

Predviđanje promene trenda vrednosti berzanskog indeksa Belex15 pomoću LS-SVM klasifikatora

Ivana Marković, Jelena Stanković

Katedra za računovodstvo, matematiku i informatiku

Ekonomski fakultet

Niš, Republika Srbija

ivana.markovic@eknfak.ni.ac.rs, jelenas@eknfak.ni.ac.rs

Miloš Stojanović

Studijski program Savremene računarske tehnologije

Visoka tehnička škola strukovnih studija

Niš, Republika Srbija

milosstojanovic10380@gmail.com

Miloš Božić

Katedra za energetiku

Elektronski fakultet

Niš, Republika Srbija

miloslbozic@gmail.com

Sadržaj— U radu je predložen model za predviđanje trenda promene vrednosti Belex15 indeksa primenom Least squares support vector machine (LS-SVM) metode za klasifikaciju. Selekcija atributa za prikazani predikcioni model zasnovana je na analizi tehničkih indikatora. Atributi koji su korišćeni za treniranje modela su izvedeni iz logaritamskih prinosa direktno i obračunom eksponencijalnog pokretnog proseka (eng. Exponential Moving Average – EMA). Rezultati testiranja su pokazali da je predloženi model pogodan za kratkoročno predviđanje promena pravca kretanja berzanskih indeksa.

Ključne reči: Predviđanje promene trenda; Metoda najmanjih kvadrata podržavajućih vektora (LS-SVM); Klasifikacija;

I. UVOD

Finansijsko tržište je kompleksan, evolutivni i dinamičan sistem, koji se ponaša izrazito nelinearno [1]. Predikciju u oblasti finansija uslovljavaju velika raznovrsnost, ali i nestacionarnost i nestruktuiranost podataka sa visokim stepenom nestabilnosti i izraženim skrivenim vezama.

Poznato je da su precizna predviđanja kretanja indeksa cena akcija veoma važna za razvoj efikasne strategije trgovanja na tržištu [2]. Većina trgovinskih praksi usvojenih od strane finansijskih analitičara se oslanja na tačna predviđanja nivoa cena finansijskih instrumenata. Međutim, novije studije, predložene u [3], su pokazale da strategije trgovanja vođene prognozama o pravcu promene cena mogu biti efikasnije i generisati veći prinos.

U mnogim studijama algoritmi mašinskog učenja pokazali su se veoma efikasnim u predviđanju pravca kretanja vrednosti berzanskih indeksa i na taj način doprineli uvećanju prinosa i smanjenju rizika usled trgovanja. U [4] je pokazano da metode podržavajućih vektora (eng. Support Vector Machines - SVMs) i metode najmanjih kvadrata podržavajućih vektora (eng. Least Squares Support Vector Machines LS-SVMs), jedne od reformulacija SVM metode, postižu bolje predikcione rezultate u ovoj oblasti, u odnosu na ostale algoritme mašinskog učenja.

Iako mnoga istraživanja ukazuju da promene cena akcija nisu potpuno nasumične, posmatrano u dužem vremenskom intervalu promena cena se aproksimira slučajnim procesom (eng. random walk). Stoga se stepen preciznosti od oko 60% koji se dobija korišćenjem metoda mašinskog učenja često opisuje kao zadovoljavajući za predviđanja na tržištima kapitala [5].

Nasuprot fundamentalnoj analizi tržišta kapitala u kojoj dominiraju makroekonomski pokazatelji, kao što su, na primer, inflacija, nezaposlenost i nivo kamatnih stopa, tehnička analiza bazira se na uverenju da sama dešavanja na berzama daju dovoljno podataka za predviđanja budućih vrednosti.

Belex 15 je vodeći indeks Beogradske berze, čiju vrednost određuju cene najlikvidnijih akcija, kojima se trguje na regulisanom tržištu Beogradske berze. Namena indeksa je da meri promene cena akcija kojima se trguje metodom kontinuiranog trgovanja, a koje su prethodno zadovoljile kriterijum za uključivanje u indeksnu korpu.

