

Predviđanje kočene snage i brzine vrtnje motora brodova za rasuti teret primjenom metoda strojnog učenja

Ivana Martić¹, Nastia Degiuli^{1*}, Marko Valčić^{2,3}, Carlo Giorgio Grlj¹

¹Sveučilište u Zagrebu Fakultet strojarstva i brodogradnje, Ivana Lučića 5, 10000 Zagreb

²Sveučilište u Zadru, Pomorski odjel, Ulica Mihovila Pavlinovića 1, 23000 Zadar

³Sveučilište u Rijeci, Tehnički fakultet, Vukovarska ulica 58, 51000 Rijeka

*nastia.degiuli@fsb.unizg.hr

Sažetak: U ovom istraživanju primijenjen je niz metoda strojnog učenja za razvoj modela za predviđanje kočene snage motora i brzine vrtnje motora brodova za rasuti teret. Poseban naglasak stavljen je na smanjenje šuma u ulaznim i izlaznim podacima primjenom metode zaglađivanja splajnom, s ciljem poboljšanja sposobnosti generalizacije razvijenih modela strojnog učenja. Provedena je detaljna usporedba rezultata dobivenih primjenom različitih metoda, pri čemu su ocijenjene njihova točnost i sposobnost generalizacije. Rezultati dobiveni na temelju zaglađenih podataka pokazuju da modeli ansambla regresijskih stabala ostvaruju uravnotežen kompromis između interpretabilnosti i točnosti, što ih čini posebno prikladnima za praktične inženjerske primjene. U učinkovitosti ih slijede optimizirani modeli regresijskih stabala te regresije Gaussovom procesom. Nadalje, modeli linearne regresije temeljeni na zaglađenim podacima također postižu zadovoljavajuću razinu točnosti za praktične primjene, što omogućuje razvoj jednostavnih prediktivnih formula za kočenu snagu i brzinu vrtnje motora, prikladnih za primjenu u okviru preliminarnog projektiranja broda.

Cljučne riječi: strojno učenje, projektiranje broda, brod za rasuti teret, kočena snaga, brzina vrtnje motora

1. Uvod

Preliminarna faza projektiranja broda, između ostalog, uključuje određivanje glavnih dimenzija i oblika trupa te odabir porivnog sustava. Procjena kočene snage i brzine vrtnje motora u preliminarnom projektiranju broda ključna je za odabir najpovoljnije projektne varijante s obzirom na energetske učinkovitost broda. U preliminarnom

projektiranju broda njegove se značajke u pravilu određuju primjenom regresijskih formula izvedenih na temelju postojećih baza podataka brodova, koje su često zastarjele i neprikladne za suvremene brodske forme. Takvi modeli su nerijetko razvijeni na temelju sirovih i zašumljenih podataka te se u tom kontekstu linearna regresija može promatrati kao niskopropusni filtar. Pouzdanost i točnost modela razlikuju se ovisno o tome primjenjuje li se model na sirove i često zašumljene podatke ili na prethodno zaglađene podatke.

Značajni naponi usmjereni su na unapređenje postojećih regresijskih formula, uključujući integraciju novo prikupljenih podataka te objedinjavanje više empirijskih modela [1]. Primjerice, Rinauro et al. [2] primijenili su višestruku regresiju u kombinaciji s algoritmom stabla odluke (engl. *forest trees*) kako bi razvili prediktivni model za procjenu glavnih dimenzija kontejnerskih brodova, brzine i kočene snage u fazi preliminarnog projektiranja. Osim toga, predloženi su relativno jednostavni modeli temeljeni na umjetnim neuronskim mrežama (engl. *artificial neural network*, ANN) za procjenu dodatnog otpora broda na pravilnim valovima [3–6]. Modeli strojnog učenja (engl. *machine learning*) sve se češće primjenjuju za rješavanje različitih problema u području brodogradnje. Neka od provedenih istraživanja uključuju između ostalog predviđanje snage glavnog pogonskog motora i emisija štetnih plinova [7], procjenu potrošnje goriva broda na temelju podataka prikupljenih tijekom plovidbe [8], predviđanje putanje broda [9], analizu učinaka obraštanja trupa [10], određivanje hidrodinamičkih značajki putem višeciljne optimizacije [11], predviđanje opterećenja koja djeluju na brodski trup uslijed plovidbe ledom [12], procjenu rizika od nasukavanja [13] te kvantifikaciju relativne vjerojatnosti incidenta u ekstremnim vremenskim uvjetima [14]. Mogućnost primjene strojnog učenja za predviđanje brzine broda primjenom metoda regresije, istražena je na temelju operativnih podataka trajekta prikupljenih tijekom plovidbe [15]. Korištene regresijske metode uključuju linearnu regresiju, regresijska stabla različitih veličina, ansamble regresijskih stabala, regresiju Gaussovima procesom te stroj s potpornim vektorima (engl. *support vector machines*). Višestruka linearna regresija, kao znatno jednostavniji algoritam, pokazala je usporedivu razinu točnosti u odnosu na složenije metode.

