



PREDVIĐANJE CENA KRIPTOVALUTA UPOTREBOM ARIMA I ARIMAKS MODELA
CRYPTOCURRENCY PRICE PREDICTION USING ARIMA AND ARIMAX MODELS

Predrag Glavaš, *Fakultet tehničkih nauka, Novi Sad*

Oblast – ELEKTROTEHNIKA I RAČUNARSTVO

Kratak sadržaj – *Ovaj rad se bavi predviđanjem cena Dogecoin, Šiba Inu i Idena kriptovaluta u kratkom i srednjem roku upotrebom ARIMA i ARIMAKS modela. Podaci su preuzimani sa različitih javno dostupnih izvora, vremenske serije preprocesirane kako bi se osigurala stacionarnost, a parametri modela birani pretragom kombinacija parametara i putem auto arima funkcije. Testirani su različiti intervali predviđanja, a modeli evaluirani Akaike informacionim kriterijumom i greškom u apsolutnom procentu. Za kratkoročna predviđanja ARIMA model je imao grešku od oko 3.5%, ARIMAKS od 7 do 19%, dok su se za srednji rok oba modela pokazala kao loša.*

Cljučne reči: *Kriptovalute, ARIMA, ARIMAKS, Dogecoin, Šiba Inu, Idena*

Abstract – *This work discusses short and mid term price prediction of Dogecoin, Shiba Inu and Idena cryptocurrencies using ARIMA and ARIMAX models. Data was sampled from various publicly available sources, preprocessed to ensure stationarity and model parameters chosen with grid search and auto arima function. Different prediction intervals were tested and models evaluated using Akaike Information Criterion and mean average percent error. For short term predictions ARIMA model had error of about 3.5%, ARIMAX 7 to 19%. Both models proved inadequate for mid term predictions.*

Keywords: *Cryptocurrencies, ARIMA, ARIMAX, Shiba Inu, Dogecoin, Idena*

1. UVOD

Od izbijanja pandemije Koronavirusa 2020. godine cene kriptovaluta su doživele nagli skok prouzrokovan ulaskom novih investitora na tržište. U takvoj situaciji mnogi investitori su uložili novac u visoko špekulativne kriptovalute i izgubili novac usled velikih fluktuacija u cenama, pa je neophodno naći adekvatan model za njihovo predviđanje kako bi se smanjili ovi gubici.

Dodžkoin kriptovaluta je kreirana 2013. kao šala i ubrzo je doživela veliki skok cene od preko 100 puta zahvaljujući popularnosti u zajednici penjući se na 0.3 dolara od 2020. do sredine 2021 [1]. Šiba Inu, kao još jedna od šala-valuta i direktni konkurent Dodžkoinu kreirana je 2020. i od tada do danas cena joj je porasla preko 2500% zahvaljujući samo velikoj popularnosti u zajednici [2].

NAPOMENA:

Ovaj rad proistekao je iz master rada čiji mentor je bio dr Aleksandar Kovačević, red. prof.

Idena s druge strane je stvorena sa idejom da podrži novi blokčejn ekosistem gde svaki validator mora dokazati da je čovek rešavajući niz logičkih zadataka u tačno određeno vreme. Iako za razliku od Dodžkoina i Šibe Inu Idena ima tehnološku i upotrebnu vrednost njena cena je sada 2022. niža nego pri puštanju u opticaj 2020 [3].

Cilj ovog rada je da se utvrdi da li je moguće tačno predvideti cene ovih kriptovaluta na osnovu podataka o ranijem kretanju cene, zajednici i razvojnom timu. Podaci o zajednici su smatrani potencijalno bitnim jer Šiba Inu i Dodžkoin duguju svoj uspeh popularnosti u zajednici, a Idena ima veliki tehnološki značaj pa podaci o razvojnom timu mogu biti bitni u tom kontekstu. ARIMA i ARIMAKS modeli su izabrani na osnovu dostupne relevantne literature. U kratkoročnim predviđanjima dan unapred ARIMA model je imao prosečnu grešku oko 3.5% za sve valute, a ARIMAKS od 7 do 19% u zavisnosti od valute. Za srednjoročna i dugoročna predviđanja oba modela su se loše pokazala jer ne uspevaju da tačno predvide amplitude cenovnih šokova.

