



УНИВЕРЗИТЕТ У НОВОМ САДУ
ФАКУЛТЕТ ТЕХНИЧКИХ НАУКА
У НОВОМ САДУ



Предраг Главаш

Предвиђање цена криптовалута употребом АРИМА и АРИМАКС модела

Мастер рад
- Мастер академске студије -

Нови Сад, 2022 година.

	UNIVERZITET U NOVOM SADU FAKULTET TEHNIČKIH NAUKA 21000 NOVI SAD, Trg Dositeja Obradovića 6	Datum:
	ZADATAK ZA IZRADU DIPLOMSKOG (MASTER) RADA	List:
		1/1

(Podatke unosi predmetni nastavnik - mentor)

Vrsta studija:	Мастер академске студије
Studijski program:	Рачунарство и аутоматика
Rukovodilac studijskog programa:	др. Мирна Капетина

Student:	Предраг Главаш	Broj indeksa:	E238-2021
Oblast:	машинско учење		
Mentor:	др. Александар Ковачевић		

NA OSNOVU PODNETE PRIJAVE, PRILOŽENE DOKUMENTACIJE I ODREDBI STATUTA FAKULTETA IZDAJE SE ZADATAK ZA DIPLOMSKI RAD, SA SLEDEĆIM ELEMENTIMA:

- problem – tema rada;
- način rešavanja problema i način praktične provere rezultata rada, ako je takva provera neophodna;
- literatura

NASLOV DIPLOMSKOG (MASTER) RADA:

Предвиђање цена криптовалута употребом АРИМА и АРИМАКС модела

TEKST ZADATKA:

Аналиzirати стање у области предвиђања временских серија цена криптовалута и идентификовати адекватан модел за њихово предвиђање. Прикупити податке о историји цена криптовалута, истренирати модел, извршити предвиђање помоћу њега и евалуирати му тачност. Прокоментарисати резултате.

Rukovodilac studijskog programa:	Mentor rada:

Primerak za: ☐ - Studenta; ☐ - Mentora

САДРЖАЈ

1.	УВОД	7
2.	ПРЕГЛЕД СТАЊА У ОБЛАСТИ.....	11
3.	ТЕОРИЈСКИ ПОЈМОВИ И ДЕФИНИЦИЈЕ	15
3.1	АРИМА и АРИМАКС модели	15
3.2	Тестови стационарности.....	18
3.2.1	АДФ тест.....	18
3.2.2	КПСС тест.....	19
4.	МЕТОДОЛОГИЈА	21
4.1	Опис методологије	21
4.2	Прикупљање и агрегација података	22
4.3	Претпроцесирање података	23
4.4	Осигуравање стационарности	24
4.5	Избор хиперпараметара модела, предвиђање и евалуација	25
5.	ЕКСПЕРИМЕНТИ	29
5.1	Скупови података.....	29
5.1.1	Доцкоин скуп.....	30
5.1.2	Шиба Ину скуп.....	32
5.1.3	Идена скуп	33
5.2	Тестирање стационарности серија.....	33
5.3	Спецификације рачунара и софтвера	37
5.4	Експеримент 1	37
5.5	Експеримент 2	39
5.6	Експеримент 3	40
5.7	Експеримент 4	41
5.8	Евалуација модела.....	42
6.	РЕЗУЛТАТИ И ДИСКУСИЈА	43
6.1	Резултати експеримента 1	43
6.2	Резултати експеримента 2	46
6.3	Резултати експеримента 3	55
6.4	Резултати експеримента 4	57
7.	ЗАКЉУЧАК.....	60
8.	ЛИТЕРАТУРА.....	63
9.	БИОГРАФИЈА.....	67
	KLJUČNA DOKUMENTACIJSKA INFORMACIJA	69
	KEY WORDS DOCUMENTATION.....	71

1. УВОД

Криптовалуте су децентрализовани дигитални систем валута које омогућавају корисницима да виртуелно плате робу и услуге без посредства централног управљачког ауторитета као што је банка [1]. Тржиште криптовалута је од почетка пандемије Корона вируса 2020. године доживело велики скок цена проузрокован инфлуксом нових инвеститора. У оваквој ситуацији многи инвеститори су уложили новац у неке високо ризичне криптовалуте на основу њихове популарности и доживели велике губитке због значајних флукуација у ценама. Зато је јако битно пронаћи адекватан модел који би могао предвидети цене ових криптовалута како би се умањио потенцијални инвестициони ризик.

Доцкоин (енг. *Dogecoin*) је криптовалута креирана као интернет шала 2013. године и убрзо је стекла велику интернет популарност и базу лојалних пратилаца и корисника. Њена цена је од 2020. до средине 2021. године порасла 100 пута, са 0.3 на 30 центи Америчког долара [2]. Још једна у низу оваквих валута са псећим називом је Шива Ину (енг. *Shiba Inu*) коју су њени креатори назвали „Убица Доцкоина“. Вођена само интернет популарношћу и великом групом пратилаца, цена јој је порасла преко 2500% од настанка 2020. године до данас [3]. Цена обе ове валуте јако варира чак и на дневном нивоу, а нови инвеститори се обично одлучују за куповину током фаза јаког раста цене. Фазе раста су неизбежно праћене великим падовима након што валута постане прецењена. Идена (енг. *Idena*) је за разлику од претходне 2 валуте креирана као подршка новом блокчејн екосистему са јасном технолошком и употребном вредношћу. Идеја иза Идене је да сваки валидатор у систему мора доказати да је човек решавајући низ логичких задатака у тачно одређено време које систем пропише [4]. Упркос својој технолошкој вредности и новини коју доноси, њена цена је данас 2022. године мања него при пуштању у оптицај 2020. Оваква нелогичност у кретању цена отвара питање да ли су валуте сличне Идени стабилније и да ли се њихова цена може боље предвидети у односу на валуте као што су Доцкоин и Шива.

Циљ овог рада је предвиђање цена горе наведених криптовалута употребом АРИМА и АРИМАКС (енг. *ARIMAX*) модела. Уз финансијске податке о кретању цене валута додати су и подаци о развојном тиму и заједници корисника како би се потенцијално

побољшала тачност јер Шиба Ину и Доцкоин базирају своју вредност на популарности у заједници а Идена на технологији која иза ње стоји. Тестирани су различити интервали предвиђања од једног до више дана унапред и различите комбинације хиперпараметара модела.

Модели су евалуирани Акаике [5] информационим критеријумом и грешком у апсолутном проценту. Тестови су показали да АРИМА и АРИМАКС модели показују најбољу тачност у краткорочном предвиђању 1 до 7 дана унапред. За предвиђање дан унапред најбољи је био АРИМА модел са грешком у апсолутном проценту од око 3.5% за све три криптовалуте. АРИМАКС модел је за исто предвиђање имао скоро duplo већу грешку. За предвиђања 3-7 дана унапред као једини валидан се показао АРИМАКС модел чије се грешке крећу од 7 до 19% у зависности од криптовалуте и броја дана унапред за које се предвиђа.

Овај рад има 9 поглавља, чији ће кратак садржај бити дат у наставку. У првом поглављу је дат опис решавањег проблема уз мотивацију за његово решавање. Друго поглавље приказује тренутно стање у области и релевантна достигнућа других аутора у предвиђању цена криптовалута и акција. Треће поглавље описује теоријске концепте иза АРИМА и АРИМАКС модела и стационарности временске серије. У четвртном поглављу је описана методологија коришћена у овом раду, а у петом њена примена кроз експерименте и евалуацију модела. У шестом поглављу су приказани резултати експеримената заједно са дискусијом, а у седмом дат закључак рада. У осмом поглављу је наведена литература на основу које је рад написан, а у последњем биографија аутора.

2. ПРЕГЛЕД СТАЊА У ОБЛАСТИ

У овом поглављу дат је приказ тренутног стања у области предвиђања цена криптовалута и акција уз посебан осврт на радове везане за АРИМА и АРИМАКС моделе и естимацију њихових параметара.