S obzirom da je cilj ovog rada predviđanje tendencija u promenama vrednosti indeksa, u radu su korišćeni isključivo tehnički indikatori za formiranje predikcionog modela. Predikcioni model zasnovan je na analizi prinosa na berzanski indeks. Atributi koji su korišćeni u ovom modelu izvedeni su iz dve prethodne vrednosti logaritamskog prinosa zasnovanih na vrednostima posmatranim u periodima od po deset dana i obračunom desetodnevno eksponencijalnog pokretnog proseka logaritamskih prinosa. Problem predviđanja pravca promene vrednosti berzanskog indeksa se zatim modeluje kao problem binarne klasifikacije.

Nastavak rada je organizovan na sledeći način: U drugom delu predstavljene su osnove LS-SVM metode za binarnu klasifikaciju. U trećem delu opisan je predloženi predikcioni model, dok su rezultati njegovog testiranja predstavljeni u četvrtom delu. U petom delu istaknuti su neki od zaključaka i pravci budućih istraživanja.

II. OSNOVE LEAST SQUARES SUPPORT VECTOR MACHINE ZA BINARNU KLASIFIKACIJU

Posmatrajmo trening skup od ukupno N primera, gde su x_k ulazni vektori, a $y_k \in \{-1, +1\}$ njima pridružene oznake binarnih klasa. LS-SVM za binarnu klasifikaciju je definisan u [6], na sledeći način:

$$\min_{w,b,e} J_{LS}(w,b,e) = \frac{1}{2} w^T w + C \frac{1}{2} \sum_{k=1}^N e_k^2 \quad (1)$$

uz ograničenja jednakosti:

$$y_k [w^T \varphi(x_k) + b] = 1 - e_k, \quad k = 1, \dots, N \quad (2)$$

gde je φ nelinearna funkcija mapiranja, koja preslikava skup ulaznih vektora u višedimenzionalan prostor. Težinski vektor hiperravni označen je sa w , dok je b skalarni pomeraj odnosno težinski prag. Promenljiva e_k , predstavlja dozvoljene greške klasifikacije, dok parametar C kontroliše process, odnosno odnos između složenosti modela i prihvatljive greške klasifikacije.

Nakon rešavanja optimizacionog problema definisanog sa (1) i (2), rešenje se može naći u [7], funkcija razdvajanja LS-SVM klasifikatora se definiše kao:

$$y(x) = \text{sign} \left[\sum_{k=1}^N \alpha_k y_k K(x, x_k) + b \right] \quad (3)$$

gde α_k predstavljaju podržavajuće vektore (odnosno Langranžove multiliktore), a b je konstanta. $K(x, x_k)$ predstavlja funkciju jezgra (Kernel function), koja je definisana skalarnim proizvodom između x i x_k .

Kako je navedeno u [6] na osnovu dvadeset različitih skupova podataka najbolju opštu stopu predviđanja dali su LS-SVM klasifikatori sa RBF (Radial basis function) kernelom. Pored toga prema [8] u slučajevima kada je broj primera za klasifikaciju mnogo veći od broja atributa (dimenzija vektora - prostora) takođe se preporučuje korišćenje RBF kernela. Stoga je u radu prilikom formiranja modela korišćen RBF kernel, definisan sa:

$$K(x, x_k) = \exp \left\{ -\frac{\|x - x_k\|_2^2}{\sigma^2} \right\} \quad (4)$$

Prilikom treniranja LS-SVM modela potrebno je odrediti vrednost skalirajućeg parametra C , kao i parametre izabranog jezgra, u ovom slučaju širinu σ . Jedan od načina za određivanje ovih parametara je postupak n unakrsnih validacija (eng. k fold Cross - Validation) u kombinaciji sa pretraživanjem po koracima (eng. Grid - Search), detaljno opisan u [8].

