenciji. Pri analizi procijenjenih budućih cijena armature u obzir je uzet odnos s promjenama ekonomskih parametara.

U ovome istraživanju strojno učenje primijenjeno je kao metoda umjetne inteligencije, a kao ekonomski parametri određeni su ugljen, plin, željezo, nafta, ferolegura, električna energija,



Slika 11. Usporedba stvarnih i predviđenih vrijednosti prema modelu glasačkog metaansambla tijekom razdoblja zakašnjenja od tri mjeseca



indeks proizvođačkih cijena, tečaj američkog dolara/turske lire i kamatni omjer. Provedena je usporedna analiza modela strojnog učenja, osnovnih modela strojnog učenja (KNN, SVR i Cart) i modela strojnog učenja temeljenih na ansamblu (slučajna šuma, regresija gradijentnim povećanjem, regresija dodatnim stablom, regresija *bagging* stablom, regresija XGB-om). Na temelju rezultata razvijen je model regresije glasačkim metaansamblom, koji je nastao testiranjem različitih varijacija osnovnih modela i modela temeljenih na ansamblu kako bi se povećala točnost procjene.

Za razliku od prethodnih istraživanja, u ovome istraživanju za procjenu cijene armature u razdobljima zakašnjenja od jedan, tri, šest, devet i dvanaest mjeseci postotna stopa pogreške (MAPE) bila je niža od 4 %, a prikladnost modela ( $R^2$ ) veća od 94 % za sva razdoblja. Najbolji rezultati za osnovni model i model strojnog učenja temeljenog na ansamblu dobiveni su u razdoblju zakašnjenja od jednog mjeseca, dok su oni za model strojnog učenja temeljenog na metaansamblu dobiveni u razdoblju zakašnjenja od tri mjeseca. Međutim, kada su uspoređeni osnovni modeli i modeli temeljeni na ansamblu i metaansamblu, tada je model glasačkog ansambla postigao najpreciznije

rezultate predviđanja tijekom razdoblja zakašnjenja od jednog mjeseca. Osim toga glasački model postigao je točnije rezultate tijekom razdoblja zakašnjenja od tri mjeseca u odnosu na jedan mjesec. Zato je model glasačkog metaansambla utvrđen kao općenito uspješniji. Predloženi model pokazuje prihvatljivu točnost predviđanja za ulaganja i odluke o kupnji armature u svim ispitanim razdobljima. Međutim, točnija predviđanja mogu se dobiti u kratkome roku.

Ukratko, analizirani su procijenjeni učinak, učinak ekonomskog parametra i postotak uspješnosti tijekom različitih razdoblja procjene te je predložen novi model procjene cijene. U budućnosti se na temelju ovih rezultata mogu napraviti različite procjene za područja u sektorima koji nisu građevinski dodavanjem ili uklanjanjem novih parametara. Očekujemo da će rezultati ovog istraživanja biti važan doprinos svima u građevinskoj industriji i akademskoj zajednici.

## Zahvala

Autori zahvaljuju tvrtki *SteelData* na pomoći u izradi skupa podataka o cijenama armature.