Вештачка интелигенција и статистика се већ преко 40 година користе за предвиђање временских серија. Првобитно коришћене методе су биле статистичке као на пример АРИМА модел изложен 1976. у раду Бокса и Џенкинса [6]. Међутим њихова мана је што се процеси у финансијском домену често не могу описати линеарним моделима те тачност и генерализација модела није адекватна. Настанком неуронских мрежа аутори су 90-их година прошлог века почели интензивније да их разматрају у циљу тачнијег предвиђања цена акција и показале су се боље од класичних регресионих метода [7][8]. Са појавом криптовалута 2009. године и наглим растом њихове популарности и цене аутори су у наредним годинама почели да истражују методе за адекватно предвиђање њихових цена. Већ 2015. године почињу масовно да се појављују радови који се баве поређењем перформанси класичних статистичких метода и различитих врста неуронских мрежа у сврху предвиђања цена криптовалута. Они су показали да боље резултате дају неуронске мреже као и код предвиђања цена акција [9][10][11]. Међутим показани резултати нису увек толико много бољи и конзистентно бољи да би се са сигурношћу могло рећи да статистичке методе немају вредност и да се не могу потенцијално побољшати.

У протеклих 10 година различити радови су се бавили поређењем алгоритама за предвиђање временских серија везаних за финансијску сферу. Рад [17] Петера, Дурке и Пасторекве из 2012. године анализира макроекономске временске серије и пореди АРИМА и АРИМАКС моделе за њихово предвиђање, у овом случају АРИМА модел се показао бољи са грешком у апсолутном проценту од 1.77% у поређењу са АРИМАКС који има грешку од 3.78%. Рад [18] из 2014. пореди АРИМА модел и вештачке неуронске мреже при предвиђању цена акција фирме Дел (енг. *Dell*). Вештачка неуронска мрежа се показала за нијансу тачнија од АРИМА модела. Нешто скорији рад [19] из 2021 који се бави поређењем три модела: АРИМАКС, Икс-г

буст (енг. XGBoost) и Фејсбук профет (енг. *Facebook prophet*) за предвиђање цене Биткоина¹ установио је да је АРИМАКС незнатно бољи од остала два са кореном просека квадрата грешке од 322.

У раду [12] Амина Азарија тестиран је АРИМА модел у сврху предвиђања цене Биткоина један дан унапред. Подаци о цени су преузети за период од 3 године, а временска серија претпроцесирана како би јој се осигурала стационарност². Перформансе модела су тестиране над различитим потпериодима временске серије како би се проверила робустност модела. Утврђено је да је АРИМА модел користан само за краткорочна предвиђања дан унапред и да не може са тачношћу предвидети велике падове и раст цена. Овај рад се у великом делу ослања на Аминове закључке и покушава побољшати тачност у краткорочном предвиђању користећи додатне предиктивне варијабле.

Скалцото Ђовани у свом раду [15] анализира утицај коментара и постова на Твитеру³ на цену Етеријум (енг. *Ethereum*) и Доцкоин криптовалута. Тестира различите хиперпараметре АРИМАКС модела како би утврдио утицај предиктивних варијабли модела на цену криптовалута. Резултати показују да је утицај Твитера на цену Доцкоина значајан за разлику од Етеријума, пре свега јер је Доцкоин валута вођена популарношћу у заједници за разлику од Етеријума који има значајну технолошку вредност. Овакав закључак је битан јер говори да увођење података о заједници у модел за предвиђање Доцкоина може побољшати резултате.

Рад [16] са 14. конференције о Пајтону у науци се бави естимацијом хиперпараметара АРИМА модела преко графика аутокорејације и парцијалне аутокорејације. Рад сугерише да је пре било каквог тренирања модела потребно проверити стационарност временске серије комбинацијом КПСС и АДФ статистичких тестова и ако се утврди нестационарност потребно ју је уклонити, за уклањање се препоручује диференцирање временске серије. Рад указује да чисто ауторегресиван процес $AR(p)$ има график аутокорејације који опада осцилаторно или експоненцијално док се график парцијалне аутокорејације одсеца након p лагова. Процес чистог покретног просека $MA(q)$ има график аутокорејације који се нагло одсеца након q лагова док му график парцијалне аутокорејације опада

¹ Биткоин је прва и најпознатија криптовалута настала 2009 године. [13]

² Стационарна временска серија је она којој се статистички параметри не мењају кроз време. [14]

³ Твитер је популарна друштвена мрежа за дељење мишљења.

експоненцијално или осцилаторно. Код мешовитих процеса оба графика опадају експоненцијално или осцилаторно. Аутори рада такође сугеришу претрагу мреже хиперпараметара модела (енг. *Grid search*) што суштински значи поређење модела са различитим комбинацијама хиперпараметара а истим тренинг и тест скуповима података. Допринос рада [16] је врло значајан за естимацију хиперпараметара АРИМА модела и разумевање процеса који стоје иза функционисања модела, те ће се закључци и сугестије из њега применити на одабир хиперпараметара модела у овом раду.

Деви, Ума, Сундар и Али у раду [20] тестирају употребљивост АРИМА модела у дугорочном предвиђању цена акција на Индијској берзи за наредних 5 година. Подаци су прикупљани од 2009 до 2011 године, а временска серија диференцирана како би се осигурала стационарност. Параметри модела су процењивани преко графика аутокорељације и парцијалне аутокорељације, а модел евалуиран Акаике информационам критеријумом и грешком у апсолутном проценту. Резултати показују да АРИМА модел при дугорочним предвиђањима цена акција има најмању грешку од 16.26% а највећу од 38.12%. Овакви резултати индицирају да АРИМА модел није довољно добар за дугорочна предвиђања, али у исто време потврђују закључке рада [16] о естимацији параметара модела и претпроцесирању временске серије због нестационарности.

3. ТЕОРИЈСКИ ПОЈМОВИ И ДЕФИНИЦИЈЕ

Модел коришћен за предвиђање цена криптовалута је АРИМА и његова подврста АРИМАКС модел. У овом поглављу ће бити описане теоријске основе оба модела као и теоријске основе статистичких тестова примењених над моделима приликом естимације хиперпараметара. Потпоглавље 3.1 даје преглед математичких основа оба модела, а 3.2 појашњава тестове примењене приликом тестирања стационарности.

3.1 АРИМА и АРИМАКС модели

Бокс-Џенкинс методологија настала 70-их година 20. века је низ активности за идентификацију, естимацију и дијагностификовање АРИМА модела. Назив АРИМА потиче од скраћенице за његов назив на енглеском *Auto Regressive Integrated Moving Average* где се могу издвојити три његове главне компоненте:

- Ауторегресивна
- Интегрисана
- Покретни просек

Код АРИМА модела вредност променљиве која се предвиђа у тренутку t рачуна се као линеарна комбинација вредности у претходним тренуцима и регресионих грешака у претходним тренуцима [21].

Ауторегресивни модел p -тог реда, или скраћено $AR(p)$ се може описати једначином (1). Чист ауторегресивни модел вредност зависне променљиве у наредном моменту одређује на основу њених вредности у претходним моментима. [22] [21]

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t \quad (1)$$

где је:

Y_t – вредност зависне променљиве која се предвиђа у моменту t .

$Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$ – вредност зависне променљиве у моментима $t-1, t-2$ и $t-p$

ε_t – насумична грешка у моменту t .

$\varphi_0, \varphi_1, \varphi_2, \dots, \varphi_p$ – коефицијенти који се одређују при тренирању модела.

Модел покретног просека q -тог реда, или скраћено $MA(q)$ описује једначина (2). Модел покретног просека вредност зависне променљиве у наредном моменту одређује на основу константе и регресионих грешака у претходним моментима. [22][21]

$$Y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (2)$$

где је:

Y_t - вредност зависне променљиве која се предвиђа у моменту t .

ε_t - насумична грешка у моменту t .

$\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ - вредност регресионих грешака у претходним моментима $t-1, t-2$ и $t-q$

μ - константа.

$\theta_1, \theta_2, \dots, \theta_q$ - коефицијенти који се одређују при тренирању модела.

Интегрисана компонента модела представља се скраћено $I(d)$ где d означава број диференцирања који је потребно применити над временском серијом како би била стационарна. Диференцирање временске серије је одузимање вредности зависне променљиве у претходном моменту Y_{t-1} од вредности у тренутном моменту Y_t и то за сваки моменат у временској серији, где t припада скупу $\{1, 2, \dots, K\}$ а K је број момената у серији. [22] Модел $I(1)$ је према томе суштински само насумична шетња у математици (енг. *Random Walk*) и може је описати једначина (3).

$$Y_t - Y_{t-1} = \mu \rightarrow Y_t = \mu + Y_{t-1} \quad (3)$$

где је:

Y_t - вредност зависне променљиве која се предвиђа у моменту t .

μ - константа.