III. FORMIRANJE MODELA

Podaci, koji su korišćeni u ovom modelu, preuzeti su sa sajta Beorgaske berze (www.belex.rs). Serija obuhvata zapise od 04. aprila 2005. do 30. septembra 2013. godine u ukupnom iznosu od 2.012 trgovinskih dana i sastoji se od šest veličina koje se utvrđuju za svaki dan: vrednost indeksa, procentualna

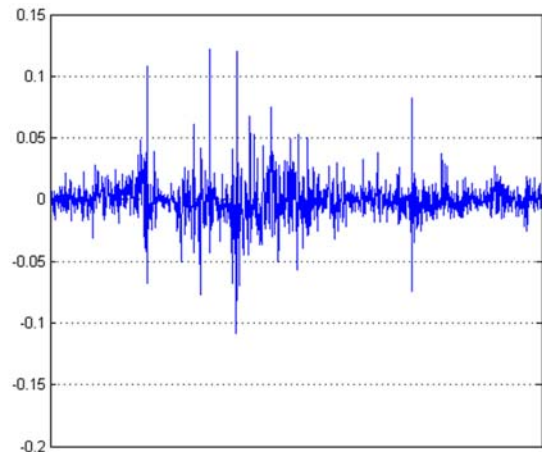
promena vrednosti indeksa u odnosu na prethodni dan trgovanja, vrednost indeksa na otvaranju berze, najviša dnevna vrednost indeksa, najniža dnevna vrednost indeksa i promet. Analitičari prate promene vrednosti i prometa kako bi doneli odluke o daljoj strategiji trgovanja.

U formiranju modela koristi se logaritamski prinos na berzanski indeks Belex15 i eksponencijalni pokretni prosek (eng. Exponential Moving Average - EMA), koji se obračunava na bazi logaritamskih prinosa na Belex15. Razlog za selekciju EMA među ostalim tehničkim indikatorima je njena karakteristika da daje veći značaj skorijim promenama cene, kao i mogućnost računanja EMA vrednosti unazad do skoro neograničenog broja koraka (npr. EMA120, EMA220), što je bitna karakteristika kod modelovanja vremenskih serija [5].

Logaritamski prinos se izračunava na osnovu dnevne vrednosti indeksa primenom sledeće formule:

$$r(t) = \ln \left[\frac{P(t)}{P(t-1)} \right] = \ln[P(t)] - \ln[P(t-1)] \quad (5)$$

gde upotrebljeni simboli imaju sledeće značenje: $r(t)$ - logaritamski prinos, $P(t)$ - vrednost indeksa na dan trgovanja t , $P(t-1)$ - vrednost indeksa za prethodni dan trgovanja. Navedena transformacija doprinosi stacionarnosti serije, a ujedno se vrši i normalizacija podataka na opseg $[-1, 1]$, što dodatno povećava efikasnost algoritama mašinskog učenja. Slika 1 predstavlja izgled logaritamskog prinosa za dati period posmatranja.



Slika 1. Prikaz vrednosti logaritamskog prinosa na Belex15 u periodu od 2005. - 2013. godine.

Eksponencijalni pokretni prosek (EMA) se izračunava na sledeći način:

$$EMA(t) = r(t) * k + EMA(t-1) * (1-k) \quad (6)$$

gde upotrebljeni simboli imaju sledeće značenje: $EMA(t)$ - eksponencijalni pokretni prosek u trenutku t , $EMA(t-1)$ - eksponencijalni pokretni prosek u trenutku $t-1$, $r(t)$ - logaritamski prinos u trenutku t , $k=2/(N+1)$, N - vremenski period za koji se obračunava EMA.

