

Dr. sc. Silvija Vlah Jerić

Izvanredna profesorica

Sveučilište u Zagrebu

Ekonomski fakultet

E-mail: svlah@efzg.hr

Orcid: <https://orcid.org/0000-0003-4738-6337>

ANALIZA FINANCIJSKE USPJEŠNOSTI MODELA STROJNOG UČENJA ZA PREDVIĐANJE SMJERA PROMJENE BURZOVNIH INDEKSA CEE I SEE REGIJE UZ RAZLIČITE METRIKE ZA VREDNOVANJE KLASIFIKATORA

UDK / UDC: 336.76:004

JEL klasifikacija / JEL classification: G17, C53

DOI: 10.17818/EMIP/2023/2.12

Prethodno priopćenje / Preliminary communication

Primljeno / Received: 8. studenoga 2022. / November 8, 2022

Prihvaćeno / Accepted: 15. ožujka 2023. / March 15, 2023

Sažetak

Cilj analize je istražiti utjecaj odabira metrike za vrednovanje klasifikatora na financijsku uspješnost sustava trgovanja temeljenih na modelima strojnog učenja za burzovne indekse iz zemalja CEE i SEE regija. Tehničkim indikatorima se koriste kao značajke za odabrane algoritme strojnog učenja pri predviđanju smjera promjena vrijednosti indeksa, tj. klasificiranje dana trgovanja u dvije klase. Istraživanje je pokazalo da odabir metrike za vrednovanje klasifikatora nema veliki utjecaj na financijsku uspješnost takvog sustava, no ipak su najveći prosječni prinosi po transakciji postignuti maksimizacijom točnosti. Nadalje, algoritam slučajne šume i naivni Bayesov klasifikator dali su najveće prosječne prinose korištenjem točnosti, dok su stroj potpornih vektora i algoritam k najbližih susjeda najveće prosječne prinose postigli pri korištenju površine ispod krivulje operativnih karakteristika. Utvrđeno je da očekivano veliki utjecaj na financijsku uspješnost ima odabir algoritma za strojno učenje te da algoritam slučajne šume daje najbolje rezultate na ovim podacima.

Ključne riječi: tehnička analiza, predviđanje promjene burzovnih indeksa, financijsko predviđanje, algoritmi za klasifikaciju, strojno učenje.

1. UVOD

Algoritmi strojnog učenja često se upotrebljavaju za predviđanje kretanja na burzi, obično za predviđanje cijena dionica ili vrijednosti burzovnog indeksa, a

još češće smjera njihove promjene. Jedan od pristupa takvim sustavima za predviđanje je korištenje skupa tehničkih indikatora kao ulaznih podataka za algoritme strojnog učenja. U ovom radu, algoritmi strojnog učenja koriste se za predviđanje smjera promjene vrijednosti odabranih burzovnih indeksa zemalja CEE i SEE regija. Smjer budućih kretanja vrijednosti indeksa predviđaju se razvrstavanjem u dvije klase: rast ili pad vrijednosti.

Indeks tržišta dionica često se upotrebljava za opisivanje ukupnog trenda tog tržišta. U ovom radu analiziraju se odabrana tržišta zbog svojih specifičnosti u odnosu na veća, razvijenija tržišta. Nekoliko je razloga zašto su CEE i SEE regija odabrani za ovu analizu. Naime, navedena tržišta dionica su još uvijek nedovoljno analizirana u literaturi (Škrinjarić, 2022). Kako na tim tržištima i dalje dominiraju neki čimbenici specifični za pojedine zemlje, njihovim uključivanjem u portfelj može se postići veća diversifikacija (Karkowska & Urjasz, 2022; Özer, Kamenković & Grubišić, 2020; Šego & Škrinjarić, 2018; Cevik, Korkmaz & Cevik, 2017). Također, prethodna istraživanja (Škrinjarić, 2018; Ferreira, 2018) pokazala su predvidljivost promjena vrijednosti indeksa na ovim tržištima, dugoročnu zavisnost prinosa na indekse, što podržava odbacivanje slabog oblika efikasnosti. Ovaj rad ispituje predvidljivost promjena vrijednosti na promatranim tržištima na alternativan način u odnosu na dosad korištene metode u literaturi za ova tržišta. Glavni cilj istraživanja je istražiti utjecaj odabira metrike za klasifikaciju na financijski učinak sustava trgovanja, odnosno strategija temeljenih na modelima strojnog učenja.