## LITERATURA

- [1] Yun, S.: Performance Analysis of Construction Cost Prediction Using Neural Network for Multioutput Regression, *Applied Sciences*, 12 (2022) 19, p. 9592.
- [2] Aram, S., Eastman, C., Beetz, J.: Qualitative and Quantitative Cost Estimation: A Methodology Analysis, *American Society of Civil Engineers*, 2014.
- [3] Baalousha, Y., Çelik, T.: An integrated web-based data warehouse and artificial neural networks system for unit price analysis with inflation adjustment, *Journal of Civil Engineering and Management*, 17 (2011) 2, pp. 157-167.
- [4] Mir, M., Kabir, H.D., Nasirzadeh, F., Khosravi, A.: Neural network-based interval forecasting of construction material prices, *Journal of Building Engineering*, 39 (2021) 102288, pp. 1-13.
- [5] Mashhadban, H., Kutanaei, S.S., Sayarinejad, M.A.: Prediction and modeling of mechanical properties in fiber reinforced self-compacting concrete using particle swarm optimization algorithm and artificial neural network, *Construction and Building Materials*, 119 (2016) 2016, pp. 277-287.
- [6] Shi, L., Lin, S., Lu, Y., Ye, L., Zhang, Y.X.: Artificial neural network based mechanical and electrical property prediction of engineered cementitious composites, *Construction and Building Materials*, 174 (2018), pp. 667-674.
- [7] Ou, T.Y., Cheng, C.Y., Chen, P.J., Perng, C.: Dynamic cost forecasting model based on extreme learning machine - A case study in steel plant, *Computers & Industrial Engineering*, 101 (2016), pp. 544-553.
- [8] Elfahham, Y.: Estimation and prediction of construction cost index using neural networks, time series, and regression, *Alexandria Engineering Journal*, 58 (2019) 2, pp. 499-506.
- [9] Shiha, A., Dorra, E.M., Nassar, K.: Neural Networks Model for Prediction of Construction Material Prices in Egypt Using Macroeconomic Indicators, *Journal of Construction Engineering and Management*, 146 (2020) 3, pp. 1-16.
- [10] Faghieh, S.A.M., Kashani, H.: Forecasting Construction Material Prices Using Vector Error Correction Model, *Journal of Construction Engineering and Management*, 144 (2018) 8, p. 4018075.
- [11] Nguyen, H., Vu, T., Vo, T.P., Thai, H.T.: Efficient machine learning models for prediction of concrete strengths, *Construction and Building Materials*, 266 (2021) 120950, pp. 1-17.
- [12] Xu, X., Zhang, Y.: Price forecasts often steel products using Gaussian process regressions, *Engineering Applications of Artificial Intelligence*, 126 (2023) 106870, pp. 1-13.
- [13] Dai, X., Gao, P., Ma, S.: Cost Forecasting for Building Rebar under Uncertainty Conditions: Methodology and Practice, *Buildings*, 14 (2024) 7, p. 1900.
- [14] Mi, J., Xie, X., Luo, Y., Zhang, Q., Wang, J.: Research on Rebar Futures Price Forecast Based on VMD—EEMD—LSTM Model, in Chen, C.H., Scapellato, A., Barbiero, A., Korzun, D.G. (Eds.): *Applied Mathematics, Modeling and Computer Simulation*, IOS Press, 2023.
- [15] Chan, C.H., Sun, M., Huang, B.: Application of machine learning for advanced material prediction and design, *EcoMat*, 4 (2022) 4, 1-28.
- [16] Karatas, I., Budak, A.: Development and comparative of a new meta-ensemble machine learning model in predicting construction labor productivity, *Engineering, Construction and Architectural Management*, 31 (2024) 3, pp. 969-9988.