U prvom poglavlju dat je uvod sa opisom problema, u drugom opis prethodnih rešenja i najrelevantnijih radova na osnovu kojih je napisan ovaj rad. U trećem poglavlju je opisan način formiranja skupa podataka, u četvrtom metodologija predviđanja korišćena u radu, u petom eksperimenti koji su sprovedeni i njihovi rezultati, a u šestom su dati zaključci rada.

2. PRETHODNA REŠENJA

U radu [4] Amina Azarija testirano je predviđanje cene Bitkoina dan unapred upotrebom ARIMA modela. Podaci su prikupljeni na dnevnom nivou za period od 3 godine, preprocesirani kako bi se osigurala stacionarnost a model testiran nad različitim potperiodima kako bi se proverila njegova robustnost. Utvrđeno je da se ARIMA model zaista može koristiti za predviđanje cene Bitkoina ali da ne može sa tačnošću predvideti velike uspone i padove. Ovaj model bi se mogao poboljšati uvođenjem varijabli sa podacima o zajednici i razvojnom timu.

Rad [5] Đovanija Sklacota proverava uticaj postova na Tviteru na Eterijum i Dodžkoin kriptovalute i testira različite ARIMAKS modele kako bi utvrdio kako dodatne varijable utiču na tačnost predviđanja cene. Podaci su preuzeti za prva 4 meseca 2021. godine skidajući tvitove sa različitim heštagovima vezanim za kriptovalute. Rezultati pokazuju da Tviter ima veliki uticaj na cenu Dodžkoina nasuprot Eterijumu u velikom delu zbog toga što je Dodžkoin vođen isključivo popularnošću u zajednici.

U radu [6], analizirane su makroekonomske vremenske serije, BDP i stopa nezaposlenosti su predviđani upotrebom ARIMA i ARIMAKS modela. Parametri modela su procenjivani preko grafika autokorelacije i parcijalne autokorelacije. ARIMA model se pokazao malo tačniji od ARIMAKS modela sa greškom od 1.77% u poređenju sa 3.78% koju pravi ARIMAKS.

Rad [7] sa konferencije u Pajtonu u nauci bavi se estimacijom parametara ARIMA modela pomoću grafika autokorelacije (ACF) i parcijalne autokorelacije (PACF). Rad sugerše da je pre treniranja modela potrebno utvrditi stacionarnost vremenskih serija kombinacijom KPSS i ADF testova, a ukoliko se utvrdi nestacionarnost potrebno je diferencirati vremensku seriju. Rad takođe sugerše da čisto autoregresivan proces ima ACF grafik koji opada eksponencijalno ili oscilatorno dok se PACF grafik naglo odseca nakon p lagova. Proces čistog pokretnog proseka ima ACF grafik koji se naglo odseca nakon q lagova dok PACF grafik opada eksponencijalno ili oscilatorno. Ovi zaključci su bitni za estimaciju hiperparametara modela.

3. FORMIRANJE SKUPA PODATAKA

Skupovi podataka za sve 3 kriptovalute su formirani preuzimanjem podataka sa Jahu finansija [8], Koingeka [9] i Gugl trendova [10]. Podaci su prikupljeni od najstarijeg dostupnog datuma do 4.2.2022.

Podaci sa Jahu finansija su preuzeti ručno skidanjem fajlova za svaki od kriptovaluta, ovi fajlovi su pružili finansijske podatke finalnom skupu. Podaci sa Koingeka su preuzeti pisanjem Pajton skripte koja poziva negov API za svaki dan u navedenom intervalu i dobavlja podatke o zajednici i razvojnom timu, agregira ih i čuva u jedinstveni CSV fajl. Podaci sa Gugl trendova su preuzeti na isti način kao i oni sa Koingeka, ključne reči koje su pretraživane su „shiba inu coin“, „idena“ i „dogecoin“. Geolokacija pretrage popularnosti je postavljena na globalnu kako bi se dobili rezultati za sve zemlje.