Za početak je potrebno identificirati relevantne podatke, provjeriti njihovu valjanost te ih statistički obraditi kako bi se osiguralo formiranje visokokvalitetnog skupa podataka. Taj postupak uključuje pažljivu analizu podataka radi uklanjanja anomalija i izdvojenih vrijednosti, uz istodobno razmatranje mogućeg utjecaja zaglađivanja na cjelovitost podataka. Skupovi sirovih podataka često sadrže šum i nedostajuće vrijednosti, što može znatno narušiti točnost modela strojnog učenja treniranih na takvim podacima [8]. Prediktivna učinkovitost bilo kojeg modela strojnog učenja uvelike ovisi o kvaliteti i količini ulaznih podataka. U mnogim slučajevima podaci pokazuju znatnu varijabilnost i nesigurnost, čemu je potrebno posvetiti dodatnu pažnju prilikom formiranja skupa podataka za učenje modela strojnog učenja. Liu i suradnici [16] primijenili su tzv. metodu teorije valića (engl. *wavelet*) kako bi uklonili šum iz

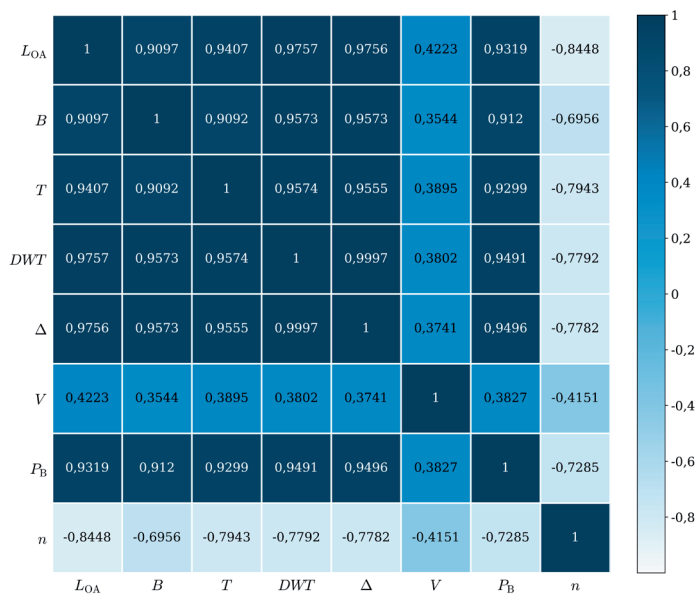
eksperimentalnih podataka gibanja KVLCC2 broda i autonomnog površinskog plovila. Slično tome, Xue i suradnici [17] koristili su Gaussov proces za uklanjanje šuma iz podataka o gibanjima broda. Nadalje, algoritam zaglađivanja podataka primijenjen je za uklanjanje šuma podataka korištenih pri razvoju ANN modela za predviđanje potrošnje goriva broda [18]. Metoda analize varijance (ANOVA) i splajnovi za zaglađivanje podataka primijenjeni su za smanjenje broja ulaznih varijabli u optimizacijskom postupku minimizacije potrošnje goriva za tipično plutajuće postrojenje za proizvodnju, skladištenje i iskrcaj (engl. *Floating Production Storage and Offloading*, FPSO) u promjenjivim radnim uvjetima [19].

U ovom radu provedena je sveobuhvatna analiza i usporedba rezultata dobivenih pomoću 18 različitih modela temeljenih na pet metoda strojnog učenja za predviđanje kočene snage i brzine vrtnje motora brodova za rasuti teret. Sirovi podaci često su značajno zašumljeni, što uvodi nesigurnost u razvoj modela. Stoga je naglašena važnost obrade zašumljenih podataka s visokom razinom nesigurnosti kako bi se razvili što pouzdaniji modeli primjenjivi na potpuno nove skupove podataka. Predstavljen je postupak pripreme podataka zaglađivanjem splajnovima s ciljem postizanja što bolje sposobnosti generalizacije, a koji se primjenjuje ne samo na složene modele strojnog učenja, već i na višestruku linearnu regresiju. Iako razvijeni modeli postižu visoku točnost u procjeni kočene snage i brzine vrtnje motora, njihova složenost onemogućuje izravnu primjenu u fazi preliminarnog projektiranja broda. Stoga su predloženi jednostavni regresijski izrazi koji omogućuju procjenu kočene snage i brzine vrtnje motora na temelju dimenzija broda i brzine plovidbe uz zadovoljavajuću točnost.

