

PREDVIĐANJE RASTA BDP-A PRIMENOM RUDARENJA PODATAKA NA PRIMERU SRBIJE

Rade Božić

Univerzitet u Istočnom Sarajevu, Fakultet poslovne ekonomije Bijeljina,

Republika Srpska, Bosna i Hercegovina

rade.bozic@fpe.ues.rs.ba

ORCID: 0000-0001-6956-8049

Članak je izlagan na X Internacionalnom naučnom skupu „EkonBiz: Ekonomski izazovi u uslovima ubrzanih globalnih promjena“, Bijeljina 16. i 17. jun 2022. godine

Apstrakt: Predviđanje ishoda različitih pojava oduvek je predstavljalo atraktivnu istraživačku temu u velikom broju naučnih disciplina, naročito u ekonomiji. Kao posebna oblast, ekonometrija pruža različite modele za predviđanje pokazatelja kao što su BDP, stopa inflacije, kamatna stopa, cena raznih roba i usluga, kao i mnogih drugih kako na mikro tako i na makro nivou. Razvoj informacionih tehnologija omogućio je obavljanje računskih operacija mnogo brže i pouzdanije. Međutim, poseban doprinos se ogleda kroz primenu rudarenja podataka u svrhu ekstrahovanja relevantnih informacija iz velikog skupa podataka. Modeli koji se razvijaju primenom rudarenja podataka pružaju dobre rezultate u predviđanju ekonomskih pokazatelja, često i uspješnije u odnosu na određene ekonometrijske modele. U ovom radu nastoji se predvideti promena (rast) BDP-a kroz primenu rudarenja vremenskih serija na primeru Republike Srbije. Analiza je obavljena u dva slučaja: u jednom modelu uključuju nezavisne atribute koji dodatno opisuju zavisnu promenljivu, dok drugi slučaj ne sadrži ove atribute. Upotrebljene su tri različite metode rudarenja u oba slučaja (linearna regresija, višeslojni perceptron i metoda slučajne šume), a dobijeni rezultati uspešnosti modela su predstavljeni i protumačeni.

Ključne riječi: rudarenje podataka, predviđanje, rast BDP-a

JEL klasifikacija: E47, O11

1. UVOD

Makroekonomski pokazatelji odražavaju stanje nacionalne ekonomije i predmet su istraživanja velikog broja radova iz različitih naučnih disciplina. Uzrok toga je njihov široki uticaj na razne društvene oblasti i aspekte, a ne samo one u ekonomskoj sferi. Ekonometrija kao naučna disciplina nastoji da statističkim tehnikama i ekonomskom teorijom istraži pomenute pokazatelje, često predviđajući njihovo kretanje ili ishod pod uticajem odabranih faktora (Stock & Watson, 2015). Pomoć u ovome pružaju informacione tehnologije kao alat za obavljanje računskih operacija. Međutim, u okviru njih poseban doprinos pruža rudarenje podataka (eng. *Data Mining*). To je proces koji se odnosi na prikupljanje, pripremu, obradu, analizu i ekstrahovanje korisnih informacija. Primenu pronalazi u različitim domenima i posmatra se kao širok aspekt koji generalno opisuje obradu podataka (Aggarwal, 2015). Obuhvata nadgledane (klasifikacija i regresija) i nenadgledane tehnike učenja (klasterisanje i pravila pridruživanja) a primenu pronalazi u različitim područjima kao što su analiza satelitskih snimaka i teksta, predviđanje vremenskih prilika, medicinska dijagnostika, predviđanje potrošnje električne energije, automatsko apstrahovanje, analiza organskih jedinjenja, opasnosti od zagađenosti itd. Pored navedenog, popularna je i primena u ekonomiji kao što je npr. marketing targetiranje, procena vrednosti nekretnina, dizajniranje proizvoda, finansijsko predviđanje, detektovanje prevara sa kreditnim karticama i sl (Brammer, 2016). Rudarenje vremenskih serija (eng. *time series*)

data mining) je posebna oblast i odnosi se na mogućnosti izdvajanja informacija iz skupa podataka koji poseduje vremensku dimenziju. One imaju vremensku oznaku i prikupljane su sa određenom frekvencijom. Primeri takvih podataka su prodaje na mesečnom nivou, berzanska trgovina na sedmičnom nivou, posete veb-sajtova evidentirane tokom svakog sata itd.

Ovakva vrsta podataka igra važnu ulogu u poslovanju preduzeća, naročito u postupku analize. Druga standardna aktivnost je predviđanje budućeg ponašanja (eng. *forecasting*). Vremenske serije podataka mogu da se prikupljaju ručno od strane čoveka, međutim, čest je slučaj i njihovog pohranjivanja putem različitih mašina ili uređaja (Dean, 2014).