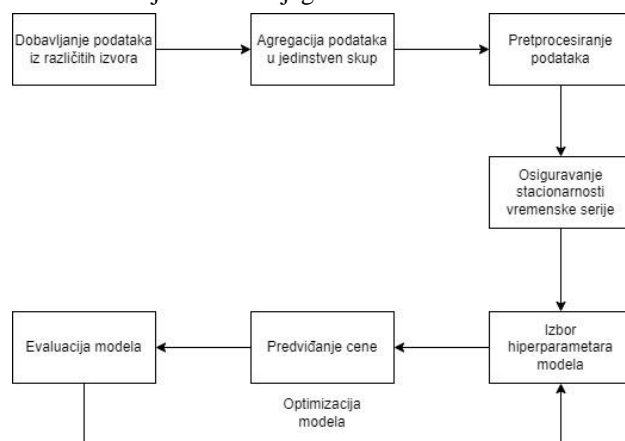
Tokom procesa eksplorativne analize tri dobijena fajla za svaku u valutu su spojena u jedan presecanjem, to je dovelo do pojave rupa u vremenskoj seriji jer za neke dane Gugl trendovi nisu imali podatke. Rupe su popunjavane propagacijom poslednjeg validnog dana unapred. Za neke atribute je takođe utvrđeno da postoje nedostajući podaci, ukoliko je procenat nedostajućih bio manji od 50% nedostajući podaci su popunjavani linearnom interpolacijom.

Ukoliko je za atribut nedostajalo više od 50% uzorka on je izbacivan iz skupa. Analizirana je i korelacija između svih nezavisnih atributa i ciljne promenljive *Close* (cena na kraju dana) kao i između nezavisnih promenljivih međusobno. Korelacija je analizirana za vrednosti od pre 3, 7, 14, 21, 30 i 60 dana. Tamo gde se utvrdilo da je korelacija između nezavisnog i zavisnog atributa manja od 0.6 on je izbacivan iz skupa, a ako je između 2 nezavisna atributa korelacija veća od 0.9 jedan je odabran i izbačen iz skupa. Na ovaj način je za svaku valutu kreirano dodatnih 6 skupova podataka nakon svake analize. U skupove su na kraju dodate varijable standardne devijacije i srednje vrednosti za svaki atribut i to računane za prozore od 3, 7, 14, 21, 30 i 60 dana.

Kako bi se vremenske serije zavisnog atributa mogle predviđati testirana je njihova stacionarnost za svaku valutu pomoću kombinacije KPSS i ADF testova. Ukoliko je bar jedan test pokazao nestacionarnost vremenska serija je smatrana nestacionarnom. Za Dodžkoin i Šiba Inu je utvrđeno da su nestacionarne, dok je za Ideninu vremensku seriju statistika testa bila na granici. Dodžkoin i Šiba Inu vremenske serije su prvo logaritmovane a zatim jedanput diferencirane nakon čega je testovima utvrđena stacionarnost, a za Idenu je odlučeno da se smatra stacionarnom iz bojazni da bi pošto je granici mogla postati prediferencirana.

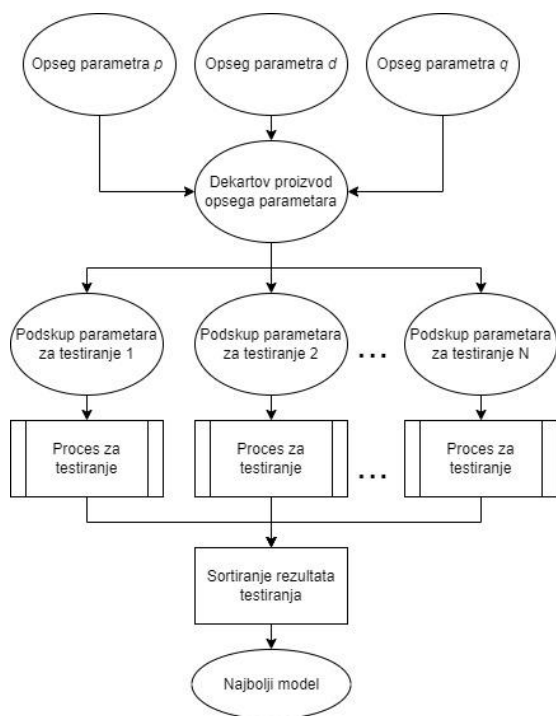
4. METODOLOGIJA

Metodologija za predviđanje cene kriptovaluta korišćena u ovom radu je data na dijagramu sa slike 1.