- [17] Karatas, I., Budak, A.: Prediction of Labor Activity Recognition in Construction with Machine Learning Algorithms, *Icontech International Journal*, 5 (2021) 3, pp. 38-47.
- [18] Üneş, F., Kaya, Y.Z., Mamak, M.: Daily reference evapotranspiration prediction based on climatic conditions applying different data mining techniques and empirical equations, *Theoretical and Applied Climatology*, 141 (2020), 1-2, pp. 763-773.
- [19] Kayadelen, C., Önal, Y., Altay, G., Öztürk, M., Serin, S.: Effects of maintenance, traffic and climate condition on International Roughness Index of flexible pavement, *International Journal of Pavement Engineering* (2022), pp. 1-13.
- [20] TUIK: Turkish Statistical Institute, <https://data.tuik.gov.tr/>, accessed September 20, 2024.
- [21] SteelData: SteelData Informatics and Information Technologies, <https://www.steel-data.com/>, accessed September 20, 2024.
- [22] Jain, S., Shukla, S., Wadhvani, R.: Dynamic selection of normalization techniques using data complexity measures, *Expert Systems with Applications*, 106 (2018), pp. 252-262.
- [23] Farooq, F., Ahmed, W., Akbar, A., Aslam, F., Alyousef, R.: Predictive modeling for sustainable high-performance concrete from industrial wastes: A comparison and optimization of models using ensemble learners, *Journal of Cleaner Production*, 292 (2021), p. 126032.
- [24] Han, J., Micheline, K., Jian, P.: *Data mining concepts and techniques third edition, The Morgan Kaufmann Series in Data Management Systems*, 2011.
- [25] Swetapadma, A., Yadav, A.: A novel single-ended fault location scheme for parallel transmission lines using k-nearest neighbor algorithm, *Computers & Electrical Engineering*, 69 (2018), pp. 41-53.
- [26] Smola, A.J., Schölkopf, B.: A tutorial on support vector regression, *Statistics and Computing*, 14 (2004) 3, pp. 199-222.
- [27] Drucker, H., Burges, C.J.C., Kaufman, L., Smola, A., Vapnik, V.: Support Vector Regression Machines, *Advances in neural information processing systems*, 9 (1997), 155-161.
- [28] Chen, R., Liang, C.Y., Hong, W.C., Gu, D.X.: Forecasting holiday daily tourist flow based on seasonal support vector regression with adaptive genetic algorithm, *Applied Soft Computing*, 26 (2015), pp. 435-443.
- [29] Kovacevic, M., Ivanisevic, N., Petronijević, P., Despotović, V.: Construction cost estimation of reinforced and prestressed concrete bridges using machine learning, *Journal of the Croatian Association of Civil Engineers*, 73 (2021) 01, pp. 1-13.
- [30] James, G., Witten, D., Hastie, T., Tibshirani, R.: *An Introduction to Statistical Learning*, Springer New York, New York, NY, 2013.
- [31] Breiman, L., Friedman, J.H., Stone, C.J., Olshen, R.A.: Classification and Regression Trees, *Biometrics*, 40 (1984) 3, p. 874.
- [32] Breiman and Leo: Random Forests, *Machine Learning*, 45 (2001), pp. 5-32.
- [33] Kayadelen, C., Altay, G., Önal, Y.: Numerical simulation and novel methodology on resilient modulus for traffic loading on road embankment, *International Journal of Pavement Engineering*, 23 (2021) 9, pp. 3212-3221.
- [34] Friedman, J.H.: Greedy function approximation: a gradient boosting machine, *Annals of statistics*, (2001), pp. 1189-1232.
- [35] Gong, M., Bai, Y., Qin, J., Wang, J., Yang, P., Wang, S.: Gradient boosting machine for predicting return temperature of district heating system: A case study for residential buildings in Tianjin, *Journal of Building Engineering*, 27 (2020), p. 100950.
- [36] Geurts, P., Ernst, D., Wehenkel, L.: Extremely randomized trees, *Machine Learning*, 63 (2006) 1, pp. 3-42.
- [37] Ahmad, M.W., Reynolds, J., Rezgui, Y.: Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees, *Journal of Cleaner Production*, 203 (2018), pp. 810-821.
- [38] Breiman, L.: Bagging Predictors, *Machine Learning*, 24 (1996) 2, pp. 123-140.
- [39] Chen, T., He, T.: xgboost: eXtreme Gradient Boosting, R package version 0.4-2, 1(4), 2015, <https://mran.microsoft.com/web/packages/xgboost/vignettes/xgboost.pdf>, accessed April 26, 2021.
- [40] Chen, T., Guestrin, C. (Eds.): *XGBoost: A Scalable Tree Boosting System*, 2016.
- [41] Alshboul, O., Shehadeh, A., Almasabha, G., Almuflih, A.S.: Extreme Gradient Boosting-Based Machine Learning Approach for Green Building Cost Prediction, *Sustainability*, 14 (2022) 11, p. 6651.
- [42] Géron, A.: *Hands-on machine learning with Scikit-Learn and TensorFlow: Concepts, tools, and techniques to build intelligent systems / Aurélien Géron*, First edition, O'Reilly, Beijing, 2017.
- [43] Stow, C.A., Jolliff, J., McGillicuddy, D.J., Doney, S.C., Allen, J.I., Friedrichs, M.A.M., et al.: Skill Assessment for Coupled Biological/Physical Models of Marine Systems, *Journal of Marine Systems*, 76 (2009), 1-2, pp. 4-15.
- [44] Zien, A., Kraemer, N., Sonnenburg, S., Raetsch, G. (Eds.): *The Feature Importance Ranking Measure*, Springer, Berlin Heidelberg, 2009.
- [45] Kayadelen, C., Önal, Y., Altay, G., Öztürk, M., Serin, S.: Effects of maintenance, traffic and climate condition on International Roughness Index of flexible pavement, *International Journal of Pavement Engineering*, 24 (2023) 1.