Bruto domaći proizvod (skraćeno BDP) predstavlja vrednost finalnih proizvoda i usluga proizvedenih u jednoj privredi tokom određenog vremenskog perioda (Blanchard, 2008). U ovom radu se nastoji predvideti rast BDP-a kao makroekonomske varijable upotrebom različitih metoda rudarenja vremenskih serija podataka. Sam ovaj pokazatelj predstavlja ekonomski rast države i od posebne je važnosti za međunarodne investitore i poverioce, a pokazuje promenu BDP u poređenju sa prethodnom kalendarskom godinom ili drugim vremenskim periodom.

U postupku analize korišten je WEKA softverski alat. Predviđanje u ovom radu obavljeno je u dva različita slučaja. Prvi slučaj obuhvatao je modele sa vremenskom komponentom i samo jednom varijablom koja se ujedno i predviđa (rast BDP-a). Ovde je izvršeno i predviđanje van odabranog

skupa za jedan kvartal unapred (prvi kvartal 2022. godine). Drugi slučaj se odnosi na modele koji su predviđanje obavili unutar prikupljenog skupa, a pored zavisne varijable (rast BDP-a) uključene su još i odabrane nezavisne varijable koje dodatno opisuju predviđanu pojavu.

One su birane iz originalno prikupljenog skupa od 14 ekonomskih varijabli (*overlay data* u WEKA alatu) u periodu od 2009. do 2021. godine na kvartalnom nivou. Kod oba načina upotrebene su tri različite metode rudarenja podataka radi međusobne komparativne analize, a rezultati do kojih se došlo su predstavljani i protumačeni u posebnom poglavlju.

2. SKUP PODATAKA

Za predviđanje porasta BDP-a u drugom slučaju prikupljen je skup od 14 različitih makroekonomskih varijabli na osnovu analize prethodno napisanih radova. Oni su se odnosili na predviđanje rasta BDP-a ili na analizu ekonomskih indikatora koji na njega utiču.

U postupku analize primenjivale su se različite kombinacije ovih atributa, a najbolji rezultati su prikazani i predstavljeni. Kako su podaci za zemlje u tranziciji nedovoljno dostupni, za ovaj rad uzete su samo one varijable koje se najčešće koriste za analizu ovog problema. Prema radovima Carreiro, Galvao i Kapetanios (2019), Carreiro, Clark i Marcellino (2015) (2019), Schorfheide i Song (2014), Smets, Warne i Wouters (2014) odabrani su sledeći indikatori koji su prikupljeni na kvartalnom nivou (tabela br. 1):

Tabela 1. *Varijable korištene u postupku analize*

| R.br. | Naziv varijable | Jedinica mere |
|-------|--|---------------|
| 1. | Rast BDP-a u odnosu na prethodni kvartal | % |
| 2. | Prihod u industriji i građevinarstvu u odnosu na prethodni kvartal | % |
| 3. | Prihod u saobraćaju i skladištenju u odnosu na prethodni kvartal | % |
| 4. | Prihod informisanje i telekomunikacije u odnosu na prethodni kvartal | % |
| 5. | Izdaci za finalnu potrošnju domaćinstava | mil. RSD |
| 6. | Izdaci za finalnu potrošnju države | mil. RSD |
| 7. | Bruto investicije u osnovna sredstva | mil. RSD |
| 8. | Uvoz robe i usluga | mil. RSD |
| 9. | Izvoz robe i usluga | mil. RSD |
| 10. | Promene u zalihama | mil. RSD |
| 11. | Prosečne bruto zarade | RSD |
| 12. | Industrijska proizvodnja - lančani indeks | % |
| 13. | Devizni kurs dinara prema dolaru | RSD |
| 14. | Građevinske dozvole za stanove | broj |

Izvor: varijable odabrane od strane autora

BDP se najčešće razlaže na sledeće kategorije: ličnu potrošnju, investicije (nestambene i stambene), državnu potrošnju, neto izvoz i investicije u zalihe (Blanchard, 2008). U datom skupu podataka ove kategorije su obuhvaćene sa barem jednim parametrom, što se podudara sa varijablama odabranim od strane prethodno pomenutih autora. Svi podaci su preuzeti sa sajta Republičkog zavoda za statistiku Srbije (2022, <https://www.stat.gov.rs>) osim deviznog kursa koji je preuzet sa *Investing* veb-prezentacije (2022, <https://www.investing.com/currencies/usd-rsd-historical-data>). Podaci vezani za novčane iznose iskazani su u milionima dinara (RSD).

Podaci u ovom radu su prikupljeni na kvartalnom nivou zbog nedovoljne dostupnosti godišnjih podataka. Sa ovom vrstom problema se susretao veliki broj istraživača koji su pokušali da primene prognoziranje ekonomskih kretanja u zemljama u tranziciji, naročito u slučajevima gde su korištene makroekonomske varijable.