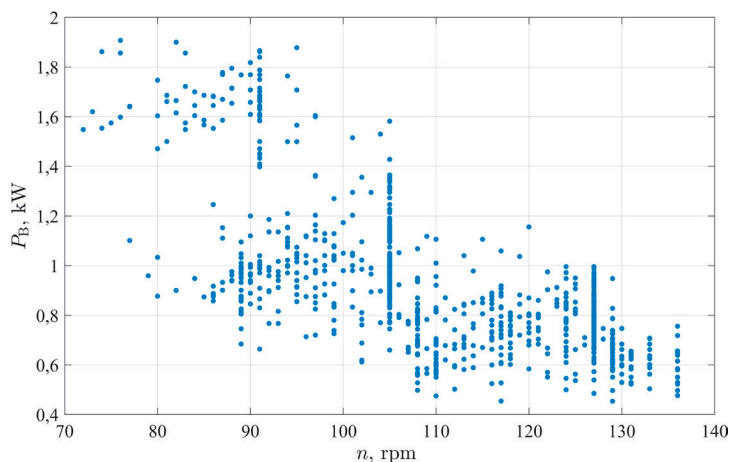
2. Baza podataka

S ciljem razvoja modela strojnog učenja korištena je IHS Fairplay [20] baza podataka brodova za rasuti teret korištena od strane MEPC-a (engl. *Marine Environment Protection Committee*) kao standardna referentna baza za određivanje referentne linije projektnog indeksa energetske učinkovitosti (engl. *Energy Efficiency Design Indeks*, EEDI). Dostupni podaci obuhvaćaju naziv broda, duljinu preko svega (L_{OA}), širinu (B), gaz (T), nosivost (DWT), istisninu (Δ), brzinu (V), kočenu snagu glavnog motora (P_B) i brzina vrtnje motora (n). Baza podataka obuhvaća brodove izgrađene u razdoblju od 2000. do 2015. godine. Nakon uklanjanja podataka koji značajnije odstupaju od skupa (eng. *outliers*), od ukupno 8253 podatka dobiveno je 6660 podataka koji su korišteni za učenje, validaciju i testiranje modela strojnog učenja. Podaci o istisnini bili su dostupni za samo 24,3 % uzoraka, stoga je ista izostavljena iz daljnje analize. Uočena je snažna pozitivna korelacija između zavisne varijable P_B i prediktorskih varijabli L_{OA} , B , T , DWT i Δ , Slika 1. Međutim, među prediktorskim varijablama prisutna je i izražena multikolinearnost, što je očekivano s obzirom na njihove fizikalne međuodnose. Kako bi se ublažio negativan utjecaj multikolinearnosti, nosivost, koja

pokazuje visoku razinu multikolinearnosti s varijablama L_{OA} , B , T , isključena je iz daljnjeg modeliranja. Nasuprot tome, zavisna varijabla n pokazuje negativnu korelaciju s L_{OA} , B , T i V , pri čemu je ta korelacija nešto manje izražena u odnosu na onu uočenu za P_B . Konačno, za procjenu kočene snage i brzine vrtnje motora brodova za rasuti teret usvojene su četiri ulazne varijable: L_{OA} , B , T i V . Distribucija podataka za P_B i n prikazana je na Slici 2.



Slika 1: Korelacijska matrica



Slika 2: Distribucija podataka za P_B i n

Statistički pokazatelji ulaznih i izlaznih varijabli prikazani su u Tablici 1. Uočena je značajna varijabilnost kočene snage i brzine vrtnje motora, što upućuje na visoku razinu šuma u ulaznim podacima. Također je primijećeno da pojedini sestrinski brodovi imaju znatno različite vrijednosti kočene snage i brzine vrtnje motora. Kako bi se smanjio šum u podacima i poboljšala sposobnost generalizacije modela, postupak zaglađivanja podataka splajnovima proveden je u okviru pripreme podataka. Na taj način je smanjena nesigurnosti u podacima te značajno povećana pouzdanost i robusnost razvijenih modela.

Tablica 1: Statistički pokazatelji uzlaznih i izlaznih varijabli

	Ulazne varijable				Izlazne varijable	
	L_{OA} , m	B , m	T , m	V , kn	P_B , kW	n , o/min
Minimum	157,00	23,10	8,70	10,00	4539,00	72,00
Aritmetička sredina	215,31	34,18	13,37	14,34	9787,24	111,34
Medijan	197,00	32,30	12,80	14,50	8753,00	110,00
Mod	190,00	32,30	12,80	14,50	9480,00	127,00
Maksimum	330,00	57,00	21,40	17,00	19073,00	136,00
Standardna devijacija	39,65	6,04	2,61	0,39	3384,26	15,91

3. Metodologija

U okviru istraživanja primijenjene su različite metode strojnog učenja poput linearne regresije, regresijskih stabala, regresije Gausovim procesom, ansambla regresijskih stabala te neuronske mreže. Široko primjenjiva linearna regresija koristi se za opisivanje i analizu odnosa između zavisne varijable i jedne ili više prediktorskih varijabli. Cilj je pronaći linearnu funkciju koja najbolje opisuje podatke, a dobiveni model omogućuje jednostavnu interpretaciju [21].

Regresijska stabla predstavljaju jednostavnu i visoko interpretabilnu metodu nadziranog učenja namijenjenu rješavanju regresijskih problema s više prediktorskih varijabli. Njihova jednostavnost i učinkovitost, zajedno s obilježjima koja ih čine pogodnima za rješavanje različitih multivarijabilnih regresijskih problema, čine ih atraktivnim alatom u praksi [22].

Modeli regresije Gausovim procesom predstavljaju nelinearnu, probabilističku metodu regresije koja se temelji na teoriji vjerojatnosti [23]. Umjesto da unaprijed definira oblik funkcije, Gaussov proces opisuje distribuciju mogućih funkcija koje su u skladu s podacima. Zbog svoje fleksibilnosti i sposobnosti modeliranja složenih

odnosa, često se koriste u strojnome učenju te su poznati po visokoj točnosti. Međutim, njihova složenost može otežati interpretaciju rezultata. Važno je napomenuti kako izlaz nije pojedinačna vrijednost, već distribucija vjerojatnosti, što omogućuje kvantifikaciju nesigurnosti u predviđanjima.

Ansambli regresijskih stabala obuhvaćaju skup algoritama strojnog učenja koji se temelje na kombiniranju više stabala odlučivanja kako bi se postigla veća točnost i robusnost modela. Umjesto oslanjanja na jedno stablo, ansambli agregiraju predviđanja više stabala, čime se smanjuje pretreniranost modela. Općenito se dijele u dvije glavne kategorije: *boosted* stabla, koja se temelje na slijednom učenju algoritama na pogreškama prethodnih algoritama, te *bagged* stabla, koja koriste pristup treniranja na podskupovima skupa za učenje [24,25].