Slika 1. Dijagram metodologije.

Dobavljanje podataka, agregacija, preprocesiranje i osiguravanje stacionarnosti su opisani u prethodnom poglavlju. Izbor hiperparametara je vršen na 2 načina: prvi je pomoću *auto arima* funkcije u Pajtonu koja testira različite kombinacije hiperparametara modela i vraća onaj koji ima najmanju vrednost Akaike informacionog kriterijuma, a drugi je pretraga različitih kombinacija hiperparametara modela. Pretraga kombinacija hiperparametara je vršena pomoću skripte napisane u Pajtonu kojoj se proslede opsezi hiperparametara ARIMA modela p , d i q a ona onda testira sve kombinacije poredeći grešku koja se pravi pri predviđanju i vrednost Akaike informacionog kriterijuma. Testiranje kombinacija se vrši prvo pravljnjenjem dekartovog proizvoda opsega hiperparametara, a zatim deljenjem tog opsega na n podskupova gde je n broj jezgara procesora na računaru i testiranjem svake kombinacije iz skupa paralelno u jezgrima procesora. Testiranje u jezgrima se vrši treniranjem modela nad delom vremenske serije odabranim za trening, zatim predviđanjem korak unapred i računanjem greške. Nakon računanja greške trening skup se proširuje stvarnom ili predviđenom vrednošću za korak unapred u zavisnosti od eksperimenta, a zatim se model ponovo trenira. Ovaj postupak se ponavlja za svaku vrednost u test podskupu vremenske serije i na kraju se računa greška u apsolutnom procentu i vrednost Akaike informacionog kriterijuma za testirani model. Kada svako jezgro završi testiranje svojih modela oni se svi sortiraju po apsolutnoj grešci i Akaike informacionom kriterijumu u rastućem redosledu i bira se najbolji. Dijagram ove metodologije je dat na slici 2.



Slika 2. Dijagram izbora hiperparametara modela.

Izbor hiperparametara, predviđanje i evaluacija su u nekim eksperimentima rađeni kao zajednički a u nekim kao odvojen proces.

5. EKSPERIMENTI I REZULTATI

Za potrebe ovog rada sprovedena su 4 eksperimenta čiji je opis u nastavku.

U prvom eksperimentu predviđane su cene sve 3 kriptovalute dan unapred upotrebom ARIMA modela. Za potrebe eksperimenta Šiba Inu i Dodžkoin skupovi su podeljeni u odnosu 90:10 za trening i testiranje, a Idena skup u odnosu 70:30. Analizom PACF i ACF grafika dobijeni su upotrebljivi rezultati samo za Idena vremensku seriju koja je pokazala znakove čisto autoregresivnog procesa ARIMA(1,0,0). Hiperparametri modela su određivani i preko pretrage kombinacija, gde je prilikom testiranja modela dodavana stvarna vrednost u trening skup nakon predviđanja. Rezultati prvog eksperimenta su u tabeli 1.

Tabela 1. Rezultati prvog eksperimenta

Kriptovaluta	p	d	q	AIC	MAPE
Idena	1	1	0	1323	3.63%
Dodžkoin	5	1	2	3350	3.34%
Šiba Inu	3	1	0	152	3.59%
Idena ¹	1	0	0	1350	3.74%

¹ Model čiji su parametri određeni preko ACF/PACF grafika

U drugom eksperimentu predviđana je cena kriptovalute više dana unapred upotrebom ARIMAKS modela. Hiperparametri modela su birani pomoću *auto arima* funkcije, a kao prediktivne varijable korišćene su vrednosti promenljivih od pre 3, 7, 14, 21, 30 i 60 dana. Takođe su i testirani različiti odnosi podele skupova za trening i test i to 70:30, 80:20 i 90:10. Rezultati eksperimenta su dati u tabeli 2.