Skup podataka obuhvata period od prvog kvartala 2009. do četvrtog kvartala 2021. godine i u njemu nema podataka koji nedostaju. Varijable u skupu su prikazane kroz period od 13 godina, što ukupno čini 52 observacije (instance). Podaci prikupljeni na mesečnim nivoima su izraženi na kvartalnom nivou. Uzrok ovakvog kratkog perioda obuhvata nepostojanje podataka vezanih za protekli period. Kao zavisna varijabla odabran je rast BDP-a koji je iskazan u procentualnom iznosu. U poslednje dve godine obuhvaćenih podataka svet se susreo sa ekonomskom krizom izazvanom pandemijom koronavirusa, što se odrazilo i na Republiku Srbiju. Ovo je bitno navesti iz razloga jer se privreda našla pod uticajem nepredviđenih faktora što je dovelo do pada BDP-a od 9,2% u drugom kvartalu 2020. godine. Pre same primene odabranih metoda podaci koji nisu iskazani procentualno su normalizovani, odnosno svedeni na interval od 0 do 1. Na ovaj način je omogućeno njihovo upoređivanje iako nisu iskazani u istim jedinicama mere (Aggarwal, 2015).

3. PREDVIĐANJE VREMENSKIH SERIJA U RUDARENJU PODATAKA, METODE I EVALUACIONI PARAMETRI

Vremenske serije podataka predstavljaju vremenski označen skup prikupljen prema određenoj frekvenciji (Dean, 2014). Za razliku od drugih skupova, ova vrsta uključuje i vremensku dimenziju koja mora biti razmatrana. Neki od primera vremenskih serija su dnevne cene hartija od vrednosti na berzi, mesečni broj poseta na veb-sajtu, sedmični promet automobila na auto-putu, dnevna emisija štenih gasova itd. Vremenske serije mogu biti stacionarne ili nestacionarne. Stacionarni

stohastički proces je onaj čiji se parametri (poput srednje vrednosti i varijanse) ne menjaju tokom vremena, dok su nestacionarni oni kod kojih su promene prisutne. Vremenske serije koje sadrže kretanje samo jedne varijable kroz vremenski period nazivaju se univarijabilne, dok one koje obuhvataju više varijabli se nazivaju multivarijabilne (Aggarwal, 2015).

Vremenske serije se često javljaju u ekonomskom domenu, naročito kod poslovanja preduzeća.

Takve vrste informacija predstavljaju osnovu za donošenje odluka menadžmenta. Međutim, ove serije imaju važnu ulogu i u makroekonomiji, naročito kada je reč o predviđanju kretanja određene pojave.

Predviđanje predstavlja jednu od najčešćih primena analize vremenskih serija a često se koristi u maloprodaji, ekonomskim indikatorima, berzama, predviđanju vremenskih prilika, kao i u drugim slučajevima. Cilj je predvideti buduću vrijednost određene varijable na osnovu njenih prethodnih vrednosti (Aggarwal, 2015). Metode korištene u ovom radu su linearna regresija (eng. *linear regression*), višeslojni perceptron (eng. *multilayer perceptron neural networks*, MLP) i metoda slučajne šume (eng. *random forest*, RF).

Linearna regresija: kada je izlaz iz modela numerčki iskazan, kao i svi ulazni atributi, prirodno je razmotriti linearnu regresiju kao metodu koju je potrebno primeniti. Ovo je osnovni metod u statistici. Ideja je iskazati zavisnu promenljivu kao linearnu kombinaciju atributa sa unapred određenim koeficijentima:

$$x = w_0 + w_1 a_1 + w_2 a_2 + \dots + w_k a_k$$

gde x predstavlja zavisnu promenljivu, a_1, a_2, \dots, a_k vrednosti atributa, dok w_0, w_1, \dots, w_k predstavljaju koeficijente koji se računaju na osnovu podataka iz skupa namenjenih za trening modela.

Važno je napomenuti da predviđena vrednost zavisne promenljive nije njena stvarna vrednost. Razlika između stvarnih i predviđenih vrednosti govori o uspešnosti modela.

Za odabir koeficijenata koristi se metod najmanjih kvadrata. Najbolji koeficijenti su oni čiji je zbir kvadrata odstupanja predviđene u odnosu na stvarnu vrednost najmanji u poređenju sa ostalim koeficijentima.