Usredotočujući se na strojno učenje, istraživanja u ovom području uglavnom su se bavila pitanjem mogu li modeli strojnog učenja (konkretnije, nadziranog učenja) dati bolje rezultate od standardnih statističkih pristupa u predviđanju tržišnog trenda, prinosa i volatilnosti (Kumbure, Lohrmann, Luukka & Porras, 2022). Također, brojni su primjeri u literaturi u kojima su se autori bavili problemom odabira najboljeg skupa tehničkih indikatora, no većina se ne fokusira na odabir vremenskih intervala za izračun indikatora (npr. Alsubaie, Hindi & Alsalman, 2019). Nadalje, neki radovi su istraživali uključivanje različitih tipova automatizacije u proces odabira temeljnih vremenskih intervala, tj. određivanja optimalnog razdoblja povijesnih podataka (Zhang, Cui, Xu, Li & Li, 2018; Chung & Shin, 2018; Vezeris, Schinas & Papaschinopoulos, 2018). Osim toga, Kovačević, Goluža, Merćep i Kostanjčar (2022) istraživali su učinak algoritama označavanja na financijski učinak dobivene strategije. S druge strane, u ovom radu ispituje se utjecaj odabira metrike za klasifikaciju na financijski učinak dobivenog sustava trgovanja, odnosno strategije. Naime, poznato je da odabir metrike za vrednovanje klasifikatora utječe na razvoj modela. Kao što su istaknuli Chrościcki i Chlebus (2022), sintetičke metrike kojima se obično koristi za razvoj modela mogu dovesti do suboptimalnih rješenja u poslovnim aplikacijama. Stoga su uspoređivali strategije razvoja modela za procjenu kreditne sposobnosti na temelju različitih metrika uspješnosti kako bi postigli najprofitabilniji kreditni portfelj te su predložili novu metriku pod nazivom Izračunata dobit (eng. *Calculated Profit*). Da bi u obzir uzeli pristanost u procjeni modela, Liu i Wang (2019) predložili su novu metriku,

srednju stopu profita (eng. *Mean Profit Rate*, MPR) i zaključili da je za predviđanje kretanja dionica MPR učinkovitija metrika od uobičajenih klasifikacijskih metrika.

Gledajući s aspekta tehničke analize, istraživanja trgovanja temeljenog na tehničkim indikatorima fokusirala su se sve više i na tržišta u razvoju, koja, kao što je prethodno već opisano, zajednički pružaju međunarodnim ulagačima alternativni izvor mogućnosti (McKenzie, 2007). Analiza koju su proveli Borowski i Pruchnicka-Grabias (2019) imala je cilj pronaći optimalne eksponencijalne pomične prosjeke za glavni oscilator tehničke analize MACD (*Moving Average Convergence Divergence*) za tvrtke iz indeksa WIG20, mWIG40 i sWIG80 indeksa, uvrštenih na Varšavsku burzu. Autori su pokazali da investitori ne bi trebali automatski slijediti standardne duljine za prosjeke koji se koriste u MACD-u, već tražiti optimalne vrijednosti za svako tržište i svaku tvrtku. Analiza koju je proveo Vlah Jerić (2020) evaluira i uspoređuje različite klasifikacijske algoritme za automatsku identifikaciju povoljnih dana za trgovanje na vrijednostima indeksa CROBEX. Ovaj rad nastavlja to istraživanje proširujući ga na više tržišta te analizirajući utjecaj izbora metrike za uspješnost klasifikacije na performanse sustava za predviđanje sa stajališta financijske uspješnosti trgovanja. Naime, strategija trgovanja kreira se na temelju naučenog modela, pri čemu se ispituje nekoliko algoritama strojnog učenja za klasifikaciju i različite metrike za vrednovanje klasifikatora. Takve strategije trgovanja temeljene na strojnom učenju zatim se uspoređuju na temelju prosječnih prinosa po transakciji, kao metrici koja se često koristi kada se ocjenjuje izvedba sustava trgovanja.

Što se tiče horizonta predviđanja, u ovom radu prognozira se samo sljedeći dan. Naime, prema hipotezi fraktalnog tržišta (Peters, 1994) kao alternativni hipoteze efikasnog tržišta (Fama, 1970), kraći investicijski horizonti tipični su za vremena i tržišta sa smanjenom likvidnošću i povećanom nestabilnošću, pa se pretpostavlja da bi predviđanje moglo biti bolje na kraći horizont za odabrana tržišta. Uz to, prijašnja istraživanja preporučuju korištenje kraćih vremenskih intervala za izračun tehničkih indikatora kao značajki algoritama strojnog učenja na kraći horizont predviđanja (Shynkevich, McGinnity, Coleman, Belatreche & Li, 2017). Prema tim preporukama su i u ovom radu izabrani vremenski intervali, tj. duljine prozora za izračun tehničkih indikatora.

Ostatak rada organiziran je na sljedeći način. U drugom poglavlju ukratko su objašnjeni podaci i metodologija korišteni za provođenje istraživanja. Treće poglavlje opisuje odabrane metrike za vrednovanje klasifikatora i financijske uspješnosti trgovanja, dok se u četvrtom poglavlju objašnjavaju detalji provedbe istraživanja i prezentiraju rezultati. Zaključci istraživanja su dani na kraju rada, u petom poglavlju.

2. OPIS PODATAKA I METODOLOGIJE

Podaci se sastoje od vremenskih serija podataka o vrijednostima na zatvaranju burze (zaključnim vrijednostima) te najvišim i najnižim dnevnim vrijednostima indeksa za svaki dan trgovanja od 2. siječnja 2011. godine do 22.

listopada 2021. godine. Riječ je o 10 indeksa CEE i SEE regija (neke zemlje su izuzete zbog nedovoljne količine podataka) za koje su vrijednosti preuzete s portala *Investing* (www.investing.com):

- BELEX (engl. *Belgrade Stock Exchange*, Beograd, Srbija; 2219 dana)
- BET (engl. *Bucharest Stock Exchange*, Bukurešt, Rumunjska, 2207 dana)
- IRS (engl. *Banja Luka Stock Exchange*, Banja Luka, Bosna i Hercegovina, 2215 dana)
- BUX (engl. *Budapest Stock Exchange*, Budimpešta, Mađarska, 2191 dana)
- CROBEX (engl. *Zagreb Stock Exchange*, Zagreb, Hrvatska, 2193 dana)
- PX (engl. *Prague Stock Exchange*, Prag, Češka, 2205 dana)
- SAX (engl. *Bratislava Stock Exchange*, Bratislava, Slovačka, 2194 dana)
- SBITOP (engl. *Ljubljana Stock Exchange*, Ljubljana, Slovenija, 2203 dana)
- SOFIX (engl. *Bulgarian Stock Exchange*, Sofija, Bugarska, 2180 dana)
- WIG20 (engl. *Warsaw Stock Exchange*, Varšava, Poljska, 2202 dana).