Prethodnom skupu je zatim dodan još jedan atribut koji služi kao oznaka klase za predviđanja pravca kretanja vrednosti indeksa. Ovaj atribut se naziva i indikator i zavisi od ostalih dostupnih atributa U našem eksperimentu kao indikator korišćen je logaritamski prinos na sledeći način:

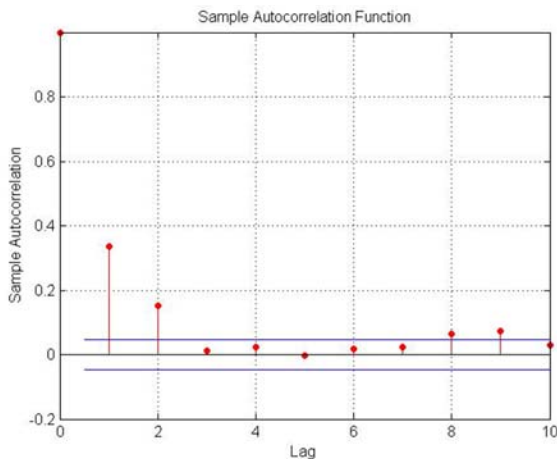
$$y_t = \begin{cases} -1 & r(t) < 0 \\ 1 & r(t) > 0 \end{cases} \quad (7)$$

Pravac (trend) je, dakle, kategorička promeljiva koja prikazuje pravac kretanja indeksa Belex15 u bilo kom trenutku vremena t.

Izbor atributa koji čine model izvršen je i po vremenskom okviru. Prema [9], analiza trenda vrednosti berzanskih indeksa i cena hartija od vrednosti u kraćim vremenskim periodima (od 5 do 25 dana) rezultira pokazateljima koji adekvatnije mere osetljivost promene vrednosti. Vremenske sekvence dodatno prate i petodnevni režim rada Beogradske berze.

Poznata je važnost uključivanja prethodnih vrednosti prinosa kako bi se predvideli budući pravci kretanja tržišta u slučaju da serije prinosa sadrže visok stepen autokorelacije. Da bi se izabrale ulazne karakteristike sa visokom vremenskom korelacijom izračunata je autokorelacija i parcijalna korelacija logaritamskih prinosa u prethodnih 5, 10, 15 i 20 trgovinskih dana na berzi. Korelogramom na Slici 2 prikazane su korelacije između tekuće vrednosti i pojedinih vrednosti vremenske serije ostvarene preko autokorelacione funkcije do 10 prethodnih vrednosti. Dobijene vrednosti autokorelacionih koeficijenata nisu visoke, ali su statistički značajne. Može se uočiti da vrednosti autokorelacione funkcije naglo opadaju u prva tri koraka i da dominiraju prva i druga prethodna vrednost logaritamskog prinosa. Zbog toga su atributima modela dodate vrednosti logaritamskih prinosa za prvu i drugu prethodnu vrednost.

Na osnovu Slike 2 može se uočiti da tačke podataka koje su bliže današnjem trenutku cena imaju veći uticaj na današnju cenu, tako da je i za period obračunavanja EMA transformacije uzeto prethodnih 10 dana.



Slika 2. Korelogram logaritamskih prinosa do 10 prethodnih vrednosti.

Na osnovu prethodnih analiza kreiran je sledeći predikcioni model:

$$y_t = f(r(t-1), r(t-2), EMA10(t-1)) \quad (8)$$

Za formiranje modela korišćena je Matlab biblioteka LS-SVMlab [10]. Rezultati testiranja su predstavljani u narednom delu.

IV. REZULTATI TESTIRANJA

Raspoloživi podaci su podeljeni u dve grupe, prvu grupu čini 1811 zapisa predviđenih za treniranje modela, od 26. oktobra 2005. do 31. decembra 2012. godine, dok su za testiranje i evaluaciju rezervisani podaci od 3. januara 2013. do 01. oktobra 2013. godine, ukupno 187 dana trgovanja

Kao opšta mera za procenu performansi predviđanja na test skupu koristi se stopa pogodaka (eng. Hit Ratio - HR) koja se izračunava na osnovu broja pravilno klasifikovanih rezultata u okviru test skupa:

$$HR = \frac{1}{m} \sum_{i=1}^m PR_i \quad (9)$$

Gde je PR predikcioni rezultat i-tog trgovinskog dana, definisan sledećom jednačinom.

$$PR_i = \begin{cases} 1 & PI_i = AI_i \\ 0 & PI_i \neq AI_i \end{cases} \quad (10)$$

pri čemu je AI_i aktuelni izlaz za i -ti trening dan i PI_i je predviđena vrednost za i -ti dan trgovanja, dok je m broj podatka u test skupu [3].