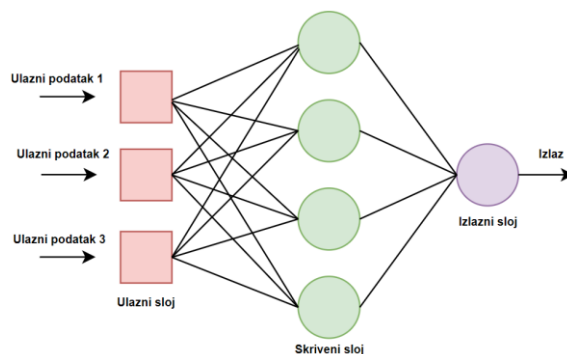
Performanse ovog metoda mogu biti umanjene ako su svi ulazni atributi u visokom stepenu korelacije. Često se koristi kao početna osnova za razvoj drugih metoda (Witten, Frank, Hall, & Pal, 2017).

Višeslojni perceptron: predstavlja veštačku neuronsku mrežu najčešće korištenu u praksi. MLP

ima tri osnovna sloja (slika 1): prvi sloj sadrži ulazne podatke za obradu, drugi sloj je „skriveni” sloj gde neuroni stvaraju težinske koeficijente i

koriste aktivacionu funkciju kako bi se dobio izlazni sloj koji zavisi od aktivnosti u prethodnom skrivenom sloju (Tsai & Wu, 2008).

Slika 1. MLP šema



Izvor: (Menzies, Kocagüneli, Peters, & Turhan, 2015)

Perceptroni sa više od jednog sloja koji dodeljuju težinske koeficijente nazivaju se višeslojni perceptroni (Kriesel, 2007). Ovaj metod pripada kategoriji nadgledanog učenja iz razloga što su izlazne vrednosti poznate MLP-u. On ima dva osnovna svojstva: poseduje nelinearne karakteristike, a jaka međuzavisnost određuje snažne veze između slojeva (Erkam, Kayakutlu, & Daim, 2011). MLP neuronska mreža zasnovana je na nepovratnoj arhitekturi, a obučava se koristeći algoritam propagacije unazad (eng. *back propagation*) (Nemes & Butoi, 2013). U radu je korišten MLP sa jednim skrivenim slojem i koeficijentom učenja od 0,3. Momentum je iznosio 0,2 a maksimalni broj epoha za učenje je 500. Kod slučaja bez uključenih dodatnih atributa ulaz predstavlja 11 ulaznih varijabli, dok skriveni sloj ima jedan neuron. Kod slučaja sa uključenim dodatnim atributima bilo je 15 ulaznih varijabli dok je u skrivenom sloju bio ponovo jedan neuron.

Metoda slučajne šume: predstavlja kombinaciju stabla odlučivanja takvu da svako stablo zavisi od vrednosti slučajnog vektora nezavisno uzorkovanog sa istom distribucijom za sva stabla u šumi. Generalizaciona greška za šumu konvergira do granice kako broj stabala u šumi postaje veći. Greška zavisi od jačine pojedinačnih stabala i korelacije među njima (Breiman, 2001). Prosečna vrednost predviđanja slučajne šume se dobija uz pomoć *bootstrap* agregacije i selekcije slučajnih varijabli. Pokazalo se da je metod slučajne šume robusan prediktor za male veličine uzoraka i

podatke velikih dimenzija (Gérard & Scornet, 2016).

Za evaluaciju modela korišteni su uobičajeni parametri: prosečna apsolutna greška (MAE), prosečna kvadratna greška (eng. *mean squared error*, MSE), prosečna apsolutna procentualna greška (MAPE), kvadratni koren prosečne kvadratne greške (RMSE). Formule su prikazane ispod:

$$MAE = \frac{1}{N} \sum_{i=1}^N |\bar{y}_i - y_i|$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (\bar{y}_i - y_i)^2$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \left| \frac{\bar{y}_i - y_i}{y_i} \right|$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\bar{y}_i - y_i)^2}$$

U navedenim formulama \bar{y}_i predstavlja predviđenu vrednost dok y_i predstavlja stvarnu vrednost za N broj opservacija. MAE predstavlja prosečnu apsolutnu razliku između vrednosti koje je predvidio model i posmatranih istorijskih

podataka, dok MSE ukazuje na prosek kvadrata njihove razlike. MAPE je prosečna apsolutna procentualna razlika između izlaza modela i stvarnih vrednosti, dok RMSE predstavlja kvadratni koren MSE pokazatelja.