Smjer budućeg kretanja vrijednosti indeksa se u ovom radu prognozira razvrstavanjem u dvije klase. Dodjeljivanje oznaka svakoj podatkovnoj točki, tj. danu izvodi se u skladu s nadolazećim ponašanjem vrijednosti sljedeći dan na zatvaranju, tako da se oznaka „Gore“ dodjeljuje podatkovnoj točki kada se zaključna vrijednost povećala, a oznaka „Dolje“ kada se odgovarajuća zaključna vrijednost smanjila ili se nije promijenila. Preciznije,

$$\text{Oznaka}_t = \begin{cases} \text{"Gore"} & \text{ako } \frac{c_{t+1} - c_t}{c_t} > 0, \\ \text{"Dolje"} & \text{inače,} \end{cases}$$

gdje su c_t i c_{t+1} su vrijednosti indeksa na zatvaranju dana t i $t + 1$, respektivno.

Na temelju podataka o trgovanju izračunavaju se tehnički indikatori kojima se koristi kao značajkama za odabrane algoritme strojnog učenja pri predviđanju smjera promjena vrijednosti indeksa, tj. klasificiranje dana trgovanja u dvije gore definirane klase. Odabrani pokazatelji često se koriste u literaturi, a detaljna objašnjenja postupaka izračuna mogu se pronaći primjerice u Bruni (2017). Kao što je u uvodnom dijelu objašnjeno, očekuje se veća uspješnost kod prognoziranja na kraće horizonte za promatrana tržišta te se u ovom radu prognozira samo sljedeći dan. Uz to je preporučljivo i korištenje kraćih vremenskih intervala za izračun tehničkih indikatora na kraći horizont (Shynkevich et al., 2017), pa se u ovoj analizi koristi sljedećim tehničkim indikatorima:

- momentum tijekom 3 razdoblja (MOM_3)
- eksponencijalni pomični prosjek tijekom tri i pet razdoblja (EMA_3 , EMA_5 , od eng. *exponential moving average*),
- konvergencija/divergencija pomičnog prosjeka s 3, 10 i 5 respektivno, kao tri parametra potrebna za izračun ($MACD_{3,10,5}$, od eng. *Moving Average Convergence/Divergence*)

- profitabilnost investicije tijekom 3, 5 i 10 razdoblja (ROI_3 , ROI_5 , ROI_{10} , od eng. *Return on Investment*),
- indeks relativne snage tijekom 3, 5 i 10 razdoblja (RSI_3 , RSI_5 , RSI_{10} , od eng. *Relative Strength Index*),
- stohastički indeks relativne snage kroz 3, 5 i 10 razdoblja ($STOCHRSI_3$, $STOCHRSI_5$, $STOCHRSI_{10}$, od eng. *Stochastic Relative Strength Index*),
- prosječni istinski raspon tijekom 3 razdoblja (ATR_3 , od eng. *Average True Range*)
- prosječni indeks smjera tijekom 3 posljednja razdoblja (ADX_3 , od eng. *Average Directional Index*)
- Williams %R tijekom tri razdoblja (WPR_3),
- indeks robnih cijena kroz tri razdoblja (CCI_3 , od eng. *Commodity Channel Index*)
- ultimativni oscilator s 3, 5 i 10 respektivno, kao tri parametra potrebna za izračun ($UO_{3,5,10}$, od eng. *Ultimate Oscillator*).

Kako bi se ispitalo utječe li izbor algoritma na učinak odabrane metrike na performanse sustava za predviđanje, tj. kao kontrola ovog faktora, testirano je nekoliko algoritama strojnog učenja:

- slučajna šuma (RF od eng. *Random Forest*) - podešena promjenom broja nasumično odabranih prediktora
- stroj potpornih vektora (SVM od eng. *Support Vector Machine*) - s jezgrom radialne bazne funkcije (RBF) uz podešavanje hiperparametra C i sigma
- naivni Bayesov klasifikator (NB)
- k najbližih susjeda (KNN od eng. *k Nearest Neighbors*) – podešen po broju susjeda.