Tabela 2. Rezultati drugog eksperimenta

Valuta	Period laga	p	d	q	AIC	MAPE
Dodžkoin	3 dana	4	0	2	-7955	7.12%
Dodžkoin	7 dana	1	0	3	-7492	8.75%
Idena	3 dana	1	0	0	-2631	17.37%
Idena	7 dana	2	0	1	-2624	18.84%
Šiba Inu	3 dana	1	0	0	-10497	7.15%
Šiba Inu	7 dana	1	0	0	-10463	14.89%

U trećem eksperimentu je testirano predviđanje sve 3 kriptovalute više dana unapred upotrebom ARIMA modela. Hiperparametri su određeni kao i u prvom eksperimentu uz to da je tokom testiranja predviđena vrednost dodavana u trening skup nakon svakog predviđanja umesto stvarne. Rezultati eksperimenta su u tabeli 3.

Tabela 3. Rezultati trećeg eksperimenta

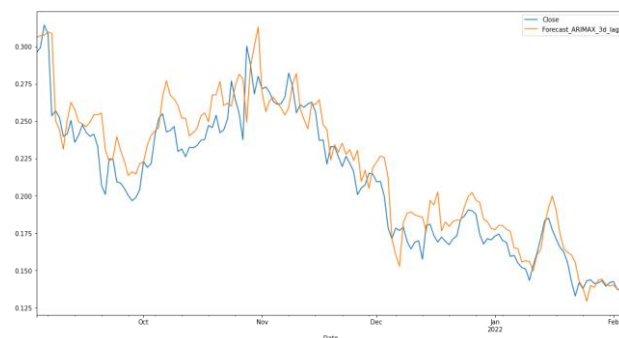
Valuta	Period	p	d	q	AIC	MAPE
Dodžkoin	155 dana	5	1	6	3399	40.55%
Idena	163 dana	4	0	4	1415	14.63%
Šiba Inu	29 dana	8	1	7	173	18.35%

U četvrtom eksperimentu testirano je predviđanje sve 3 kriptovalute dan unapred upotrebom ARIMAKS modela. Za predviđanje su korišćene srednje vrednosti i standardne devijacije varijabli računate za prozore od 3 i 7 dana. Skupovi su podeljeni za trening i testiranje u istim odnosima kao u eksperimentu 1. Za određivanje hiperparametara modela korišćena je *auto arima* funkcija. Rezultati četvrtog eksperimenta su dati u tabeli 4.

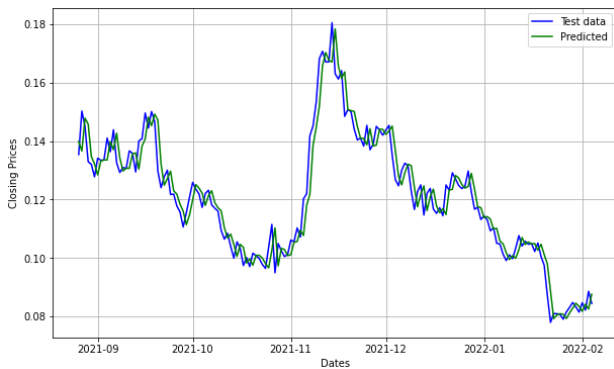
Tabela 4. Rezultati četvrtog eksperimenta

Valuta	Period laga	p	d	q	AIC	MAPE
Dodžkoin	3 dana	1	0	0	-8418	6.21%
Dodžkoin	7 dana	2	0	3	-8351	12.73%
Idena	3 dana	1	0	0	-2132	7.46%
Idena	7 dana	2	0	1	-2119	12.38%
Šiba Inu	3 dana	1	0	0	-10493	6.46%
Šiba Inu	7 dana	1	0	0	-10496	10.50%

Najbolji rezultati kod ARIMA i AIRMAKS modela za kratkoročno predviđanje su dobijeni sa modelima ARIMAKS(1,0,0) za Dodžkoin dan unapred (slika 3) i ARIMA(5,1,2) za Dodžkoin dan unapred (slika 4).



Slika 3. Dijagram predviđanja cene Dodžkoina ARIMAKS modelom.



Slika 4. Dijagram predviđanja cene Dodžkoina ARIMA modelom.