LS-SVM parametri C i σ određeni su procesom 10 unakrsnih validacija uz pomoć grid-search tehnike, čime se uspešno prevazilazi problem overfitting-a modela. Nakon trening faze, uspešnost modela je proverena nad test podacima.

U Tabeli I predstavljene su stope pogodaka modela po mesecima u dostupnom test skupu podataka zajedno sa brojem dana koji u testiranom periodu imaju rastući i opadajući trend.

TABELA I. STOPA POGODAKA I BROJ DANA TRENDU PO MESECIMA

Mesec (2013. godina)	Hit Rate (HR)	Broj dana	
		Rastući trend	Opadajući trend
Januar	0.700	14	6
Februar	0.526	10	9
Mart	0.619	16	5
April	0.591	10	12
Maj	0.684	5	14
Jun	0.600	6	14
Jul	0.522	11	12
Avugust	0.545	14	8
Septembar	0.571	12	9

Stopa pogodaka prediktora na nivou celog test skupa je 0.568. Na osnovu Tabele I može se definisati opseg kretanje stope preciznosti od 0.522 do 0.700. Iako je dostupni test skup izbalansiran, ukupan broj dana u test skupu sa rastućim i opadajućim trendom je redom 98 i 89, u konkretnim periodima posmatranja dolazi do pojava neizbalansiranih sekvenci podataka. U izbalansiranom test skupu podataka stope pogodaka su bliže stopi pogotka na celom test skupu. Dok je model nestabilnijih performansi na neizbalansiranom skupu podataka. Na nivou celog test skupa ipak je stopa pogodaka u zapisima sa rastućom promenom trenda 0.55, dok je stopa pogodaka u danima opadajuće promene trenda 0.58.

Oscilacije preciznosti predviđanja mogu se dodatno objasniti redovnim i vanrednim revizijama korpe berzanskog indeksa, kao i obelodanjivanjem informacija o poslovanju preduzeća, čije akcije čine indeksu korpu, usled čega dolazi do promena u vrednosti indeksa i obima trgovanja. U periodima rasta prinosa investitori su zainteresovani za ulaganja, međutim, u periodima pada prinosa investitori menjaju svoje strategije i finansijska sredstva realociraju u relativno sigurne investicione alternative.

Predloženi predikcioni model dodatno je testiran u vremenskom periodu od 5, 10, 15 i 20 dana od ranije prepoznatih kritičnih momenta na Beogradskoj berzi usled kvartalnog preračunavanja berzanskog indeksa. U Tabeli II prikazani su rezultati testiranja.

TABELA II. STOPA POGODAKA PO DANIMA KRITIČNIH PERIODA

Period testiranja (2013. godina)	Sekvence u danima				
	5	10	15	20	ceo opseg
15. mart – 15.april	0.60	0.55	0.66	0.62	0.53
15. jun -15. jul	0.60	0.61	0.60	0.55	0.52

Na osnovu Tabele II može se uočiti da obračunate stope pogodaka u graničnim slučajevima ostaju u opsegu prethodno definisanih vrednosti prediktora. Dobijeni rezultati potvrđuju stabilnost predloženog modela.

S obzirom na sve navedene osobenosti berzanskog tržišta, dobijene vrednosti stope pogodaka spadaju u očekivani opseg preciznosti i komparativne su sa rezultatima prikazanim u drugim studijama [1-2], [5],[11].

V. ZAKLJUČAK

Primenom predloženog modela dobijaju se konkurentni rezultati predviđanja pravca kretanja vrednosti berzanskih indeksa. Naročito je značajno uzeti u obzir da se u ovom radu ispituje stepen predvidljivosti kretanja indeksa cena akcija na tržištu u razvoju, kao što je tržište kapitala Republike Srbije, nasuprot većini radova iz ove oblasti, koji se bave predviđanjem indeksa cena akcija na razvijenim tržištima.