Algoritam slučajnih šuma temelji se na modifikaciji *bagging* metode te gradi veliki broj nekoreliranih stabala odluke koja u slučaju klasifikacije glasaju, nakon čega se odabire klasa s najviše glasova. Jednostavan je za učenje i podešavanje te stoga vrlo popularan. Stroj potpornih vektora također je vrlo popularan i učinkovit algoritam, originalno razvijen baš za klasifikaciju. Osnovna ideja je pronalaženje hiperravnine u prostoru značajki koja klasificira podatkovne točke tako da udaljenost između točaka različitih klasa bude maksimalna, tj. da se maksimizira margina razdvajanja. Naivan Bayesov klasifikator je jednostavan algoritam temeljen na Bayesovom pravilu koji pretpostavlja uvjetnu nezavisnost između značajki unutar svake klase, što značajno pojednostavnjuje problem, a zbog čega se i kaže da je „naivan“. Usprkos tome i činjenici da je ta pretpostavka u stvarnosti često narušena, u praksi se pokazao vrlo učinkovitim. Ideja algoritma k najbližih susjeda vrlo je jednostavna i sastoji se od odabira one klase za određenu podatkovnu točku koja je najčešća među k najbližih susjeda. Svi korišteni algoritmi dobro su opisani u literaturi pa se detalji mogu pronaći primjerice u (Hastie, Tibshirani & Friedman, 2001).

3. ODABRANE METRIKE ZA VREDNOVANJE KLASIFIKATORA I STRATEGIJE TRGOVANJA

Kao metrike za vrednovanje klasifikatora u ovom radu upotrebljavaju se točnost, Cohenova kappa te površina ispod krivulje operativnih karakteristika. Kao financijska mjera uspješnosti trgovanja, koje se temelji na dobivenom modelu, upotrebljava se prosječni prinos po transakciji kao jedna od često primjenjivanih mjera za evaluaciju takvih sustava (Shynkevich et al., 2017). U nastavku su objašnjene navedene mjere kao i pripadajuća strategija trgovanja.

Točnost (Acc od eng. *Accuracy*) je vjerojatno najpoznatija metrika za vrednovanje klasifikatora, a predstavlja udio točno klasificiranih primjera u skupu svih primjera. Preciznije, točnost se računa kao

$$Acc = \frac{TP + TN}{TP + TN + FP + FN'}$$

gdje su TP stvarno pozitivni, FP lažno pozitivni, TN stvarno negativni, a FN lažno negativni primjeri u matrici zabune.

Cohenova kappa mjeri slaganje između dva procjenjivača koji svaki klasificiraju N stavki u C međusobno isključivih kategorija. Ako je p_a udio opažanja koja se slažu, a p_ε udio koji se slažu zbog slučajnosti, tada je Cohenova kappa

$$\kappa = \frac{p_a - p_\varepsilon}{1 - p_\varepsilon}$$

te se može shvatiti kao mjera korigirana za koliko često se ocjenjivači mogu slučajno složiti, pa se stoga uporaba ove metrike preporučuje za nebalansirane skupove podataka. U terminima klasične matrice zabune koja se koristi kod binarne klasifikacije Cohenova kappa se može izraziti kao

$$\kappa = \frac{2 \cdot (TP \cdot TN - FP \cdot FN)}{(TP + FP) \cdot (FP + TN) + (TP + FN) \cdot (FN + TN)}$$

Minimalna vrijednost Cohenove kappe je -1 (savršeno pogrešno predviđanje), a maksimalna vrijednost $+1$ (savršena klasifikacija). Nadalje, ako je $\kappa \approx 0$, takvo predviđanje je slično nasumičnom pogađanju. Više o toj metrici može se pronaći npr. u radu autora Chicco, Warrens i Jurman (2021).

Površina ispod krivulje operativnih karakteristika, tj. ROC krivulje (*AUC* od eng. *area under ROC curve*) mjeri površinu ispod krivulje koja prikazuje stvarno pozitivnu stopu $TPR = TP / (TP + FP)$ (od eng. *True Positive Rate*) u odnosu na lažno pozitivnu stopu $FPR = FP / (TP + FP)$ (od eng. *False Positive Rate*). AUC ima raspon vrijednosti od 0 do 1. Model čija su predviđanja 100 % pogrešna ima AUC jednak 0, dok onaj čija su predviđanja 100 % točna ima AUC jednak 1. Više o ovoj mjeri može se pročitati primjerice u radu autora Ling, Huang, Zhang (2003) koji preporučuju njezino korištenje umjesto točnosti.

Prosječni prinosi po transakciji računaju se u fazi testiranja kao prosječna vrijednost prinosa svih pojedinačnih transakcija, za svaki indeks. U ovom radu bit

će prikazani u obliku postotka. Koja transakcija će biti učinjena određuje se na temelju rezultata klasifikacije budućih vrijednosti indeksa, odnosno naučenog modela strojnog učenja. Kada je sustav predvidio kretanje vrijednosti indeksa „Gore“, on se u trenutku predviđanja kupuje te se na kraju sljedećeg dana prodaje, a prinos od ove transakcije (u postotku) izračunava se kao

$$\frac{c_{t+1} - c_t}{c_t} \cdot 100.$$

Kada je sustav predvidio kretanje vrijednosti indeksa „Dolje“, on se u trenutku predviđanja prodaje te se na kraju sljedećeg dana ponovno kupuje, a prinos od ove transakcije izračunava se kao