Y_{t-1} - вредност зависне променљиве у претходном моменту.

Комбинација ауторегресивног модела p -тог реда $AR(p)$ и модела покретног просека q -тог реда $MA(q)$ је АРМА модел реда p, q , скраћено $ARMA(p, q)$. Сходно томе, а имајући у виду једначине за ауторегресивни и модел покретног просека, општа једначина којом се може описати АРМА⁴ модел је (4):

$$Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q} \quad (4)$$

⁴ Под АРМА се овде подразумева АРИМА модел без специфицираног броја диференцирања.

Додатно, уколико желимо да укључимо и интегрисану компоненту d -тог реда $I(d)$ и специфицирамо број диференцирања над моделом, једначина (5) представља уопштење АРИМА модела реда p, d, q . [23]

$$\phi(L)(1-L)^d Y_t = \theta(L)\varepsilon_t \quad (5)$$

где је:

- Y_t - вредност зависне променљиве која се предвиђа у моменту t .
- $\phi(L)$ – полином ауторегресивне компоненте.
- $\theta(L)$ – полином компоненте покретног просека.
- d – оператор диференцирања d -тог степена.
- L – оператор лага (претходних момената)
- ε_t – насумична грешка у моменту t .

Пошто се често процеси, а нарочито они финансијски не могу добро описати само регресијом над претходним вредностима и регресионим грешкама некада је потребно увести додатне варијабле у једначину као предикторе. АРИМАКС модел представља проширење основног АРИМА модела додавањем варијабли које су исто временске серије као и зависна варијабла. Над додатим временским серијама врши се регресија на сличан начин као код претходних вредности зависне променљиве. У формули (6) је дата једначина АРИМАКС модела са k предиктивних варијабли [23].

$$\phi(L)(1-L)^d Y_t = \theta^1(L)X_t^1 + \dots + \theta^k(L)X_t^k + \theta(L)\varepsilon_t \quad (6)$$

где је:

- Y_t - вредност зависне променљиве која се предвиђа у моменту t .
- $\phi(L)$ – полином ауторегресивне компоненте.
- $\theta(L)$ – полином компоненте покретног просека.
- $\theta^k(L)$ – полином k -те⁵ предиктивне варијабле
- X_t^k – вредност k -те предиктивне варијабле у моменту t
- d – оператор диференцирања d -тог степена.
- L – оператор лага (претходних момената)
- ε_t – насумична грешка у моменту t .

⁵ Слово k овде не представља степен него редни број предиктивне варијабле

3.2 Тестови стационарности

Шок у временској серији представља изненадну промену у варијабли или грешци предвиђања у одређеном временском периоду. Уколико се шок деси у стационарној временској серији његов ефекат се губи након коначно много времена, али код нестационарне серије ефекат шока је трајан. Ако код процеса $AR(p)$ описаног једначином $\phi(L)Y_t = \mu + \varepsilon_t$, где је $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ важи да је $\Phi(1) = 0$ или $\phi_1 + \phi_2 + \dots + \phi_p = 1$ каже се да Y_t има **јединични корен** (енг. *Unit Root*) и да није стационаран [24]. Јединични корен у временској серији се може уклонити диференцирањем серије.

У наредна 2 потпоглавља биће описани АДФ и КПСС тестови стационарности који се наслањају на хипотезу о јединичном корену.

3.2.1 АДФ тест

АДФ (енг. *Augmented Dickey Fuller*) тест је један од најстаријих и најпопуларнијих тестова који проверавају јединични корен временске серије изложен у радовима [25][26] Дикија и Фулера. АДФ је проширење основне верзије ДФ теста ради тестирања регресивних процеса вишег реда. У наставку је дат пример ДФ теста над моделом првог реда, АДФ је само проширење описаног модела.

Ако имамо процес $AR(1)$ описан једначином $y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t$ за њега можемо издвојити 2 битна случаја:

1. $\phi_1 = 1$ где процес има јединични корен и није стационаран
2. $\phi_1 < 1$ где нема јединични корен и стационаран је

Даљом репараметризацијом једначине процеса и одузимањем y_{t-1} од y_t добијамо $\Delta y_t = \mu + (\phi_1 - 1)y_{t-1} + \varepsilon_t$ где је Δ оператор диференцирања а $(\phi_1 - 1)$ се краће записује као α_0 које мора бити 0 како би била потврђена хипотеза о јединичном корену. Тестирање хипотезе се врши t тестом (7) [24]

$$DF_T = \frac{\phi - 1}{SE(\phi)} \quad (7)$$

где је:

DF_T – статистика теста.

$SE(\phi)$ – стандардна грешка при тестирању.

Статистика теста се упоређује са критичним вредностима из табеле вредности за t расподелу теста. Уколико је добијена вредност мања од

критичних вредности из табеле за тестирани број лагова хипотеза о постојању јединичног корена се одбацује [25][26].

У програмском језику Пајтон АДФ тест се може извршити помоћу библиотеке *statsmodels* и њене функције *adfuller* [27] која враћа статистику теста, p вредност и број лагова над којима је вршено тестирање. Добијена p вредност се тумачи на следећи начин:

1. за $p < 0.05$ хипотеза о јединичном корену се одбацује
2. за $p \geq 0.05$ хипотеза о јединичном корену се потврђује

P вредност се добија упоређивањем статистике теста као што је раније објашњено и може се засебно посматрати ради утврђивања стационарности.

3.2.2 КПСС тест

КПСС (енг. *Kwiatkowski-Phillips-Schmidt-Shin*) тест је као и АДФ тест начин да се провери постојање јединичног корена у временској серији. Битна разлика између њих је што је хипотеза КПСС теста да временска серија **нема** јединични корен, то јест да је стационарна. КПСС за разлику од АДФ-а може тестирати стационарност у присуству детерминистичког тренда, дакле чак и ако постоји константан нагиб у временској серији КПСС неће нужно показати да је серија нестационарна. Овакво понашање се објашњава тиме да стационарне временске серије са детерминистичким трендом имају тенденцију да се опораве од шока након коначно много времена. [24][28].

У Пајтону КПСС тест се може извршити помоћу библиотеке *statsmodels* и функције *kps* која враћа статистику теста, p вредност и број лагова над којима је вршено тестирање. Добијена p вредност се тумачи обрнуто у односу на АДФ тест:

1. за $p < 0.05$ хипотеза о непостојању јединичног корена се одбацује
2. за $p \geq 0.05$ хипотеза о непостојању јединичног корена се потврђује

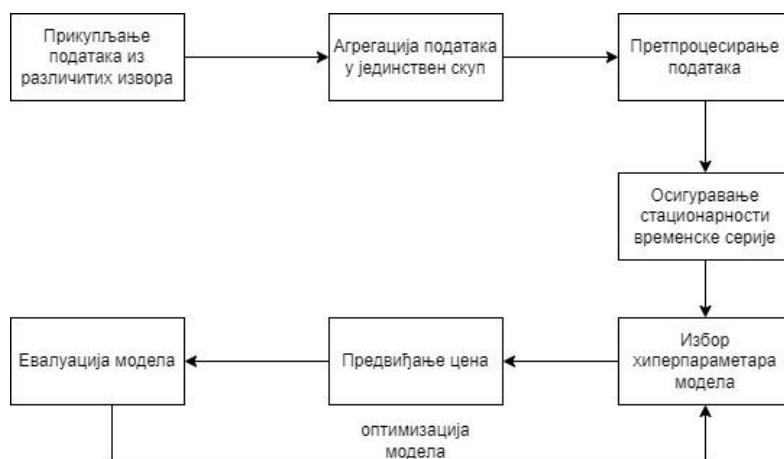
Статистика теста се упоређује са критичним вредностима из табеле теста. Уколико је добијена вредност већа од критичне вредности из табеле хипотеза о непостојању јединичног корена се одбацује [28]

4. МЕТОДОЛОГИЈА

У овом поглављу ће бити описана методологија предвиђања цена криптовалута коришћена у овом раду. Прво ће у 4.1 бити дат опис целог тока предвиђања и набројани сви кораци од којих се састоји, а затим редом у 4.2 процес прикупљања и агрегације података, у 4.3 претпроцесирање, у 4.4 осигуравање стационарности временске серије и у 4.5 избор хиперпараметара предвиђање и евалуација.