U nastavku istraživanja bi se moglo definisati više pravaca koji bi mogli dovesti do poboljšanja preciznosti modela. Najpre bi se moglo razmatrati uvođenje novih atributa iz skupa

tehničkih indikatora, npr. momentum-a, a zatim i uvođenje makroekonomskih pokazatelja. Pored toga, od interesa je formiranje takozvanih kombinovanih modela predviđanja, gde bi se izlazi iz više modela kombinovali u finalnom modelu. Na kraju, važno je napomenuti da se u predstavljenom modelu predviđanja, svako povećanje efikasnosti od 1% u odnosu na trenutne domete smatra izuzetnim doprinosom, dok se na tržištu preslikava u visoke novčane dobitke [11].

LITERATURA

- [1] W. Huang; Y. Nakamori and SY. Wang, "Forecasting stock market movement direction with support vector machine," Computers & Operations Research vol. 32, 2005, pp. 2513–2522
- [2] Y. Kara, M. A. Boyacioglu b, Ö. K. Baykan, „Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange“,Expert systems with Applications, vol. 38, no. 5, 2011, pp. 5311-5319
- [3] M. Kumar, M. Thenmozhi, "Forecasting stock index movement: a comparasion of support vector machines and random forest" Indian Institute of Capital Markets 9th Capital Markets Conference Paper, 2006, available at SSRN: <http://ssrn.com/abstract=876544> or <http://dx.doi.org/10.2139/ssrn.876544>, 2006.
- [4] O. Phichhang and H. Wang, "Prediction of Stock Market Index Movement by Ten Data Mining Techniques" Modern Applied Science, Vol. 3, No. 12, 2009, pp. 28-42.
- [5] S. Lahmiri, "A Comparison of PNN and SVM for Stock Market Trend Prediction using Economic and Technical Information", International Journal of Computer Applications, vol. 29, no.3, 2011.
- [6] T.V. Gestel, J. A.K. Suykens at all. "Benchmarking Least Squares Support Vector Machine Classifiers", Machine Learning, 54, 5–32, 2004
- [7] J. Suykens and J. Vandewalle, "Least Squares Support Vector Machines", Neural processing letters, vol. 9, no. 3, 1999, pp. 293-300
- [8] M. Božić, Z. Stajić i M. Stojanović, "Kratkoročno predviđanje električnog opterećenja primenom metoda podržavajućih vektora", Infoteh Jahorina, vol. 10, 2010, pp. 326-329
- [9] A. Bradić-Martinović, „Predviđanje cena akcija pomoću tehničke analize“, Economic Anals, no 170, 2006
- [10] K. De Brabanter, P. Karsmakers, F. Ojeda, C. Alzate, J. De Brabanter, K. Pelckmans, B. De Moor, J. Vandewalle, J.A.K. Suykens "LS-SVMlab Toolbox User's Guide", ESAT-SISTA Technical Report 10-146, 2011, <http://www.esat.kuleuven.be/sista/lssvmlab/>
- [11] L. Yuling, H. Guo and J. Hu, "An SVM-based Approach for Stock Market Trend Prediction", Neural Networks (IJCNN), The 2013 International Joint Conference on. IEEE, 2013, pp. 1-7.

ABSTRACT

The paper proposes a trend prediction model of BELEX15 stock market values using the Least Squares Support Vector Machines (LS-SVMs) for classification. The feature selection was based on the analysis of technical indicators. The features that are used for model training are derived directly from the logarithmic returns and calculation of exponential moving average (Exponential Moving Average - EMA). The test results showed that the proposed model is suitable for short-term stock market trend prediction.

STOCK EXCHANGE TREND PREDICTION OF BELEX15 INDEX WITH LS-SVM CLASSIFIER

Ivana Marković, Jelena Stanković, Miloš Stojanović, Miloš Božić