Neuronske mreže predstavljaju napredne algoritme učenja čiji postupak treniranja uključuje podešavanje težinskih faktora i pomaka korištenjem odgovarajućeg algoritma učenja. U nadziranom učenju ulazni podaci su dostupni mreži, a izlazne vrijednosti se uspoređuju s ciljnim vrijednostima kako bi se smanjila pogreška i poboljšala točnost predikcije [26].

Svaka metoda ima svoje specifične značajke, poput početnih uvjeta i topologije modela. Kako bi se uzeli u obzir ti specifični aspekti, u ovom su istraživanju za svaku metodu razvijeni višestruki modeli.

3.1. Zaglađivanje podataka splajnom

Kao što je prethodno navedeno, sirovi podaci pripremljeni su zaglađivanjem splajnom $s(x_i)$ temeljenom na parametru zaglađivanja p i težinama w_i . Zaglađivanje podataka splajnom minimizira sljedeći član:

$$p \sum_i w_i (y_i - s(x_i))^2 + (1-p) \int (d^2 s / dx^2)^2 dx \quad (1)$$

Parametar zaglađivanja odabran je tako da bude što bliži nuli kako bi se postiglo glatko prilagođavanje krivulje metodom najmanjih kvadrata:

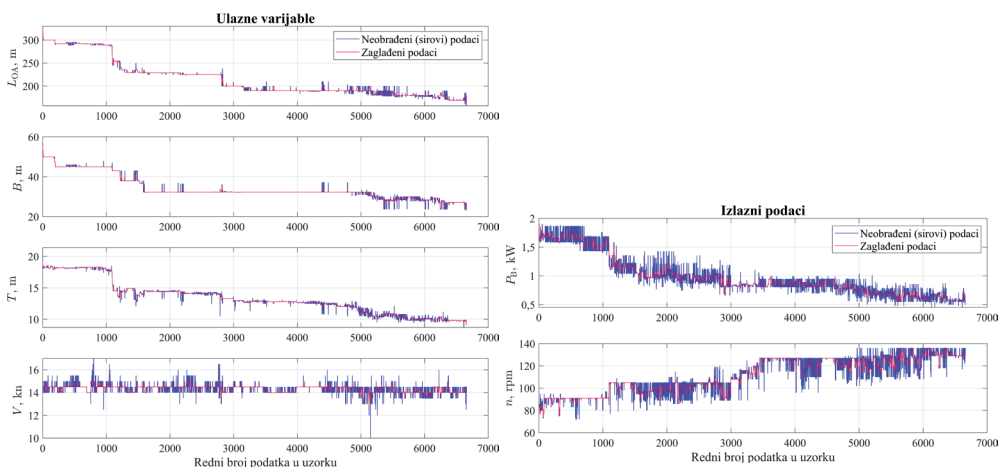
$$\begin{cases} \min_{p_j} (R_{j,\text{test}}^2 - R_{j,\text{val}}^2) \\ \text{s.t. } 0 \leq p_j \leq 1, \quad j = 1, 2, \dots, N \end{cases} \quad (2)$$

gdje R_{test}^2 i R_{val}^2 predstavljaju koeficijente determinacije za testni i validacijski skup podataka, a N broj analiziranih modela strojnog učenja.

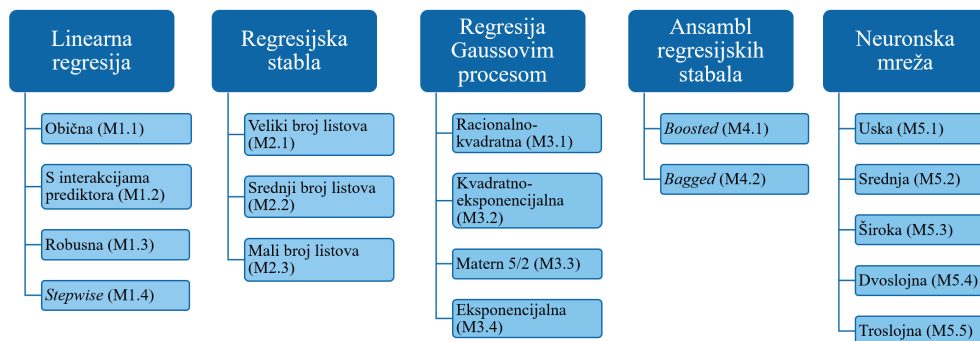
Cjelokupni postupak, koji uključuje zaglađivanje podataka i treniranje modela strojnog učenja započinje korakom inicijalizacije, u kojem se definiraju granice parametra zaglađivanja, zajedno s veličinom koraka za iterativni proces. Osim toga, za k -struku unakrsnu validaciju potrebno je odrediti broj dijelova skupa podataka (k) te postotak podataka namijenjenih za testiranje. Potom se odabrani trenutačni parametar zaglađivanja primjenjuje za zaglađivanje izvornog skupa podataka u tekućoj iteraciji te se definiraju ulazni i ciljani podaci koristeći zaglađeni skup podataka, zajedno s njihovom raspodjelom prema principu k -struke unakrsne validacije. Unutar istog koraka provodi se treniranje i validacija na temelju zaglađenog skupa podataka. Pokazatelji učinkovitosti računaju se na temelju trenutačnog parametra zaglađivanja, dok se prosječni pokazatelj validacije računa za sve dijelove skupa podataka k i pohranjuje za daljnju analizu. Validirani modeli strojnog učenja testiraju se na preostalom dijelu nasumično odabranih podataka iz zaglađenog skupa. Navedeni postupak ponavlja se iterativno za svaku vrijednost parametra zaglađivanja, dok se u završnom koraku odabire vrijednost parametra zaglađivanja koja minimizira funkciju cilja.