4.1 Опис методологије

Методологија за предвиђање цена криптовалута у овом раду састоји се из више корака, интерне имплементације корака нису нужно исте за АРИМА и АРИМАКС модел али је скуп корака исти. У наставку на слици 4.1.1 је дат дијаграм који приказује кораке у методологији, скуп корака је исти за сваку од три криптовалуте па се неће описивати за сваку појединачно.



слика 4.1.1 Кораци методологије предвиђања

Прикупљање и агрегација података су вршени једном и тако прикупљени подаци су даље претпроцесирани и коришћени на различит начин у зависности од модела и криптовалуте. Осигуравање стационарности је поступак испитивања стационарности временске серије и њена трансформација одговарајућим техникама како би се осигурала стационарност. Избор хиперпараметара, предвиђање и

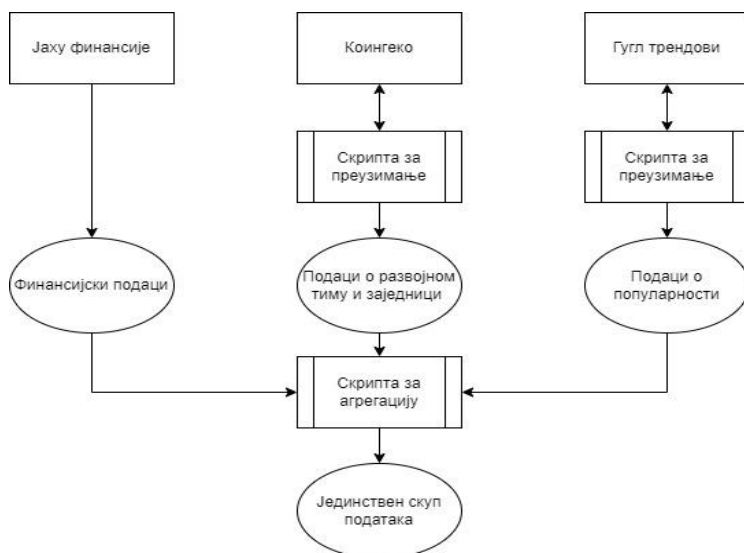
евалуација су у различитим експериментима изведени и као спојен и као одвојен процес.

4.2 Прикупљање и агрегација података

Формиран је скуп података за сваку од криптовалута преузимањем података из различитих извора и њиховим спајањем у јединствен скуп. Пошто су ова два поступка прикупљања и агрегације доста повезана биће описани заједно. За потребе овог рада подаци о криптовалутама Шиба Ину, Идена и Доцкоин су прикупљани из три извора:

1. Јаху финансије [29] (енг. *Yahoo Finance*) одакле су преузети финансијски подаци
2. Коингеко [30] (енг. *CoinGecko*) одакле су преузети подаци о заједници и развојном тиму
3. Гугл трендови [31] (енг. *Google trends*) одакле је преузета метрика популарности

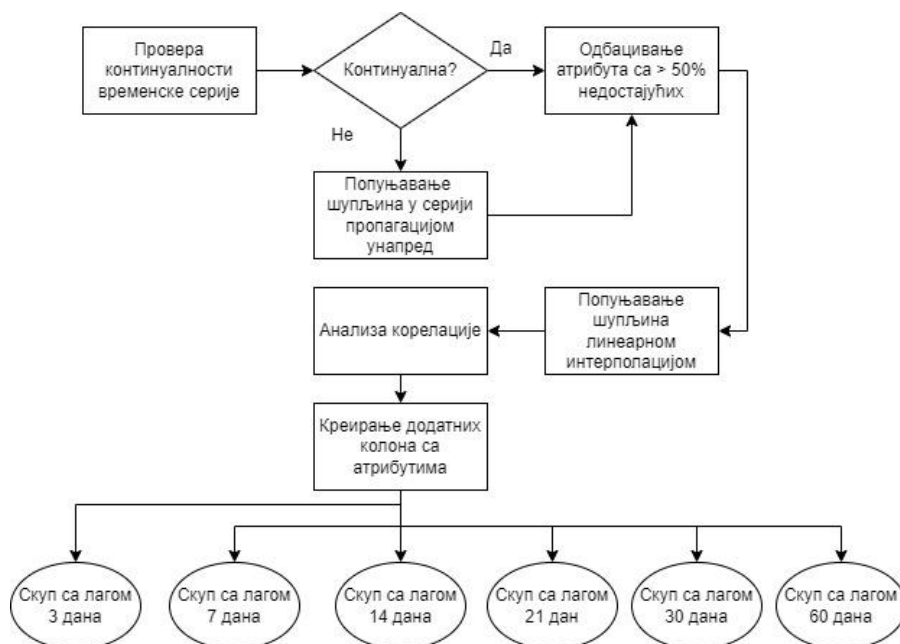
Подаци са Јаху финансија су преузети ручно, скидањем ЦСВ фајла са подацима за одабрани временски период. Подаци са Коингека и Гугл трендова су скинути писањем Пајтон скрипта које шаљу упите ка њиховим апијима и преузимају податке за сваки дан из специфицираног интервала. Агрегација података је вршена према датуму на начин да су подаци из сва три овако преузета скупа спојени пресеком по датуму. Овај процес је представљен дијаграмом на слици 4.2.1.



слика 4.2.1 Дијаграм процеса преузимања и агрегације података

4.3 Претпроцесирање података

Претпроцесирање података представља процес припреме агрегираних скупова података за даље убацивање у модел ради тренирања и евалуације модела. Пре претпроцесирања података скупови су анализирани статистички и визуелно како би се утврдио број обележја и број недостајућих података за свако обележје. На слици 4.3.1 је дат дијаграм поступка претпроцесирања а у наставку ће бити описане методе примењене у фази претпроцесирања.



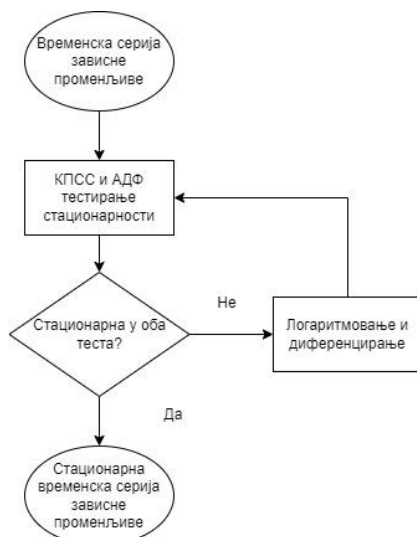
слика 4.3.1 Дијаграм претпроцесирања података

Пре свега је проверена континуалност временске серије и уколико је за серију утврђено да постоје дани који недостају попуњавани су пропагацијом последњег постојећег дана унапред (енг. *Forward fill*). Након утврђивања континуалности уклањани су атрибути који имају више од 50% недостајућих вредности јер би се њиховим попуњавањем могао направити нетачан модел. Атрибути са мање од 50% недостајућих вредности су попуњени линеарном интерполацијом јер се претпоставља да се на малим интервалима шупљина вредности крећу линеарно. У наредном кораку анализирана је корелација међу атрибутима у скупу података помоћу топлотних мапа (енг. *heatmap*), корелација је анализирана за лагове од 3, 7, 14, 21,

30 и 60 дана. Тамо где је утврђено да је корелација независног и зависног атрибута мања од 0.6, независни атрибут је избациван из скупа због мале релевантности. Ако је утврђено да је корелација два независна атрибута преко 0.9 један је избациван како би се избегао проблем мултиколинеарности. Битно је напоменути да пошто се ради о 6 засебних анализа корелације излаз из овог корака је 6 различитих скупова података, на дијаграму су засебни скупови приказани на крају како се не би понављали. Последњи корак у фази претпроцесирања је креирање додатних колона које чувају средњу вредност и стандардну девијацију за сваки атрибут и то за покретне прозоре од 3, 7, 14, 21, 30 и 60 дана.

4.4 Осигуравање стационарности

Пошто је један од предуслова за употребу АРИМА фамилије модела стационарност временске серије која се предвиђа, битан корак је провера и осигуравање стационарности. Иако суштински део претпроцесирања овде је издвојен због свог значаја и приказан на дијаграму 4.4.1.



слика 4.4.1 Дијаграм процеса осигуравања стационарности временске серије

Први корак у овом поступку је провера стационарности временске серије зависне променљиве КПСС и АДФ тестовима описаним у поглављу 3.2. Пошто један тест не искључује други овде се користе заједно. Ако се на основу метрике оба теста утврди стационарност временска серија се не процесира и ту је овај поступак

готов. Међутим ако барем један од тестова покаже нестационарност серија се прво логаритмује а затим диференцира онолико пута колико је потребно да би оба теста показала стационарност. Битно је напоменути да се пре овог процеса прави копија временске серије и над њом врше диференцирања јер је значај овог корака у утврђивању броја диференцирања потребних за достизање стационарности, што је суштински параметар d у АРИМА фамилији модела.