4. POSTUPAK ANALIZE I REZULTATI

Postupak analize sproveden je pomoću WEKA alata, odnosno njegove ekstenzije namenjene za predviđanje vremenskih serija podataka. Ovaj softver je razvijen od strane univerziteta u Waikatu (Novi Zeland) i predstavlja skup algoritama za mašinsko učenje. Korištena verzija je 3.8.6. Postupak analize sproveden je na dva različita načina. Prvi način obuhvata predviđanje bez korištenja dodatnih varijabli koje objašnjavaju rast BDP-a, dok drugi način podrazumeva njihovo uključivanje u modele. Prva analiza je sprovedena zbog mogućnosti predviđanja buduće vrednosti zavisne varijable izvan testnog seta. U ovom radu odabran je jedan kvartal unapred, odnosno prvi kvartal u 2022. godini. Kada se u model uključe i dodatne varijable, onda u WEKA alatu nije moguće izvršiti predviđanje za period unapred van prikupljenog seta. Razlog tome je što dodatne nezavisne varijable (ne predviđaju se) moraju biti poznate u budućnosti kako bi se predviđanje

izvršilo. Međutim, one se uključuju u model putem simulacije predviđanja u okviru postojećeg skupa na taj način što se isključuje predviđanje van njega. Ukoliko se ova opcija ne isključi, onda će program javiti grešku. Kod oba slučaja korišteni su isti metodi, a sledeća podešavanja su zajednička: period je podešen na kvartalni interval dok je koeficijent pouzdanosti od 95% standardno određen u softverskim podešavanjima i nije menjan. Za trening modela korišteno je 80% instanci skupa, dok je poslednjih 20% korišteno za evaluaciju modela. Potrebno je napomenuti da u 20% poslednjih instanci spada 10 kvartala gde je u 8 vladala pandemija koronavirusa (od 2020. godine).

4.1 Predviđanje rasta BDP-a bez dodatnih varijabli

Za vremensku oznaku odabrano je korištenje veštačkog vremenskog indeksa (eng. *artificial time index*) gde WEKA automatski dodeljuje vremenske oznake vrstama, dok je jedan kvartal odabran kao vremenski interval za koji će se obaviti predviđanje van skupa. U tabeli 2. su prikazane predviđene, stvarne vrednosti i njihova razlika (greška), dok su dobijeni rezultati evaluacije sva tri primenjena metoda prikazani u tabeli 3.

Tabela 2. Prikaz stvarnih i predviđenih vrednosti primenjenih metoda

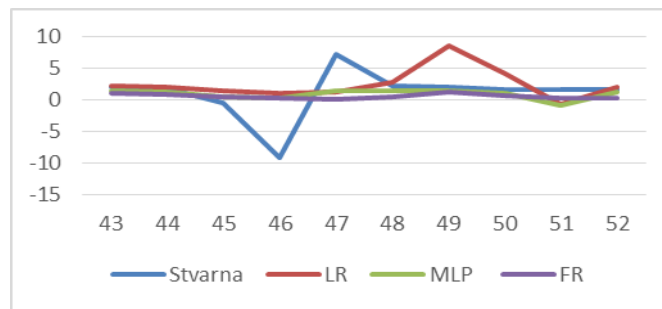
| Kvartal | Stvarna vrednost | LR | | MLP | | RF | |
|---------|------------------|---------------------|---------|---------------------|---------|---------------------|---------|
| | | Predviđena vrednost | Razlika | Predviđena vrednost | Razlika | Predviđena vrednost | Razlika |
| 43 | 2,1 | 2,2452 | 0,1452 | 1,4259 | -0,6741 | 1,1558 | -0,9442 |
| 44 | 1,8 | 1,9853 | 0,1853 | 1,2905 | -0,5095 | 0,96 | -0,84 |
| 45 | -0,5 | 1,4892 | 1,9892 | 0,5021 | 1,0021 | 0,572 | 1,072 |
| 46 | -9,2 | 1,1762 | 10,3762 | 0,27 | 9,47 | 0,3806 | 9,5806 |
| 47 | 7,2 | 1,2644 | -5,9356 | 1,4306 | -5,7694 | 0,2284 | -6,9716 |
| 48 | 2,2 | 2,7704 | 0,5704 | 1,4581 | -0,7419 | 0,5543 | -1,6457 |
| 49 | 2 | 8,6321 | 6,6321 | 1,4562 | -0,5438 | 1,2295 | -0,7705 |
| 50 | 1,7 | 4,1812 | 2,4812 | 1,0395 | -0,6605 | 0,7495 | -0,9505 |
| 51 | 1,7 | -0,5497 | -2,2497 | -0,8383 | -2,5383 | 0,3022 | -1,3978 |
| 52 | 1,7 | 2,0963 | 0,3963 | 1,3076 | -0,3924 | 0,3812 | -1,3188 |
| 53* | - | 2,5828 | - | 1,4075 | - | 0,4173 | - |

Izvor: istraživanje autora

U tabeli broj 2. prikazane su stvarne i predviđane vrednosti prema metodama, kao i njihova razlika, odnosno greška. Pored 10 kvartala na kojima se vršila evaluacija modela (43-52 kvartal), prikazan je i 53 kvartal (označen zvezdicom) koji predstavlja prvi kvartal 2022. godine, odnosno predviđani kvartal van prikupljenog originalnog skupa. Pošto stvarni podaci za ovaj kvartal još nisu

dostupni u javnosti, ta vrednost je izostavljena kao i razlika između nje i predviđene. Sva tri metoda su predvidela rast BDP-a u ovom kvartalu, najviše LR od 2,5828%, dok je najmanji rast predvidio RF algoritam u iznosu od 0,4173%. Tabelarni podaci su prikazani i na grafikonu 1, odnosno predviđene i stvarne vrednosti rasta BDP-a:

Grafikon 1. odnos stvarnog i predviđenog rasta BDP-a



Izvor: istraživanje autora

Tabela 3. Prikaz grešaka kod primenjenih metoda

| Greške | Metode | | |
|--------|----------|----------------|---------|
| | LR | MLP | RF |
| MAE | 3,0961 | 2,2302 | 2,5492 |
| MAPE | 126,9408 | 71,6045 | 83,6026 |
| RMSE | 4,501 | 3,6409 | 3,8863 |
| MSE | 20,2593 | 13,2564 | 15,1036 |

Izvor: istraživanje autora

U tabeli 3. prikazani su podaci koji se odnose na evaluaciju performansi modela, a izračunati su na osnovu podataka iz prethodne tabele. Najmanji iznosi grešaka ukazuju da odabrani metod pruža najbolje rezultate u odnosu na ostale metode. Sva četiri odabrana pokazatelja na najnižem nivou ima MLP algoritam, dok najveće ima linearna regresija.

4.2. Predviđanje sa uključenim nezavisnim varijablama

U ovom slučaju, modeli uključuju i nezavisne promenljive. Kod predviđanja sa uključenim dodatnim varijablama koje naknadno objašnjavaju pojavu nije moguće izvršiti predviđanje za period van skupa. Razlog tome je što u WEKA alatu dodatne varijable moraju da imaju svoju zadatu vrednost u periodu za koji se zavisna varijabla predviđa. U testnom delu skupa (20%) postoje ove vrednosti pa je moguće izvršiti predviđanje. Pre

postupka implementacije algoritma primenjeni su različiti filteri koji se odnose na selekciju atributa, međutim ručno odabrana kombinacija pružila je bolje rezultate. Oni su predstavljeni u tabeli broj 4. i na grafikonu 2. Kod linearne regresije najbolje rezultate pružio je model sa uključenim prihodima iz industrije i građevinarstva, saobraćaja i skladištenja, informisanja i telekomunikacija u odnosu na prethodni period i izvoz robe i usluga. Sa uključenim pomenutim atributima svi pokazatelji evaluacije su postigli bolje rezultate. Kada je u pitanju MLP algoritam, uz uključene attribute prihodi iz industrije i građevinarstva, saobraćaja i skladištenja, informisanja i telekomunikacija u odnosu na prethodni period i izdaci za državnu potrošnju, pružio je bolje rezultate kod MAPE pokazatelja, dok je ostale evaluacione parametre imao blago veće nego bez uključenih dodatnih atributa.

Tabela 4. stvarne vrednosti i predviđene prema kvartalima

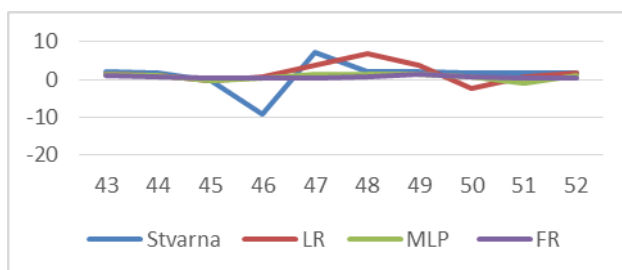
| Kvartal | Stvarna vrednost | LR | | MLP | | RF | |
|---------|------------------|---------------------|---------|---------------------|---------|---------------------|---------|
| | | Predviđena vrednost | Razlika | Predviđena vrednost | Razlika | Predviđena vrednost | Razlika |
| 43 | 2,1 | 1,5194 | -0,5806 | 1,4293 | -0,6707 | 0,945 | -1,155 |
| 44 | 1,8 | 0,9239 | -0,8761 | 1,1559 | -0,6441 | 0,745 | -1,055 |
| 45 | -0,5 | -0,2579 | 0,2421 | -0,2661 | 0,2339 | 0,4703 | 0,9703 |
| 46 | -9,2 | 0,5813 | 9,7813 | 0,4159 | 9,6159 | 0,3801 | 9,5801 |
| 47 | 7,2 | 3,7005 | -3,4995 | 1,4621 | -5,7379 | 0,525 | -6,675 |
| 48 | 2,2 | 7,0178 | 4,8178 | 1,4735 | -0,7265 | 0,6218 | -1,5782 |
| 49 | 2 | 3,9144 | 1,9144 | 1,4654 | -0,5346 | 1,214 | -0,786 |
| 50 | 1,7 | -2,541 | -4,241 | 0,7345 | -0,9655 | 0,8693 | -0,8307 |
| 51 | 1,7 | 0,7904 | -0,9096 | -0,982 | -2,682 | 0,3959 | -1,3041 |
| 52 | 1,7 | 1,8471 | 0,1471 | 1,1879 | -0,5121 | 0,3419 | -1,3581 |