4. Rezultati

Kako bi se osigurala objektivna usporedba rezultata razvijenih modela, ključno je dosljedno koristiti isti skup podataka tijekom faza treniranja, validacije i testiranja. Stoga je za sve analizirane modele strojnog učenja usvojena srednja vrijednost parametra zaglađivanja. Zaglađeni podaci, generirani iz izvornog skupa podataka, prikazani su za ulazne i izlazne varijable na Slici 3. Pet metoda strojnog učenja analizirano je kroz ukupno 18 modela s različitim konfiguracijama, kako je prikazano na Slici 4. Svakom modelu dodijeljena je jedinstvena oznaka radi lakšeg praćenja i usporedbe rezultata.



Slika 3: Sirovi i zaglađeni podaci za ulazne i izlazne varijable



Slika 4: Modeli strojnog učenja

Kod *stepwise* linearne regresije početni članovi prediktorskih varijabli su linearni, uz interakcije prediktora te ograničenje na broj članova koje model može uključiti u maksimalno 1000 koraka. Regresijska stabla primijenjena su bez zamjenskih podjela odluka. Kod velikog broja listova najmanja veličina lista iznosi 4, kod srednjeg 12, a kod malog broja listova 36. U regresiji Gausovim procesom primijenjena je konstantna bazna funkcija i izotropna jezgrena funkcija, uz automatsko određivanje standardne devijacije signala te sigma parametra. Za ansamble regresijskih stabala minimalna veličina lista postavljena je na osam, s ukupno 30 učenika. Kod neuronskih mreža primijenjena je ReLU aktivacijska funkcija, a postupak treniranja ograničen je na 1000 iteracija. Uska, dvoslojna i troslojna neuronska mreža sadrže po 10 neurona u skrivenom sloju, dok srednja i široka mreža imaju 25, odnosno 100 neurona u skrivenom sloju.

Skup podataka podijeljen je na podskupove za treniranje, validaciju i testiranje. Za validaciju je korištena k -struka unakrsna validacija s pet podjela. Poseban testni podskup, koji obuhvaća 10 % ukupnog broja podataka, rezerviran je za procjenu učinkovitosti svih modela, dok su preostali podaci podijeljeni u pet jednakih dijelova za potrebe unakrsne validacije, gdje je svaki dio obuhvatio 1199 brodova, odnosno 18 % ukupnog broja podataka. Treniranje je u svakoj iteraciji provedeno na temelju podataka u četiri dijela skupa podataka, odnosno za 4795 brodova ili 72 % ukupnog broja podataka. Učinkovitost modela ocijenjena je na temelju koeficijenta determinacije R^2 , korijena srednje kvadratne pogreške *RMSE* (engl. *Root Mean Square Error*), srednje kvadratne pogreške *MSE* (engl. *Mean Square Error*) i srednje apsolutne pogreške *MAE* (engl. *Mean Absolute Error*), koji su izračunati kao prosjek vrijednosti preko svih pet skupova podataka za validaciju. Korišten je programski paket MATLAB R2022b [27], a analiza je provedena na sirovim i zaglađenim skupovima podataka. Usporedbom rezultata pokazalo se kako je postupak zaglađivanja podataka značajno poboljšao kvalitetu podataka, što je dovelo do povećane točnosti i bolje sposobnosti generalizacije [28].

Tablica 2: Pokazatelji učinkovitosti modela strojnog učenja u predviđanju kočene snage motora P_B

Model strojnog učenja		<i>RMSE</i>	<i>R</i> ²	<i>MSE</i>	<i>MAE</i>
M1.1	validacija	576,21	0,9694	332018,61	438,25
	testiranje	577,28	0,9675	333252,15	440,02
M1.2	validacija	536,11	0,9735	287413,95	405,93
	testiranje	537,92	0,9718	289356,79	404,93
M1.3	validacija	576,70	0,9694	332578,93	437,65
	testiranje	577,11	0,9675	333053,72	438,83
M1.4	validacija	536,11	0,9735	287413,95	405,93
	testiranje	537,92	0,9718	289356,79	404,93
M2.1	validacija	308,77	0,9912	95338,18	185,36
	testiranje	308,14	0,9907	94950,49	174,53
M2.2	validacija	352,31	0,9886	124122,02	236,88
	testiranje	334,22	0,9891	111704,40	219,53
M2.3	validacija	417,52	0,9839	174324,78	291,20
	testiranje	407,25	0,9838	165852,54	287,52
M3.1	validacija	364,91	0,9877	133159,05	245,29
	testiranje	322,45	0,9899	103974,26	212,53
M3.2	validacija	427,20	0,9832	182503,99	302,85
	testiranje	421,61	0,9827	177751,39	300,51
M3.3	validacija	411,72	0,9844	169511,17	289,31
	testiranje	408,27	0,9838	166685,83	287,50
M3.4	validacija	346,70	0,9889	120201,68	231,78
	testiranje	326,19	0,9896	106398,34	217,03
M4.1	validacija	604,42	0,9663	365317,89	477,73
	testiranje	598,56	0,9651	358277,24	474,02
M4.2	validacija	276,44	0,993	76416,40	185,70
	testiranje	243,53	0,9942	59304,92	170,96
M5.1	validacija	505,35	0,9765	255376,82	369,25
	testiranje	530,13	0,9726	281039,21	387,72
M5.2	validacija	491,35	0,9778	241421,89	357,20
	testiranje	492,82	0,9763	242870,22	355,44
M5.3	validacija	448,84	0,9814	201452,99	324,68
	testiranje	450,10	0,9803	202592,71	322,14
M5.4	validacija	507,12	0,9763	257172,18	373,21
	testiranje	504,12	0,9752	254137,13	371,41
M5.5	validacija	499,15	0,977	249152,02	365,71
	testiranje	717,92	0,9498	515407,28	548,13