6. ZAKLJUČAK

U ovom radu je predstavljeno kratkoročno i srednjeročno predviđanje cena 3 kriptovalute Dodžkoin, Šiba Inu i Idena upotrebom ARIMA i ARIMAKS modela. Testirani su različiti odnosi podele skupova za trening i test i različite kombinacije hiperparametara modela.

Analiza grafika autokorelacije i parcijalne autokorelacije se pokazala nedovoljno pouzdanom u slučaju Šiba Inu i Dodžkoina dok je kod Idene njome dobijen model skoro isto tačan kao i onaj dobijen pretragom kombinacija hiperparametara. Pretraga kombinacija hiperparametara se pokazala dosta pouzdanom ali i znatno sporijom jer je trajala u proseku više od 25 časova za sve 3 kriptovalute. Iako je paralelizacija procesa pretrage dosta ubrzala stvari, pojavilo se usko grlo jer su modeli višeg reda koji se najduže treniraju završili u poslednja 2 jezgra procesora i nisu ravnomerno raspoređeni.

Pri podeli skupova za trening i test potrebno je obratiti pažnju da veliki skokovi i padovi cene budu sadržani u trening skupu jer i ARIMA i ARIMAKS modeli ne mogu adekvatno da predvide amplitude takvih cenovnih šokova. Prilikom analize korelacije promenljivih ARIMAKS modela potrebno je uzeti u obzir njihovu vrednost od pre najviše 7 dana jer su raniji periodi pokazali veoma slab uticaj, a modeli trenirani sa tim vrednostima slabu tačnost.

Uočeno je da oba modela prilikom predviđanja dan unapred daju grafike predviđanja pomerene blago unapred. Ovo može biti posledica toga da modeli ne mogu zaista da uhvate cenovne oscilacije već samo pokušavaju predviđanje na osnovu trenda od par dana unazad.

Oba modela pri dugoročnim predviđanjima prave velike greške i čini se da ne mogu biti korišćeni u tu svrhu. ARIMA model kod dugoročnih predviđanja opisanih u

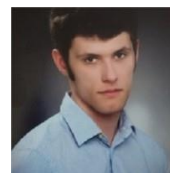
eksperimentu 3 daje grafik nalik na linearnu regresiju i ne može uopšte da uhvati trend kretanja iako se to ne bi reklo samo gledajući prosečnu grešku.

Za moguće pravce nadogradnje predlaže se poređenje ARIMA modela i linearne regresije za predviđanje više dana unapred kako bi se utvrdilo koliko su slični. Predlaže se i pokušaj predviđanja na nedeljnom, mesečnom i godišnjem nivou računanjem proseka dnevnih podataka za nedelju, mesec i godinu dana.

7. LITERATURA

- [1] Chohan, Usman W. "A history of Dogecoin." *Discussion Series: Notes on the 21st Century* (2021).
- [2] <https://shibatoken.com/> (pristupljeno u maju 2022.)
- [3] <https://docs.idena.io/docs/wp/summary/> (pristupljeno u maju 2022.)
- [4] Azari, Amin. "Bitcoin price prediction: An ARIMA approach." arXiv preprint arXiv:1904.05315 (2019).
- [5] Scalzotto, Giovanni. "Social Media Impact on Cryptocurrencies." (2021).
- [6] Peter, Đurka, and Pastoreková Silvia. "ARIMA vs. ARIMAX—which approach is better to analyze and forecast macroeconomic time series." *Proceedings of 30th international conference mathematical methods in economics*. Vol. 2. 2012.
- [7] Mahan, Margaret & Chorn, Chelley & Georgopoulos, Apostolos. (2015). White Noise Test: detecting autocorrelation and nonstationarities in long time series after ARIMA modeling. 10.25080/Majora-7b98e3ed-00f.
- [8] <https://finance.yahoo.com/> (pristupljeno u maju 2022.)
- [9] <https://www.coingecko.com/> (pristupljeno u maju 2022.)
- [10] <https://trends.google.com/> (pristupljeno u maju 2022.)

Kratka biografija:



Predrag Glavaš rođen je u Beogradu 1998. god. Master rad na Fakultetu tehničkih nauka iz oblasti Elektrotehnike i računarstva – Elektronsko poslovanje odbranio je 2022. god.

kontakt: pedjaglavas98@gmail.com