Izvor: istraživanje autora

Model sa RF metodom je postigao bolje rezultate na svim evaluacionim parametrima ali na malom nivou uz uključivanje izvoza robe i usluga i

deviznog kursa dolara i dinara. Dobijeni rezultati predstavljani su na grafikonu broj 2.

Grafikon 2. stvarne i predviđene vrednosti po kvartalima



Izvor: istraživanje autora

Evaluacija modela predstavljena je u tabeli broj 5.

Tabela 5. Prikaz grešaka kod primenjenih metoda

| Greške | Metode | | |
|--------|---------|----------------|---------|
| | LR | MLP | RF |
| MAE | 2,701 | 2,2323 | 2,5293 |
| MAPE | 90,6011 | 60,3155 | 82,1012 |
| RMSE | 3,9344 | 3,6812 | 3,8349 |
| MSE | 15,4795 | 13,5509 | 14,7067 |

Izvor: istraživanje autora

U tabeli 5. uočava se da MLP algoritam ponovo ima najbolje rezultate kod svih evaluacionih parametara u odnosu na ostale primenjene metode

iako su 3 od 4 pokazatelja imali blago veći nivo greške. U tabeli broj 6. prikazane su razlike u evaluaciji modela sa i bez uključenih atributa.

Tabela 6. Prikaz razlike grešaka kod primenjenih metoda

| | LR | MLP | RF |
|------|---------|---------|--------|
| MAE | 0,3951 | -0,0021 | 0,0199 |
| MAPE | 36,3397 | 11,289 | 1,5014 |
| RMSE | 0,5666 | -0,0403 | 0,0514 |
| MSE | 4,7798 | -0,2945 | 0,3969 |

Izvor: istraživanje autora

Posmatrajući tabelu broj 6. uočava se da su LR i RF poboljšali sve evaluacione parametre uvođenjem novih atributa koji dodatno objašnjavaju zavisnu varijablu. RF je postigao skoro beznačajan napredak dok je LR uvođenje novih atributa imalo znatan uticaj na model. MLP je poboljšalo samo MAPE pokazatelj dok su ostali imali blago niže rezultate u odnosu na model bez dodatnih varijabli.

ZAKLJUČAK

Predviđanje makroekonomskih pokazatelja predstavlja izazovan zadatak, prvenstveno zbog nepredviđenih uticaja. U ovom radu predviđanje rasta BDP-a je izvršeno nad vremenskom serijom podataka pomoću tri različite metode rudarenja u dva slučaja. Prvi slučaj je obuhvatao samo rast BDP-a i predviđanje na osnovu njega, dok su u drugom slučaju bile uključene i dodatne varijable u cilju boljeg objašnjavanja zavisne promenjive. Za

obuku modela korišteno je 80% skupa, dok je poslednjih 20% instanci (10 kvartala) korišteno za evaluaciju. Skup je bio prukupljen na kvartalnom nivou i obuhvatao je period od 1. januara 2009. godine do 1. januara 2022. godine. Originalni skup je imao 14 dodatnih varijabli (nezavisnih promenljivih) od kojih su samo određene odabrane za izgradnju modela. Dodatne varijable su kod LR i RF algoritama pružile bolje rezultate kod svih evaluacionih parametara. Međutim, kod MLP metode MAPE pokazatelj je imao manju vrednost, dok su ostali imali blago veće iznose. Rezultati do kojih se došlo ukazuju da je u oba slučaja najuspešni bio MLP algoritam. Iako je on pružio najbolje rezultate, oni nisu na zavidnom nivou. Jedan od razloga za ovo je pandemija koronavirusa koja se javila u kvartalima kod kojih je obavljano predviđanje i evaluacija modela. Nastavak istraživanja može da se bazira na poboljšanju rezultata predviđanja kroz primenu novih metoda ili izmenu postojećih, novih dodatnih atributa (naročito onih koji prikazuju razvoj pandemije koronavirusa), obuhvatanje više instanci u skupu podataka i slično.