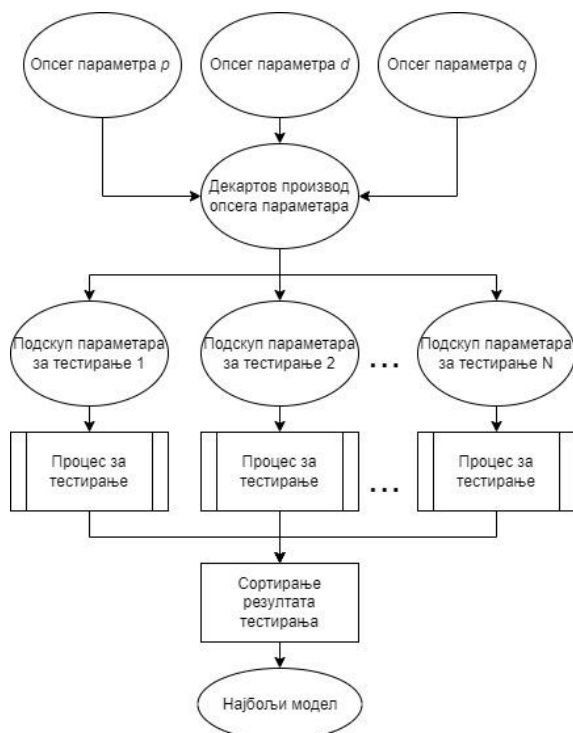
4.5 Избор хиперпараметара модела, предвиђање и евалуација

У овом раду примењене су две стратегије избора хиперпараметара модела чији је опис у наставку.

Прва стратегија подразумева коришћење функције *auto_arima* из Пајтонове библиотеке *pmdarima* [32] која аутоматски претражује комбинације хиперпараметара модела и бира најбољу комбинацију према Акаике информационом критеријуму. Функцији је могуће одредити опсеге вредности у којима ће вршити претрагу, али не постоји гаранција да ће она заиста претражити све могуће комбинације и наћи најбољу. Након што функција врати најбољу комбинацију хиперпараметара и вредност Акаике информационог критеријума врши се тренирање модела са тим параметрима над тренинг скупом. Тренирани модел се даље користи за предвиђање K корака унапред и рачуна се грешка у процентима на основу које се евалуира модел.

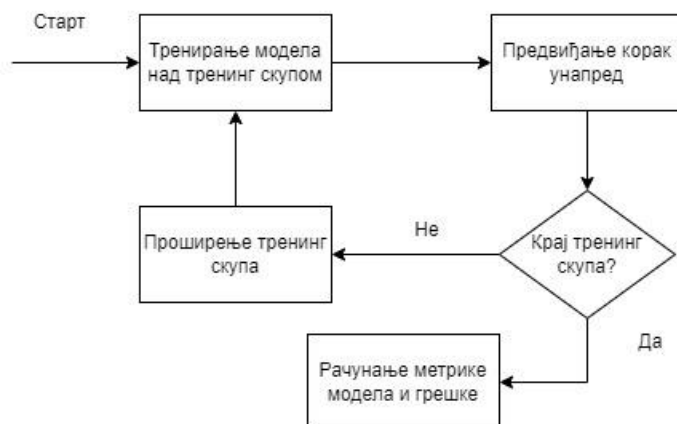
Друга стратегија подразумева имплементацију претраге свих комбинација хиперпараметара из задатог опсега и бира најбоље комбинације према просечној релативној грешци предвиђања и Акаике информационом критеријуму. Најпре се одреде опсези вредности ауторегресивне, интегрисане и компоненте покретног просека које ће бити тестиране. Затим се рачуна декартов производ та три опсега и добија скуп комбинација хиперпараметара који је потребно тестирати. Овако добијен скуп се дели на N подскупова и сваки се шаље посебном процесу на рачунару који врши тестирање свих комбинација из тог подскупа. Подела се врши на онолико подскупова колико процесор рачунара има језгара како би се постигао потпуни паралелизам тестирања и убрзао цео процес избора хиперпараметара. Процеси након завршетка враћају за сваку тестирану комбинацију хиперпараметара вредност Акаике информационог критеријума и апсолутног процента грешке. Резултати се затим

сортирају у растућем редоследу по вредности обе метрике и узима прва комбинација као најбоља. На слици 4.5.1 је дат дијаграм ове стратегије.



слика 4.5.1 Дијаграм процеса избора хиперпараметара

Тестирање комбинације хиперпараметара се врши динамичким предвиђањем над тест скупом података. Прво се тренира модел са задатом комбинацијом хиперпараметара над делом временске серије одабраним за тренинг. Онда се врши предвиђање цене корак унапред и добијена вредност памти. Затим се стварна вредност додаје у тренинг скуп и модел поново тренира и врши предвиђање, цео процес се понавља за сваку вредност у тест скупу и на крају рачуна Акаике информациони критеријум финалног модела и грешка у процентима за сва предвиђања из тест скупа. Поред ове, у раду је примењена и мало измењена верзија ове стратегије где се предвиђена вредност цене додаје у тренинг скуп уместо стварне. Уколико се стварна вредност дода у тренинг скуп то подразумева само могућност тестирања модела за предвиђање корак унапред, међутим ако се дода предвиђена вредност онда се може тестирати предвиђање и више дана унапред. Скица овог процеса је дата на слици 4.5.2.



слика 4.5.2 Дијаграм процеса тестирања комбинације хиперпараметара

5. ЕКСПЕРИМЕНТИ

Циљ поглавља експерименти је да опише све спроведене експерименте и скупове података над којима су спроведени. У првом потпоглављу ће бити детаљно описани скупови података, њихови атрибути и извршена претпроцесирања над њима. У другом ће бити дат опис тестирања стационарности све 3 временске серије, у трећем спецификација рачунара над којим су спроведени експерименти, а у наредна 4 потпоглавља биће описани спроведени експерименти. У последњем потпоглављу је објашњена метрика евалуације модела.

5.1 Скупови података

Пошто су у раду предвиђане цене криптовалута Шиб Ину, Доцкоин и Идена креирани су скупови података за све три валуте према методологији описаној у 4.2. Скупови података, са изузетком метрике популарности, садрже исте атрибуте и то:

- Преузето са Јаху финансија
 - Date – датум
 - Open – цена на почетку дана
 - High – највећа дневна цена
 - Low – најнижа дневна цена
 - Close – цена на крају дана
 - Adj Close – коригована цена на крају дана
 - Volume – дневни промет валуте
- Преузето са Коингека
 - facebook_likes – број свиђања на Фејсбуку⁶
 - twitter_followers – број пратилаца на Твитеру
 - reddit_average_posts_48h – просечан број нових тема за 48 сати у Редит⁷ заједници криптовалуте

⁶ Фејсбук је популарна друштвена мрежа [33]

⁷ Редит је друштвена мрежа налик на форум, свака већа криптовалута имају своју заједницу на њој [34]

- `reddit_average_comments_48h` – просечан број коментара за 48 сати у редит заједници.
- `reddit_subscribers` – број претплатника на Редиту
- `reddit_accounts_active_48h` – број корисничких налога активних за 48 сати
- `forks` – број форкова⁸ репозиторијума кода криптовалуте
- `stars` – број звезда на Гитхабу⁹
- `subscribers` – број претплатника на Јутјубу¹⁰
- `total_issues` – укупан број пријављених проблема на Гитхабу
- `closed_issues` – број разрешених проблема на Гитхабу
- `pull_requests_merged` – број одобрених захтева за повлачење кода
- `pull_request_contributors` – број људи који су допринели развоју кода
- `commit_count_4_weeks` – број комитова за 4 недеље

➤ Преузето са Гугл трендова

- `(dogecoin, idena, shiba_inu)`¹¹ – метрика популарности за криптовалуту скалирана на опсег [0,1]
- `(dogecoin, idena, shiba_inu)_unscaled` – нескалирана метрика популарности криптовалуте
- `(dogecoin, idena, shiba_inu)_monthly` – месечна метрика популарности
- `scale` – фактор скалирања метрике популарности

У наредна три потпоглавља дат је опис сваког од три скупа података.