Pokazatelji učinkovitosti razvijenih modela strojnog učenja za procjenu P_B i n prikazani su u Tablici 2 odnosno Tablici 3. Ansambl regresijskih stabala, tj. *bagged* stabla pokazuju najveću učinkovitost u predviđanju P_B i za validacijski i za testni skup podataka, a slijede ih stabla s velikim brojem listova. Najmanju učinkovitost pokazuju tzv. *boosted* stabla za validacijski skup podataka, odnosno troslojna neuronska mreža za testni skup podataka.

Tablica 3: Pokazatelji učinkovitosti modela strojnog učenja u predviđanju brzine vrtnje motora n

Model strojnog učenja		<i>RMSE</i>	R^2	<i>MSE</i>	<i>MAE</i>
M1.1	validacija	6,18	0,8305	38,21	4,93
	testiranje	6,29	0,8365	39,52	5,02
M1.2	validacija	5,47	0,8675	29,88	4,07
	testiranje	5,41	0,879	29,25	4,01
M1.3	validacija	6,22	0,8283	38,71	4,85
	testiranje	6,32	0,8346	39,98	4,93
M1.4	validacija	5,46	0,8675	29,86	4,07
	testiranje	5,41	0,8789	29,28	4,01
M2.1	validacija	2,42	0,9741	5,85	1,28
	testiranje	2,03	0,9829	4,12	1,08
M2.2	validacija	2,75	0,9665	7,56	1,69
	testiranje	2,43	0,9756	5,90	1,44
M2.3	validacija	3,21	0,9542	10,33	2,18
	testiranje	2,95	0,9639	8,72	2,01
M3.1	validacija	2,57	0,9707	6,61	1,63
	testiranje	2,27	0,9786	5,17	1,46
M3.2	validacija	3,07	0,9583	9,40	2,12
	testiranje	3,10	0,9603	9,60	2,17
M3.3	validacija	2,80	0,9652	7,85	1,88
	testiranje	2,64	0,9711	6,99	1,79
M3.4	validacija	2,50	0,9723	6,24	1,57
	testiranje	2,32	0,9778	5,38	1,49
M4.1	validacija	5,89	0,846	34,71	5,21
	testiranje	5,74	0,8635	32,99	5,08
M4.2	validacija	2,16	0,9792	4,68	1,35
	testiranje	1,86	0,9856	3,48	1,19
M5.1	validacija	4,54	0,9084	20,65	3,38
	testiranje	4,30	0,9236	18,46	3,17
M5.2	validacija	3,80	0,936	14,43	2,80
	testiranje	3,62	0,9458	13,10	2,62
M5.3	validacija	3,30	0,9516	10,91	2,36
	testiranje	3,35	0,9536	11,22	2,42
M5.4	validacija	3,94	0,9312	15,51	2,87
	testiranje	3,67	0,9444	13,45	2,63
M5.5	validacija	3,79	0,9363	14,37	2,78
	testiranje	3,49	0,9495	12,20	2,49

U slučaju brzine vrtnje motora, vidljivo je da *bagged* stabla nadmašuju preostale modele kako za validacijski, tako i za testni skup podataka, a slijede ih modeli eksponencijalne regresije Gausovim procesom i stabla sa srednjim brojem listova. Općenito, modeli regresijskih stabala te regresije Gausovim procesom pokazuju najveću učinkovitost u predviđanju brzine vrtnje motora n , a slijedi ih široka neuronska mreža. Ostali modeli, osobito linearna regresija i *boosted* stabla, imaju najniži koeficijent determinacije te najveće vrijednosti *MSE* i *RMSE*. S druge strane, ako se za pokazatelj učinkovitosti uzme *MAE*, stabla s velikim brojem listova pokazuju najnižu vrijednost za oba skupa podataka.