LITERATURA

- [1] Aggarwal, C. C. (2015). *Data Mining - The Textbook*. London: Springer International Publishing.
- [2] Blanchard, O. (2008). *Macroeconomics*. Massachusetts: Pearson College Div.
- [3] Brammer, M. (2016). *Principles of Data Mining*. Oxford, UK: Springer-Verlag London.
- [4] Breiman, L. (2001). Random Forest. *Machine learning*, 5-32.
- [5] Carreiro, A., Clark, T., & Marcellino, M. (2015). Bayesian vars: specification choices and forecast accuracy. *Journal of applied econometrics*, 46-73.
- [6] Carreiro, A., Clark, T., & Marcellino, M. (2019). Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 137-154.
- [7] Carriero, A., Galvao, A. B., & Kapetanios, G. (2019). A comprehensive evaluation of macroeconomic forecasting methods. *International Journal of Forecasting*, 1226-1239.
- [8] Chatfield, C. (2016). *The Analysis of Time Series: An Introduction*. CRC Press.
- [9] Dean, J. (2014). *Data mining and machine learning, Value Creation for Business Leaders and Practitioners*. New Jersey: SAS Institute.
- [10] Erkam, G., Kayakutlu, G., & Daim, T. (2011). Using artificial neural network models in stock market index prediction. *Expert Systems with Applications*, 10389-10397.
- [11] Frank, E., Hall, M. A., Witten, I. H., & Pal, C. J. (2016). *The WEKA Workbench*. Retrieved from Waikato: https://waikato.github.io/weka-wiki/citing_weka/
- [12] Gérard, B., & Scornet, E. (2016). A random forest guided tour. *Springer Nature SharedIt*, 197-227.
- [13] Hall, M. (2014). Time Series Analysis and Forecasting with Weka - Pentaho Data Mining. Retrieved from https://pentaho-community.atlassian.net/wiki/spaces/DATA_MINING/pages/293700841/Time+Series+Analysis+and+Forecasting+with+Weka#TimeSeriesAnalysisandForecastingwithWeka-3.2.2Lagcreation
- [14] Kriesel, D. (2007). *A Brief Introduction to Neural Networks*. Retrieved from https://www.dkriesel.com/_media/science/neuronalenetze-en-zeta2-2col-dkrieselcom.pdf
- [15] Limited, F. M. (2022). *investing.com*. Retrieved from <https://www.investing.com/currencies/usd-rsd-historical-data>
- [16] Menzies, T., Kocagüneli, E., Peters, F., & Turhan, B. (2015). Using Goals in Model-Based Reasoning. *Sharing Data and Models in Software Engineering*, 321-353.
- [17] Nemes, M., & Butoi, A. (2013). Data Mining on Romanian Stock Market Using Neural Networks for Price. *Informatica Economică*, 125-136.
- [18] Schorfheide, F., & Song, D. (2014). Real-Time Forecasting with a Mixed-Frequency VAR. *Journal of Business & Economic Statistics*.
- [19] Shmueli, G., Bruce, C. P., & Patel, R. N. (2016). *Data Mining For Business Analytics, Concepts, Techniques, and Applications with XLMiner*. New Jersey: John Wiley & Sons.
- [20] Smets, F., Warne, A., & Wouters, R. (2014). Professional forecasters and real-time forecasting with a DSGE model. *International Journal of Forecasting*, 981-995.
- [21] R. z. (2022). *Republički zavod za statistiku*. Retrieved from <https://www.stat.gov.rs>
- [22] Stock, H. J., & Watson, W. M. (2015). *Introduction to econometrics - third edition*. New Jersey: Pearson Education.
- [23] Tsai, C.-F., & Wu, J.-W. (2008). Using neural network ensembles for bankruptcy prediction and credit scoring. *Expert Systems with Applications*, 2639-2649.

- [24] Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2017). *Data mining - Practical Machine Learning Tools and Techniques*. Cambridge: Elsevier.

SUMMARY

Forecasting macroeconomic indicators is a challenging task, primarily due to unforeseen impacts. In this paper, GDP growth forecasting was performed over a time series of data using three different data mining methods in two cases. The first case included only GDP growth and forecasting based on it, while in the second case additional variables were included in order to better explain the dependent variable. Part of set (80%) was used for model training, while the last 20% of instances (10 quarters) were used for evaluation. The data set was collected on a quarterly basis and covered the period from January 1, 2009 to January 1, 2022. The original data set had 14 additional variables (independent variables) where only certain were selected to build the model. Additional variables with LR and RF algorithms provided better results on all evaluation parameters. However, in the MLP method MAPE indicator had a lower value, while the others had slightly higher. The results indicate that the MLP algorithm was the most successful in both cases. Although it provided the best results, they are not at a good level. One of the reasons for this is the COVID-19 pandemic that occurred in the quarters where model prediction and evaluation was performed. This unforeseen impact had a global character, so that a large number of countries recorded turbulent trends in GDP growth. Continuation of research can be based on improving the results of prediction by applying new methods or changing existing, new additional attributes (especially those that show the development of the COVID-19 pandemic), collecting more instances in a data set and etc.