5.1.1 Доцкоин скуп

Подаци за Доцкоин су прикупљани на дневном нивоу за период од 9.11.2017. до 4.2.2022 са сва три апија. Пре претпроцесирања у скупу је било 1549 редова са подацима и 25 атрибута. Статистичком анализом скупа утврђено је да 12 атрибута има недостајуће податке,

⁸ Форковање је копирање туђег целог репозиторијума на сопствени налог

⁹ Гитхаб је платформа за контролу верзије кода [35]

¹⁰ Јутјуб је платформа за дељење видео садржаја [36]

¹¹ Атрибут има различит назив у складу са криптовалutom

од тога атрибут *facebook_likes* више од 50% недостајућих и он је избачен из скупа. Недостајуће вредности осталих атрибута: *reddit_subscribers*, *reddit_accounts_active_48h*, *forks*, *stars*, *subscribers*, *total_issues*, *closed_issues*, *pull_requests_merged*, *pull_request_contributors*, *commit_count_4_weeks*, *twitter_followers* су попуњене линеарном интерполацијом.

Анализом корелације лагованих вредности атрибута за лаг од 3 дана утврђено је да је потребно избацити атрибуте: *Volume*, *reddit_average_comments_48h*, *dogecoin*, *reddit_accounts_active_48h*, *commit_count_4_weeks*, *stars*, *High*, *Low*, *twitter_followers*, *reddit_subscribers*, *forks*, *subscribers*, *total_issues*, *pull_requests_merged*.

Анализом корелације лагованих вредности атрибута за лаг од 7 дана утврђено је да је потребно избацити атрибуте: *Volume*, *reddit_average_comments_48h*, *dogecoin*, *reddit_accounts_active_48h*, *commit_count_4_weeks*, *stars*, *High*, *Low*, *twitter_followers*, *reddit_subscribers*, *forks*, *subscribers*, *total_issues*, *pull_requests_merged*.

Анализом корелације лагованих вредности атрибута за лаг од 14 дана утврђено је да је потребно избацити атрибуте: *Volume*, *reddit_average_comments_48h*, *High*, *reddit_accounts_active_48h*, *commit_count_4_weeks*, *pull_requests_merged*, *dogecoin*, *twitter_followers*, *Low*, *reddit_subscribers*, *forks*, *stars*, *subscribers*, *total_issues*.

Анализом корелације лагованих вредности атрибута за лаг од 21 дан утврђено је да је потребно избацити атрибуте: *Volume*, *reddit_average_comments_48h*, *dogecoin*, *reddit_accounts_active_48h*, *commit_count_4_weeks*, *High*, *Low*, *twitter_followers*, *reddit_subscribers*, *forks*, *stars*, *subscribers*, *total_issues*, *pull_requests_merged*.

Анализом корелације лагованих вредности атрибута за лагове од 30 и 60 дана утврђено је да је потребно избацити атрибуте: *Volume*, *reddit_average_comments_48h*, *dogecoin*, *dogecoin_monthly*, *reddit_accounts_active_48h*, *commit_count_4_weeks*, *High*, *Low*, *twitter_followers*, *reddit_subscribers*, *forks*, *stars*, *subscribers*, *total_issues*, *pull_requests_merged*.

У скуп података су на крају убачене нове колоне колоне које садрже средњу вредност и стандардну девијацију за сваку варијаблу из

скупа без обзира на анализу корелације. Исто је урађено за св остале криптовалуте.

5.1.2 Шибa Ину скуп

Подаци из Шибa Ину скупа су прикупљани на дневном нивоу за период од 1.8.2020. до 4.2.2022. Пре претпроцесирања у скупу су била 553 реда са подацима и 25 атрибута. Утврђено је да 17 атрибута има недостајуће податке, од тога 10 више од 50% недостајућих и то атрибуту: *facebook_likes*, *forks*, *stars*, *subscribers*, *total_issues*, *closed_issues*, *pull_requests_merged*, *pull_request_contributors*, *commit_count_4_weeks*, *twitter_followers*. Недостајуће вредности атрибута: *reddit_average_posts_48h*, *reddit_average_comments_48h*, *reddit_subscribers*, *reddit_accounts_active_48h*, *shiba_inu_coin_monthly*, *scale*, *shiba_inu_coin* су попуњене линеарном интерполацијом.

Анализом корелације лагованих вредности атрибута за лаг од 3 дана утврђено је да је потребно избацити атрибуте: *shiba_inu_coin_unscaled*, *High*, *Low*, *shiba_inu_coin_monthly*.

Анализом корелације лагованих вредности атрибута за лаг од 7 дана утврђено је да је потребно избацити атрибуте: *shiba_inu_coin_unscaled*, *High*, *Low*, *shiba_inu_coin_monthly*.

Анализом корелације лагованих вредности атрибута за лагове од 14 и 21 дана утврђено је да је потребно избацити атрибуте: *Volume*, *reddit_average_comments_48h*, *reddit_accounts_active_48h*, *shiba_inu_coin_unscaled*, *shiba_inu_coin*, *High*, *Low*, *shiba_inu_coin_monthly*.

Анализом корелације лагованих вредности атрибута за лаг од 30 дана утврђено је да је потребно избацити атрибуте: *Volume*, *reddit_average_comments_48h*, *reddit_accounts_active_48h*, *shiba_inu_coin_monthly*, *scale*, *shiba_inu_coin_unscaled*, *shiba_inu_coin*, *High*, *Low*.

Анализом корелације лагованих вредности атрибута за лаг од 60 дана утврђено је да је услед слабе корелације са циљном променљивом потребно избацити све осим *reddit_subscribers*.

5.1.3 Идена скуп

Подаци за Идена скуп су прикупљани на дневном нивоу за период од 11.8.2020. до 4.2.2022. Пре претпроцесирања у скупу су била 543 реда са подацима и 25 атрибута. Утврђено је да 10 атрибута има недостајуће податке, од тога 9 више од 50% недостајућих и то атрибути: *facebook_likes*, *forks*, *stars*, *subscribers*, *total_issues*, *closed_issues*, *pull_requests_merged*, *pull_request_contributors*, *commit_count_4_weeks*. Недостајуће вредности атрибута: *twitter_followers* су попуњене линеарном интерполацијом.

Анализом корелације лагованих вредности атрибута за лаг од 3 дана утврђено је да је потребно избацили атрибуте: *Volume*, *twitter_followers*, *idena_unscaled*, *reddit_average_posts_48h*, *reddit_average_comments_48h*, *idena_monthly*, *scale*, *reddit_subscribers*, *reddit_accounts_active_48h*, *idena*, *High*, *Low*.

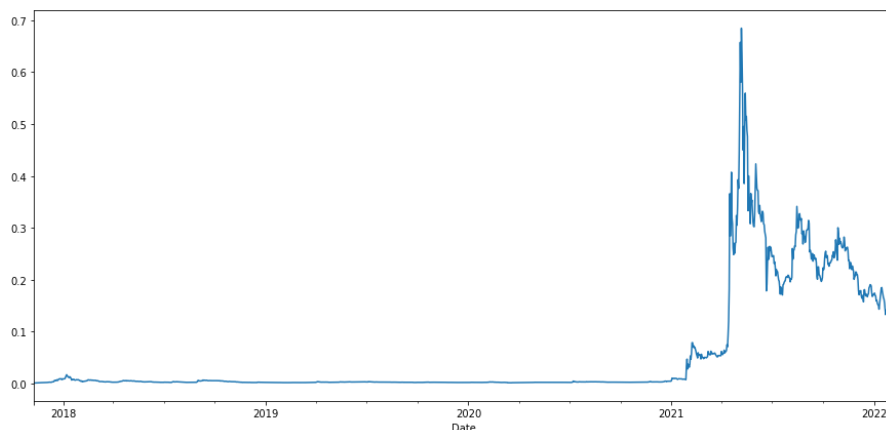
Анализом корелације лагованих вредности атрибута за лаг од 7 дана утврђено је да је потребно избацили атрибуте: *Volume*, *twitter_followers*, *idena_unscaled*, *reddit_average_posts_48h*, *reddit_average_comments_48h*, *idena_monthly*, *scale*, *reddit_subscribers*, *reddit_accounts_active_48h*, *idena*, *High*, *Low*.

Анализе корелација за лагове од 14, 21, 30 и 60 дана показале су да све независне променљиве у слабој корелацији са зависном те их је све потребно избацили.