Za potrebe inženjerske primjene obična linearna regresija može biti zadovoljavajuća. Izraz za procjenu P_B i n kao funkcije brzine broda i glavnih dimenzija, koji se može primijeniti u preliminarnom projektiranju broda, glasi:

$$\hat{P}_B = -10617,7260 + 28,2814 \cdot L_{OA} + 195,9338 \cdot B + 381,9788 \cdot T + 175,3274 \cdot V \quad (3)$$

$$\hat{n} = 222,0986 - 0,3412 \cdot L_{OA} + 1,3351 \cdot B - 2,7910 \cdot T - 3,1832 \cdot V \quad (4)$$

Treba napomenuti da je točnost u ovom slučaju osjetno manja u usporedbi s prethodno spomenutim složenijim modelima strojnog učenja, što je potrebno uzeti u obzir pri praktičnoj primjeni ovih jednadžbi. Nešto veća točnost može se postići primjenom *stepwise* linearne regresije. Jednadžbe za procjenu P_B i n u tom su slučaju sljedeće:

$$\begin{aligned} \hat{P}_B = & 21017,6420 - 257,9705 \cdot L_{OA} - 1298,1847 \cdot B + 5988,9754 \cdot T \\ & - 1737,4671 \cdot V - 2,4606 \cdot L_{OA} \cdot B + 1,3145 \cdot L_{OA} \cdot T + 23,7887 \cdot L_{OA} \cdot V \\ & + 39,3344 \cdot B \cdot T + 104,2699 \cdot B \cdot V - 494,5938 \cdot T \cdot V \end{aligned} \quad (5)$$

$$\begin{aligned} \hat{n} = & 478,7795 - 8,1115 \cdot L_{OA} + 2,3376 \cdot B + 91,1797 \cdot T - 13,1283 \cdot V \\ & - 0,0064 \cdot L_{OA} \cdot B + 0,0496 \cdot L_{OA} \cdot T + 0,5013 \cdot L_{OA} \cdot V - 7,1916 \cdot T \cdot V \end{aligned} \quad (6)$$

5. Zaključak

U ovom istraživanju provedena je usporedba rezultata pet metoda strojnog učenja za procjenu kočene snage motora i brzine vrtnje motora brodova za rasuti teret. Razvijeno je 18 različitih modela strojnog učenja: četiri modela linearne regresije, tri modela regresijskih stabala, četiri modela regresije Gausovim procesom, dva modela ansambla regresijskih stabala te pet modela neuronskih mreža. Kako bi se ublažio utjecaj šuma u podacima i poboljšala sposobnost generalizacije modela, u okviru

pripreme podataka primijenjena je metoda zaglađivanja splajnom, čime je učinkovito smanjena nesigurnost u izvornom skupu podataka te značajno povećana pouzdanost i robusnost razvijenih modela.

Usporedba rezultata pokazala je da modeli ansambla regresijskih stabala ostvaruju uravnotežen kompromis između interpretabilnosti i točnosti, što ih čini posebno prikladnima za praktične inženjerske primjene. U učinkovitosti ih slijede optimizirani modeli regresijskih stabala te regresije Gaussovima procesom. Iako su napredniji modeli strojnog učenja očekivano postigli višu razinu učinkovitosti, modeli linearne regresije također su pokazali zadovoljavajuću točnost, čime je potvrđena njihova praktična primjenjivost u preliminarnom projektiranju broda, gdje su jednostavnost i interpretabilnost od ključne važnosti. Kao rezultat, predložene su regresijske formule za procjenu kočene snage i brzine vrtnje motora na temelju glavnih značajki broda i brzine plovidbe. Ove jednostavne formule predstavljaju učinkovite alate za brzu procjenu te doprinose kvalitetnijem donošenju odluka u ranim fazama projektiranja broda.

Zahvala

Ova rad je financirala Hrvatska zaklada za znanost projektima IP-2025-02-4779 i IP-2022-10-2821.

6. Literatura

- [1] Zhang, C.; Ringsberg, J. W., Thies, F.: Development of a ship performance model for power estimation of inland waterway vessels, *Ocean Engineering*, **287** (2023) 115731, <https://doi.org/10.1016/j.oceaneng.2023.115731>
- [2] Rinauro, B.; Begovic, E., Mauro, F., Rosano, G.: Regression analysis for container ships in the early design stage, *Ocean Engineering*, **292** (2024) 116499, <https://doi.org/10.1016/j.oceaneng.2023.116499>
- [3] Martić, I.; Degiuli, N., Grlj, C. G.: Prediction of added resistance of container ships in regular head waves using an artificial neural network, *Journal of Marine Science and Engineering*, **11** (2023) 7, 1293, <https://doi.org/10.3390/jmse11071293>
- [4] Martić, I.; Degiuli, N., Farkas, A., Grlj, C. G.: The application of ANN in estimating added resistance of container ships in regular head waves, *Sustainable Development and Innovations in Marine Technologies - Proceedings of the 19th International Congress of the International Maritime Association of the Mediterranean*, IMAM 2022, 26-29 September, Istanbul, Turkey, 175–182 (2023), <https://doi.org/10.1201/9781003358961-23>
- [5] Martić, I.; Degiuli, N., Majetić, D., Farkas, A.: Artificial neural network model for the evaluation of added resistance of container ships in head waves, *Journal of Marine Science and Engineering*, **9** (2021) 8, 826, <https://doi.org/10.3390/jmse9080826>