$$\frac{c_t - c_{t+1}}{c_t} \cdot 100.$$

Iako bi se mogao očekivati jak linearni odnos između klasifikacijske izvedbe modela i financijske izvedbe strategije temeljene na istom modelu, specifičnosti pojedinog tržišta, kao npr. stupanj balansiranoosti podataka, mogu utjecati na rezultate pri korištenju različitih metrika. Iako je točnost jedna od onih koja se čini najprirodnijom mjerom i koristi se desetljećima, po mnogim autorima ona nije učinkovita jer, između ostalog, ne uzima u obzir distribuciju pogrešne klasifikacije među klasama niti marginalne distribucije (Delgado & Tibau, 2019). Kappa je također jedna od najraširenijih mjera koja se koristi u različitim područjima i disciplinama. Međutim, neki autori, uključujući i samog tvorca Kappa statistike, Jakoba Cohena, upozorili su da bi Kappa mogla biti neadekvatna u različitim okolnostima, primjerice zbog problema koji mogu nastati jer nije jasno kako definirati hipotetsku vjerojatnost slučajnog dogovora (npr. Brennan & Prediger, 1981). Warrens (2010) opisuje takozvani Kappa paradoks. S druge strane, Ling i sur. (2003) po prvi put uspostavljaju rigorozne kriterije za usporedbu mjera evaluacije algoritama, sugeriraju da bi AUC trebao zamijeniti točnost pri mjerenju i usporedbi klasifikacijskih sustava te potiču istraživače da preispitaju zaključke temeljene na točnosti u strojnom učenju. Odnos između ROC krivulja i Cohenove kappe istražio je Ben-David (2008).

4. REZULTATI

Prvih 1500 dana trgovanja uzeto je kao skup za učenje (otprilike 60 % podataka za svaki indeks) i svi podaci su standardizirani (upotrebom vrijednosti iz skupa za učenje). Ugađanje algoritma provedeno je unakrsnom provjerom za vremenske serije, tj. uz pomoć tehnike koja pomiče skupove za učenje i testiranje u vremenu (eng. *rolling forecasting origin technique*, Hyndman & Athanasopoulos, 2021). Pritom se u ovom istraživanju koristi opcija fiksne duljine vremenskog intervala, tj. duljina skupa za učenje uvijek je ista, dok je omjer između duljine intervala za učenje i horizonta do kojeg se skup za testiranje (broj uzastopnih vrijednosti u skupu za učenje u odnosu na broj uzastopnih vrijednosti u skupu za testiranje) 80 : 20 bez preklapanja. Sve analize provedene su uz pomoć programa R (R Core Team, 2021).

Rezultati za prosječnu točnost i pripadajuće prosječne prinose po transakciji ovisno o algoritmu prikazani su u Tablici 1 (uprosječeni kroz promatrane indekse za svaki par metrika/algoritam). Može se vidjeti da je najveća prosječna točnost postignuta RF algoritmom (0,710), a pripadajući prosječni prinosi po transakciji su 0,388 %. Najveća prosječna kappa vrijednost postignuta je također RF algoritmom (0,413), uz pripadne prosječne prinose po transakciji od 0,382 %. Najviša AUC vrijednost postignuta je pak SVM algoritmom (0,787), uz postignute prosječne prinose po transakciji od 0,384 %. Međutim, u ovom slučaju je vrlo mala razlika u odnosu na rezultate postignute RF algoritmom. Stoga se čini da je RF najuspješniji u ovako postavljenim uvjetima. Uspoređujući metrike za vrednovanje klasifikatora na temelju dobivenih prosječnih prinosa po transakciji, može se uočiti da su najveći prinosi (0,388 %) postignuti maksimizacijom točnosti (0,710), uz također i najmanju pripadnu standardnu devijaciju. Rezultati su s obzirom na postignute vrijednosti mjera za klasificiranje te prosječnih prinosa po transakciji prihvatljivi.

Tablica 1.

Prosječna točnost, kappa i površina ispod ROC krivulje te pripadajući prosječni prinosi po transakciji (%) ovisno o algoritmu (standardna devijacija u zagradama)

	RF		SVM		NB		KNN	
Acc	0.710	(0.076)	0.697	(0.093)	0.648	(0.072)	0.665	(0.066)
Pr. prinos	0.388	(0.173)	0.362	(0.204)	0.301	(0.154)	0.318	(0.151)
Kappa	0.413	(0.155)	0.378	(0.191)	0.293	(0.137)	0.320	(0.141)
Pr. prinos	0.382	(0.176)	0.360	(0.202)	0.299	(0.156)	0.318	(0.151)
AUC	0.780	(0.096)	0.787	(0.101)	0.704	(0.098)	0.725	(0.094)
Pr. prinos	0.381	(0.177)	0.384	(0.180)	0.289	(0.149)	0.319	(0.151)

Kako bi se promotriilo koliko se prosječni prinosi po transakciji razlikuju primjenom različitih metrika za vrednovanje klasifikacije, Tablica 2 prikazuje broj indeksa za koje su različiti prinosi dobiveni variranjem tih metrika. Uz to, dani su i podaci sažeti iz prethodne tablice o najvećim postignutim prosječnim prinosima po algoritmu te metrika za vrednovanje klasifikatora kojom su oni postignuti. Može se vidjeti da se korištenjem SVM algoritma u najvećem broju slučajeva (8 od 10) dobivaju različiti prinosi korištenjem različitih metrika, dok se korištenjem KNN algoritma dobivaju većinom jednaki prinosi neovisno o korištenoj metrici (samo je u jednom slučaju dobivena različita vrijednost). Uz to, SVM i KNN najveće prosječne prinose postižu pri korištenju AUC metrike, dok RF i NB daju najbolje vrijednosti prosječnih prinosa uz Acc.

Tablica 2.