5.2 Тестирање стационарности серија

Пошто је предуслов за употребу АРИМА и АРИМАКС модела стационарност временске серије, а сви експерименти се наслањају на та 2 модела онда је за њихово извођење потребно тестирање и осигуравање стационарности временских серија циљних варијабли.

За временску серију циљног атрибута Доцкоин скупа (*Close*) (слика 5.2.1) АДФ тест је показао статистику из табеле 5.2.1. Статистика теста има вредност већу од свих критичних вредности, а p вредност је већа од 0.05 што потврђује хипотезу о нестационарности серије.

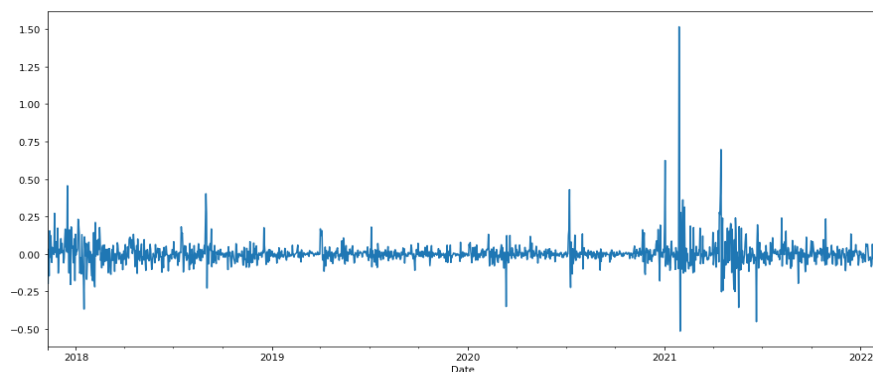


слика 5.2.1 Временска серија цене Доџкоина.

Тест	P вредност	Статистика теста	1% критична вредност	5% критична вредност	10% критична вредност
АДФ	0.287	-1.998	-3.434	-2.863	-2.567
КПСС	0.010	2.901	0.739	0.463	0.347

табела 5.2.1 Статистика АДФ и КПСС тестова Доџкоин временске серије.

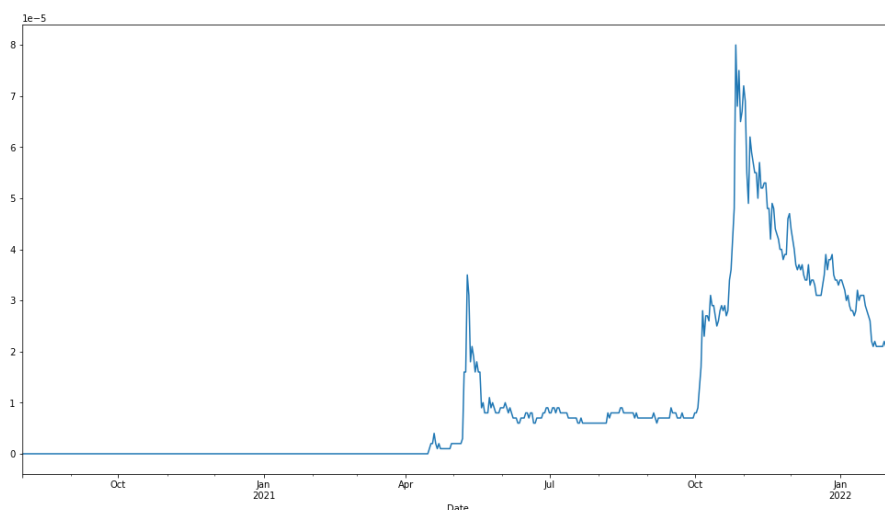
КПСС тест је показао статистику из табеле 5.2.1 са p вредношћу мањом од 0.05 и статистиком теста већом од свих критичних вредности, што говори да се хипотеза о стационарности серије одбацује. Након логаритмовања временске серије и диференцирања једном добијена је временска серија са слике 5.2.2 а АДФ и КПСС тестови су показали стационарност што се види у табели 5.2.2.



слика 5.2.2 Временска серија цене Доџкоина након логаритмовања и диференцирања.

Тест	<i>P</i> вредност	Статистика теста	1% критична вредност	5% критична вредност	10% критична вредност
АДФ	0.000	-20.718	-3.434	-2.863	-2.567
КПСС	0.100	0.181	0.739	0.463	0.347

табела 5.2.2 Статистика АДФ и КПСС тестова Доцкоина након диференцирања и логаритмовања



слика 5.2.3 Временска серија цене ШИБА Ину.

Анализом временске серије циљног атрибута ШИБА Ину скупа (слика 5.2.3) АДФ и КПСС тестовима утврђена је нестационарности (табела 5.2.3) у оба теста.

Тест	<i>P</i> вредност	Статистика теста	1% критична вредност	5% критична вредност	10% критична вредност
АДФ	0.595	-1.371	-3.442	-2.866	-2.569
КПСС	0.010	2.206	0.739	0.463	0.347

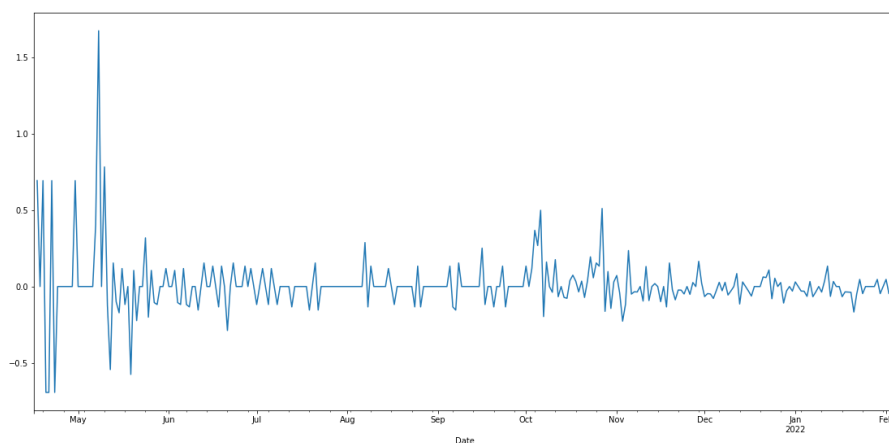
табела 5.2.3 Статистика АДФ и КПСС тестова ШИБА Ину

Након логаритмовања и једног диференцирања статистика тестова из табеле 5.2.4 показала је стационарност. Међутим како би логаритмовање могло бити урађено уклоњени су сви дани када је цена била 0, што је период од почетка временске серије до краја априла

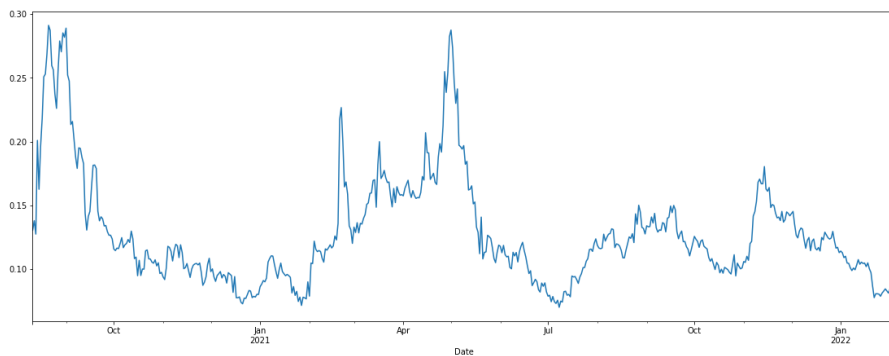
2021. Изглед временске серије након диференцирања и логаритмовања је дат на слици 5.2.4.

Тест	<i>P</i> вредност	Статистика теста	1% критична вредност	5% критична вредност	10% критична вредност
АДФ	5.382e-6	-5.303	-3.454	-2.872	-2.572
КПСС	0.146	0.100	0.739	0.463	0.347

табела 5.2.4 Статистика АДФ и КПСС тестова Шиба Ину након логаритмовања и диференцирања.



слика 5.2.4 Временска серија цене Шиба Ину након логаритмовања и диференцирања.



слика 5.2.5 Временска серија затварајуће цене Идена криптовалуте.