- [6] Cepowski, T.: The prediction of ship added resistance at the preliminary design stage by the use of an artificial neural network, *Ocean Engineering*, **195** (2020), 106657, <https://doi.org/10.1016/j.oceaneng.2019.106657>
- [7] Ozsari, I.: Predicting main engine power and emissions for container, cargo, and tanker ships with artificial neural network analysis, *Brodogradnja*, **74** (2023) 2, 77–94, <https://doi.org/10.21278/brod74204>
- [8] Fan, A.; Wang, Y., Yang, L., Tu, X., Yang, J., Shu, Y.: Comprehensive evaluation of machine learning models for predicting ship energy consumption based on onboard sensor data, *Ocean & Coastal Management*, **248** (2024) 106946, <https://doi.org/10.1016/j.ocecoaman.2023.106946>
- [9] Li, H.; Jiao, H., Yang, Z.: Ship trajectory prediction based on machine learning and deep learning: A systematic review and methods analysis, *Engineering Applications of Artificial Intelligence*, **126** (2023) 107062, <https://doi.org/10.1016/j.engappai.2023.107062>
- [10] Mittendorf, M.; Nielsen, U. D., Bingham, H. B.: Capturing the effect of biofouling on ships by incremental machine learning, *Applied Ocean Research*, **138** (2023) 103619 <https://doi.org/10.1016/j.apor.2023.103619>
- [11] Liu, J.; Wu, S., Yue, X., Yue, Q.: Hydrodynamic shape optimization of an autonomous and remotely-operated vehicle via a multi-surrogate model, *Brodogradnja*, **75** (2024) 3, 75301, <https://doi.org/10.21278/brod75301>
- [12] Lee, S. J.; Jung, K. H., Ku, N., Lee, J.: A comparison of regression models for the ice loads measured during the ice tank test, *Brodogradnja*, **74** (2023) 3, 1–15, <https://doi.org/10.21278/brod74301>
- [13] Zhang, M.; Kujala, P., Hirdaris, S.: A machine learning method for the evaluation of ship grounding risk in real operational conditions, *Reliability Engineering & System Safety*, **226** (2022) 108697, <https://doi.org/10.1016/j.res.2022.108697>
- [14] Rawson, A.; Brito, M., Sabeur, Z., Tran-Thanh, L.: A machine learning approach for monitoring ship safety in extreme weather events, *Safety Science*, **141** (2021) 105336, <https://doi.org/10.1016/j.ssci.2021.105336>
- [15] Bassam, A. M.; Phillips, A. B., Turnock, S. R., Wilson, P. A.: Ship speed prediction based on machine learning for efficient shipping operation, *Ocean Engineering*, **245** (2022) 110449, <https://doi.org/10.1016/j.oceaneng.2021.110449>
- [16] Liu, S. Y.; Ouyang, Z. L., Chen, G., Zhou, X., Zou, Z. J.: Black-box modeling of ship maneuvering motion based on Gaussian process regression with wavelet threshold denoising, *Ocean Engineering*, **271** (2023) 113765, <https://doi.org/10.1016/j.oceaneng.2023.113765>
- [17] Xue, Y.; Liu, Y., Ji, C., Xue, G., Huang, S.: System identification of ship dynamic model based on Gaussian process regression with input noise, *Ocean Engineering*, **216** (2020) 107862, <https://doi.org/10.1016/j.oceaneng.2020.107862>
- [18] Jeon, M.; Noh, Y., Shin, Y., Lim, O. K., Lee, I., Cho, D.: Prediction of ship fuel consumption by using an artificial neural network, *Journal of Mechanical Science and Technology*, **32** (2018) 5785–5796, <https://doi.org/10.1007/s12206-018-1126-4>

- [19] Allahyarzadeh-Bidgoli, A.; Dezan, D. J., Salviano, L. O., de Oliveira Junior, S., Yanagihara, J. I.: FPSO fuel consumption and hydrocarbon liquids recovery optimization over the lifetime of a deep-water oil field, *Energy*, **181** (2019) 927–942, <https://doi.org/10.1016/j.energy.2019.05.146>
- [20] IHS Fairplay, World Register of Ships Manual (2013)
- [21] Kutner, M.H.; Nachtsheim, C.J., Neter, J., Li, W.: *Applied Linear Statistical Models*, 5th Ed., McGraw-Hill Professional, New York, USA (2004)
- [22] Breiman, L.; Friedman, J., Olshen, R., Stone, C.: *Classification and regression trees*, Statistics/probability series, Chapman and Hall/CRC, Belmont (1984)
- [23] Rasmussen, C.E.; Williams, C.K.I.: *Gaussian Processes for Machine Learning*, MIT Press, Cambridge, Massachusetts, USA (2006) <https://doi.org/10.7551/mitpress/3206.001.0001>
- [24] Friedman, J.: Greedy function approximation: A gradient boosting machine, *Annals of Statistics*, **29** (2001) 5, 1189–1232, <https://doi.org/10.1214/aos/1013203451>
- [25] Breiman, L.: Bagging Predictors, *Machine Learning*, **26** (1996) 123–140, <https://doi.org/10.1023/A:1018054314350>
- [26] Beale, M.H.; Hagan, M.T., Demuth, H.B.: Deep Learning Toolbox User's Guide, The MathWorks, Inc., Natick, MA, USA (2023).
- [27] The MathWorks: MATLAB Software Package, R2022b, *Dostupno na* <https://uk.mathworks.com/>, *Pristupljeno*: 2025-12-19
- [28] Valčić, M.; Martić, I., Degiuli, N., Grlj, C.G., Zhang, M.: Comparative analysis of machine learning methods for the prediction of brake power and rate of revolution for bulk carriers, *Ocean Engineering*, **322** (2025) 120517, <https://doi.org/10.1016/j.oceaneng.2025.120517>