Broj indeksa za koje se variranjem metrike za vrednovanje klasifikatora dobivaju različiti prosječni prinosi (%), najveći prosječni prinosi te metrika uz koju je isti postignut

	RF	SVM	NB	KNN
Br. razl. vr. prinosa	4	8	5	1
Najveći. pr. prinos	0.388	0.384	0.301	0.319
(postignut je uz)	Acc	AUC	Acc	AUC

Tablica 3 raščlanjuje dosadašnje rezultate prikazujući ih za svaki pojedinačni indeks kako bi se uočile njihove specifičnosti te navodi najviši prosječni prinos po transakciji među pojedinačnim prinosima. Uz to su prikazane i metrike za vrednovanje klasifikatora kojima su ti prinosi postignuti. U nekim slučajevima više različitih metrika daje istu vrijednost. Primjerice, za BELEX, BIRS i BUX sve tri metrike daju iste prosječne prinose po transakciji, dok za PX i SAX samo po jedna metrika daje najbolji rezultat za prinose (AUC i Acc respektivno).

U Tablici 3 dani su i podaci o postotku dana kada se vrijednost indeksa na zatvaranju povećala u odnosu na prethodni dan, tj. dana označenih s „Gore“ u odnosu na ukupan broj dana trgovanja. Navedeni postotci su prilično ujednačeni i podaci ne pokazuju značajnu nebalansiranost, osim možda vrlo malo u primjerice SAX i BIRS indeksa. Za oba spomenuta slučaja korištenje točnosti omogućilo je postizanje najvećih prosječnih prinosa po transakciji.

Tablica 3.

Najviši prosječni prinos po transakciji (%) među pojedinačnim indeksima

	RF	SVM	NB	KNN	Metrike	% „Gore“
BELEX	0.2330	0.2556	0.1609	0.1917	Acc, Kappa, AUC	49 %
BET	0.4147	0.4131	0.3375	0.3828	Acc, Kappa	46 %
BIRS	0.2481	0.2342	0.1646	0.2142	Acc, Kappa, AUC	42 %
BUX	0.6223	0.6021	0.5083	0.4084	Acc, Kappa, AUC	47 %
CROBEX	0.0248	-0.0061	0.0032	0.0159	Acc, Kappa	49 %
PX	0.5137	0.5633	0.3589	0.4943	AUC	47 %
SAX	0.3200	0.2895	0.2941	0.2311	Acc	37 %
SBITOP	0.5510	0.5455	0.3795	0.4159	Acc, AUC	48 %
SOFIX	0.4308	0.4169	0.3929	0.3539	Acc, AUC	49 %
WIG20	0.5245	0.5544	0.4790	0.4800	Acc, AUC	50 %

Može se primijetiti da se prinosi na različiti tržištima podosta razlikuju. Najmanji prosječni prinos po transakciji ostvaren RF algoritmom je 0,0248 % (za CROBEX), a najveći je 0,6223 % (za BUX). Brojni su mogući razlozi za to. Moguće je da su niži prinosi dobiveni zbog skrivene strukture podataka zbog kojih

algoritam lošije radi za neka tržišta. Promatrajući razlike među algoritmima za pojedinačno tržište, razlike između RF i SVM rezultata ne čine se velikima. S druge strane, NB i KNN algoritmi su znatno manje uspješni za većinu tržišta, konzistentno, pa se čini da ipak nešto drugo izaziva razlike u prinosima na različitim tržištima. Još jedan mogući i više vjerojatan razlog za podosta različite postignute prinose na različitim tržištima leži u odabiru jednakih tehničkih indikatora i njihovih pripadnih parametara. Konačno, očiti razlog je i samo kretanje vrijednosti promatranih indeksa, koje za veće raspone između vrijednosti dva uzastopna dana rezultira i većim prinosima, naravno uz pretpostavku da je dan točno klasificiran. Naime, algoritam prognozira samo rast ili pad, a ne govori ništa o intenzitetu.

Uspješnost opisanih sustava za predviđanje možda bi bilo moguće poboljšati koristeći se klasifikacijom na tri klase umjesto dvije kao u Shynkevich et al. (2017). Naime, potencijalno bi bilo korisno uvesti i klasu „Bez pomicanja“, što bi značilo da za odgovarajući dan nema transakcije. Naime, u tom bi slučaju bili uvedeni negativni i pozitivni prag te ako se relativna promjena nalazi u rasponu između tih pragova, smatrala bi se beznačajnom i oznaka „Bez pomicanja“ dodijelila bi se toj podatkovnoj točki, tj. danu. Međutim, iako bi se uvođenje tri klase umjesto dvije moglo opravdati sa stajališta profitabilnosti jer točno predviđanje rasta ili pada ne daje nužno profitabilnu strategiju (zbog transakcijskih troškova i dr.), trebalo bi ispitati i kakva bi bila uspješnost algoritama za klasifikaciju u tom slučaju. Trebalo bi pritom u obzir uzeti i transakcijske troškove.

5. ZAKLJUČAK

U ovom radu ispitana su četiri algoritma strojnog učenja za predviđanje smjera promjene vrijednosti odabranih burzovnih indeksa zemalja CEE i SEE regija. Na temelju tehničkih indikatora dani su klasificirani u dvije klase. Uspješnost klasifikacije mjerila se korištenjem točnosti, Cohenove kappe te površinom ispod ROC krivulje, a time ostvarena financijska uspješnost prosječnim prinosima po transakciji.

Istraživanje je pokazalo da odabir metrike za vrednovanje klasifikatora nema veliki utjecaj na financijsku uspješnost takvog sustava za predviđanje temeljenog na strojnom učenju. Ipak, određene razlike postoje te je pokazano da su najveći prinosi postignuti maksimizacijom točnosti. Nadalje, RF i NB dale su najbolje vrijednosti prosječnih prinosa korištenjem točnosti kao metrike za vrednovanje klasifikatora, dok su SVM i KNN najveće prosječne prinose postigli uz korištenje površine ispod ROC krivulje. Također, pokazano je da se korištenjem SVM algoritma u najvećem broju slučajeva dobivaju različiti prinosi korištenjem različitih metrika, dok se korištenjem KNN algoritma dobivaju većinom jednaki prinosi neovisno o korištenoj metrici.

S druge strane, veliki utjecaj na financijsku uspješnost sustava za predviđanje ima odabir algoritma za strojno učenje, što je i očekivano. Naime, RF i SVM algoritmi su na svim odabranim tržištima ostvarili znatno veće prinose, nego NB i KNN algoritmi, što je velikim dijelom posljedica veće uspješnosti tih

algoritama na temelju metrika za vrednovanje klasifikatora uz koje su i učeni. Međutim, pokazalo se da razlike postoje i između samih RF i SVM algoritama te da RF daje nešto bolje rezultate na ovim podacima. Također, RF algoritam je i inače popularan jer može dobro funkcionirati bez puno podešavanja (tzv. „*out of the box*“), a ovo istraživanje ponovno to potvrđuje.

Buduća istraživanja mogla bi ispitati uspješnost sustava za prognoziranje na tržištima promatranim u ovom radu uz klasificiranje u tri klase umjesto dvije korištenjem negativnog i pozitivnog praga, tj. dodatak kategorije koja bi označavala da se na promatrani dan uopće ne trguje. Također, korisno bi bilo ispitati druge kombinacije tehničkih indikatora i njihovih parametara.

LITERATURA

Alsubaie, Y., Hindi, K.E. & Als Salman, H. (2019). Cost-Sensitive Prediction of Stock Price Direction: Selection of Technical Indicators. *IEEE Access*, 7, 146876–146892. <https://doi.org/10.1109/ACCESS.2019.2945907>

Ben-David, A. (2008). About the relationship between ROC curves and Cohen's kappa. *Engineering Applications of Artificial Intelligence*, 21(6), 874-882. <https://doi.org/10.1016/j.engappai.2007.09.009>.

Borowski, K. & Pruchnicka-Grabias, I. (2019). Optimal lengths of moving averages for the MACD oscillator for companies listed on the Warsaw Stock Exchange. *Bank i Kredyt*, 50(5), 457–478.

Brennan, R. L. & Prediger, D. J. (1981). Coefficient Kappa: Some Uses, Misuses, and Alternatives. *Educational and Psychological Measurement*, 41(3), 687–699. <https://doi.org/10.1177/001316448104100307>

Bruni, R. (2017). Stock Market Index Data and indicators for Day Trading as a Binary Classification problem. *Data in Brief*, 10, 569–575. <https://doi.org/10.1016/j.dib.2016.12.044>

Cevik, E. I., Korkmaz, T. & Cevik, E. (2017). Testing causal relation among central and eastern European equity markets: evidence from asymmetric causality test. *Economic Research-Ekonomska Istraživanja*, 30(1), 381–393. <https://doi.org/10.1080/1331677X.2017.1305774>

Chicco, D., Warrens, M. J. & Jurman, G. (2021). The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment, *IEEE Access*, 9, 78368–78381, <https://doi.org/10.1109/ACCESS.2021.3084050>

Chrościcki, D. & Chlebus, M. (2022). The Advantage of Case-Tailored Information Metrics for the Development of Predictive Models, Calculated Profit in Credit Scoring. *Entropy*, 24(9):1218. <https://doi.org/10.3390/e24091218>

Chung, H. & Shin, K. (2018). Genetic Algorithm-Optimized Long Short-Term Memory Network for Stock Market Prediction. *Sustainability*, 10(10), 3765. <https://doi.org/10.3390/su10103765>

Delgado, R. & Tibau, X.-A. (2019). Why Cohen's Kappa should be avoided as performance measure in classification. *PLoS ONE* 14(9): e0222916. <https://doi.org/10.1371/journal.pone.0222916>

Fama, E. (1970). Efficient capital markets: A review of theory and empirical work. *Journal of Finance*, 25(2), 383–417. <https://doi.org/10.2307/2325486>

Ferreira, P. (2018). Long-range dependencies of Eastern European stock markets: A dynamic detrended analysis, *Physica A: Statistical Mechanics and its Applications*, 505, 454–470. <https://doi.org/10.1016/j.physa.2018.03.088>.

Hastie, T., Tibshirani, R., & Friedman, J. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer. <https://doi.org/10.1007/978-0-387-21606-5>

- Hyndman, R. J., & Athanasopoulos, G. (2021). *Forecasting: principles and practice*, 3rd edition, OTexts: Melbourne, Australia. <https://otexts.com/fpp3/> [pristupljeno: 2/3/2023]
- Karkowska, R. & Urjasz, S. (2022). Linear and Nonlinear Effects in Connectedness Structure: Comparison between European Stock Markets. *Entropy*, 24(2), 303. <https://doi.org/10.3390/e24020303>.
- Kovačević, T., Goluža, S., Merćep, A. & Kostanjčar, Z. (2022). Effect of labeling algorithms on financial performance metrics, 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), 980-984, <https://doi.org/10.23919/MIPRO55190.2022.9803522>.
- Kumbure, M., Lohrmann, C., Luukka, P. & Porras, J. (2022.) Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems With Applications*, 197, 116659, <https://doi.org/10.1016/j.eswa.2022.116659>
- Ling, C. X., Huang, J. & Zhang, H. (2003). AUC: A Better Measure than Accuracy in Comparing Learning Algorithms. In: Xiang, Y., Chaib-draa, B. (eds) *Advances in Artificial Intelligence. Canadian AI 2003. Lecture Notes in Computer Science*, vol 2671. Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-44886-1_25
- Liu, G. & Wang, X. (2019). A new metric for individual stock trend prediction. *Engineering Applications of Artificial Intelligence*, 82, 1-12. <https://doi.org/10.1016/j.engappai.2019.03.019>.
- McKenzie, M. D. (2007). Technical Trading Rules in Emerging Markets and the 1997 Asian Currency Crises. *Emerging Markets Finance and Trade*, 43(4), 46–73. <https://doi.org/10.2753/REE1540-496X430403>
- Özer, M., Kamenković, S. & Grubišić, Z. (2020). Frequency domain causality analysis of intra- and inter-regional return and volatility spillovers of South-East European (SEE) stock markets. *Economic research-Ekonomska istraživanja*, 33(1), 1-25. <https://doi.org/10.1080/1331677X.2019.1699138>
- Peters, E. (1994). *Fractal Market Analysis – Applying Chaos Theory to Investment and Analysis*. New York: John Wiley & Sons, Inc.
- R Core Team (2021). *R: A language and environment for statistical computing* [Computer software]. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Shynkevich, Y., McGinnity, T.M., Coleman, S., Belatreche, A. & Li, Y. (2017). Forecasting Price Movements using Technical Indicators: Investigating the Impact of Varying Input Window Length. *Neurocomputing*, 264, 71–88. <https://doi.org/10.1016/j.neucom.2016.11.095>
- Šego, B. & Škrinjarić, T. (2018). Quantitative Research Of Zagreb Stock Exchange-Literature Overview For The Period From Establishment Until 2018. *Ekonomski pregled*, 69(6), 655-743. <https://doi.org/10.32910/ep.69.6.3>
- Škrinjarić, T. (2018). Testing for Seasonal Affective Disorder on Selected CEE and SEE Stock Markets. *Risks*, 6(4), 140. <https://doi.org/10.3390/risks6040140>
- Škrinjarić, T. (2022). Higher Moments Actually Matter: Spillover Approach for Case of CESEE Stock Markets. *Mathematics*, 10(24), 4811. <https://doi.org/10.3390/math10244811>
- Vezeris, D. T., Schinas, C. J. & Papanichopoulos, G. (2018). Profitability Edge by Dynamic Back Testing Optimal Period Selection for Technical Parameters Optimization, in *Trading Systems with Forecasting*. *Computational Economics*, 51, 761–807. <https://doi.org/10.1007/s10614-016-9640-x>
- Vlah Jerić, S. (2020). Comparing classification algorithms for prediction on CROBEX data. *Croatian Review of Economic, Business and Social Statistics*, 6(2), 4–11. <https://doi.org/10.2478/crebss-2020-0007>
- Warrens, M. J. (2010). A Formal Proof of a Paradox Associated with Cohen's Kappa. *Journal of Classification*, 27, 322–332 <https://doi.org/10.1007/s00357-010-9060-x>
- Zhang, J., Cui, S., Xu, Y., Li, Q. & Li, T. (2018). A novel data-driven stock price trend prediction system. *Expert Systems with Applications*, 97, 60–69. <https://doi.org/10.1016/j.eswa.2017.12.026>

Silvija Vlah Jerić, PhD

Associate Professor
University of Zagreb
Faculty of Economics and Business
E-mail: svlah@efzg.hr
Orcid: <https://orcid.org/0000-0003-4738-6337>

ANALYSIS OF THE FINANCIAL PERFORMANCE OF MACHINE LEARNING MODELS FOR PREDICTING THE DIRECTION OF CHANGES IN CEE AND SEE STOCK MARKET INDICES WITH DIFFERENT CLASSIFICATION EVALUATION METRICS***Abstract***

The aim of the analysis is to investigate the influence of the selection of classification evaluation metrics on the financial performance of trading systems based on machine learning models for stock market indices from CEE and SEE regions. Technical indicators are used as features for selected machine learning algorithms when predicting the direction of index value changes, i.e. classifying trading days into two classes. The research showed that the choice of classifier evaluation metrics does not have a great impact on the financial performance of such a system. However, the highest average returns per trade were achieved by maximizing accuracy. Furthermore, the random forest algorithm and the naive Bayesian classifier gave the highest average returns using accuracy, while the support vector machine and the k-nearest neighbor algorithm achieved the highest average returns when using the area under the receiver operating characteristic curve. It was determined that the choice of machine learning algorithm has an expectedly large impact on financial performance and that the random forest algorithm gives the best results on this data.

Keywords: technical analysis, forecasting stock index movement, financial forecasting, classification algorithms, machine learning

JEL classification: G17, C53