Провером временске серије циљне варијабле Идене (слика 5.2.5) АДФ и КПСС тестовима добијени су мешани резултати (табела 5.2.5). КПСС тест је показао потпуну стационарност, међутим АДФ тест је на

граници што се тиче p вредности, а упоређивањем статистике теста са критичним вредностима може се рећи да је статистика теста између 5 и 10% критичних вредности. Како би се у овом случају избегло предиференцирање серије одлучено је да се временска серија сматра стационарном и не диференцира даље.

Тест	P вредност	Статистика теста	1% критична вредност	5% критична вредност	10% критична вредност
АДФ	0.050	-2.855	-3.442	-2.866	-2.569
КПСС	0.289	0.100	0.739	0.463	0.347

табела 5.2.5 Статистика АДФ и КПСС тестова Идена валуте.

5.3 Спецификације рачунара и софтвера

Спецификација рачунара над којим су спроведени експерименти за потребе овог рада је дата у табели 5.3.1.

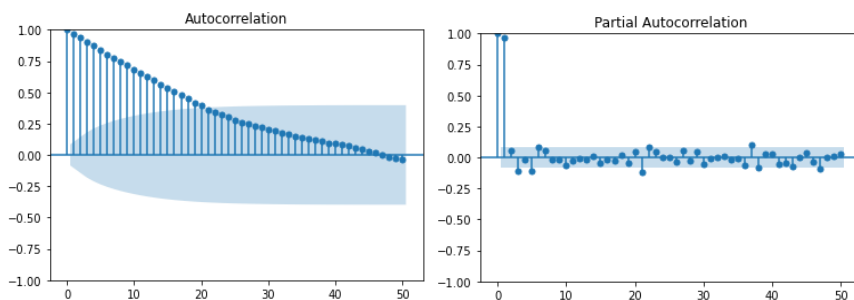
процесор	AMD FX-6300
број језгара	6
радни такт	3.5 GHz
опер. систем	Windows 10 21H1
РАМ меморија	24 GB, 1600 MHz, DDR3
окружење	Pycharm 2021.3.3
језик	Python 3.10

табела 5.3.1 Спецификације софтвера и хардвера тест рачунара.

5.4 Експеримент 1

У првом експерименту предвиђане су цене све три криптовалуте дан унапред употребом АРИМА модела. Коришћени су раније описани скупови података с тим да је употребљен једино атрибут *Close* који је у овом и свим наредним експериментима циљна варијабла за предвиђање. За потребе експеримента Шиба Ину и Доцкоин скупови су подељени у односу 90% тренинг – 10% тест, а Идена скуп у односу 70% тренинг – 30% тест. Доцкоин и Шиба Ину имају 2 велика скока у цени па је узето у обзир да те скокове треба ставити у тренинг скуп како би предвиђање било тачније. Сви скупови су креирани континуалном поделом, првих $x\%$ за тренинг остатак за тест. Из временске серије Шиба Ину криптовалуте су уклоњени дани с почетка код којих је цена 0.

Анализом графика аутокорејације и парцијалне аутокорејације за Доцкоин и Шиба Ину утврђено је да оба имају графике који оштро одсецају након нултог лага, а затим вредности осцилују и повремено се издижу изнад области поверења. Оваква сазнања нису допринела сужавању опсега параметара, те је опсег претраге за ове 2 валуте одређен произвољно. Анализом графика аутокорејације и парцијалне аутокорејације за Идена временску серију (слика 5.4.1) примећено је одсецање након једног лага на графику парцијалне аутокорејације и експоненцијално опадање на графику аутокорејације. Према томе, а у складу са закључцима рада [16] може се рећи да је адекватан модел $AR(1)$.



слика 5.4.1 Графици аутокорејације и парцијалне аутокорејације за Идену.

Хиперпараметри модела су претраживани методологијом описаном у 4.5, а опсези параметара који су тестирани су:

- За Доцкоин $p[0-9]$, $q[0-9]$ и $d=1$ због броја диференцирања до стационарности.
- За Шиба Ину $p[0-9]$, $q[0-9]$ и $d=1$ због броја диференцирања до стационарности.
- За Идену $p[0-9]$, $q[0-9]$ и $d[0-1]$ како би се тестирало понашање са и без диференцирања.

Опсези су одређени произвољно, али имајући у виду број диференцирања потребан до стационарности и чињеницу да се модели вишег реда где су вредности параметара изнад 7 тренирају много дуже и захтевају много више процесорског времена. Претрага је трајала више од 25 часова за све 3 криптовалуте у збиру и укупно је тестирано по 100 различитих модела за Шиба Ину и Доцкоин криптовалуте и 200 модела за Идену.

Прво је тестирано предвиђање дан унапред за Идена криптовалуту са параметрима $ARIMA(1,0,0)$ одређеним преко графика

са слике 5.4.1, по корацима са слике 4.5.2. Добијени чисто ауторегресивни модел је над тест скупом имао процентуалну грешку од 3.74%.

Претрага хиперпараметара за Доцкоин пронашла је да је најбољи модел за предвиђање дан унапред $ARIMA(5,1,2)$ са грешком у апсолутном проценту од 3.34% и вредношћу Акаике информационог критеријума од 3350.

Најбољи Шиба Ину модел је био $ARIMA(3,1,0)$ са Акаике информационим критеријумом у вредности од 152 и грешком у апсолутном проценту од 3.59%.

Претрага хиперпараметара модела за Идену је дала модел $ARIMA(1,1,0)$ као најбољи, са процентуалном грешком од 3.63% и Акаике информационим критеријумом од 1323.

5.5 Експеримент 2

У оквиру другог експеримента предвиђана је цена криптовалута више дана унапред помоћу АРИМАКС модела. Стратегија одабира хиперпараметара модела била је помоћу *auto_arima* функције. У оквиру експеримента су тестирани и различити односи поделе скупова података на тренинг и тест подскупове како би се проверило који однос даје најбоље резултате. Односи поделе скупа који су тестирани су 70:30, 80:20 и 90:10. Као предиктивне варијабле коришћене су лаговане вредности атрибута из скупова за лагове од 3, 7, 14, 21, 30, 60 дана.

Прва подела скупа за Шиба Ину била је у односу 70% тренинг – 30% тест где се у тренинг скупу нису нашли подаци о другом великом скоку цене из октобра 2021. Претрагом хиперпараметара узимајући у обзир све периоде лагова *auto arima* је показала да је за лагове од 3, 7, 14 и 21 дана најбољи модел $ARIMA(1,0,0)$. За лагове од 30 и 60 дана претрага је вратила модел $ARIMA(0,1,0)$. За другу поделу од 80%-20% претрага у тренинг скупу се нашао први део скока цене из октобра који је трајао 2 недеље и завршен је заравнањем. Претарага хиперпараметара је показала да је за лагове од 3, 7, 14, 21 и 30 дана најбољи модел $ARIMA(1,0,0)$ док је за лаг од 60 дана најбољи модел био $ARIMA(0,1,0)$. Подела у односу 90:10 је је обухватила оба велика скока цене у целости у тренинг скуп, а претрага хиперпараметара је вратила исте моделе као и за 80:20.

Пошто је анализом корелације за Идену из поглавља 5.1.3 утврђено да је за лагове од 14, 21, 30 и 60 дана потребно избацити све независне променљиве овде су тестирана предвиђања само за лагове од 3 и 7 дана. Код поделе у односу 70:30 за лаг од 3 дана претрага је вратила модел $ARIMA(1,0,0)$, а за 7 дана модел $ARIMA(2,0,1)$. За поделу у односу 80:20 су се као најбољи показали исти модели као и за 70:30, док су за поделу у односу 90:10 најбољи модели били $ARIMA(1,0,2)$ за лаг од 3 дана и $ARIMA(2,0,1)$ за 7 дана.

За предвиђање цене Доцкоина код поделе у односу 70:30 у тренинг скупу се нису нашли подаци о великом скоку цене који се десио у почео у фебруару 2021 и наставио се у априлу исте године. Претрага хиперпараметара модела *auto arima* функцијом је као најбоље моделе вратила редом $ARIMA(2,0,0)$, $ARIMA(3,0,0)$, $ARIMA(0,1,1)$, $ARIMA(0,1,1)$, $ARIMA(1,1,2)$, $ARIMA(1,1,0)$ за лагове од 3, 7, 14, 21, 30 и 60 дана. Тренинг тест подела у односу 80:20 је у тренинг скупу обухватила први део великом пораста цене из фебруара и почетак другог дела пораста у априлу. Претрага хиперпараметара је за лагове од 3, 14 и 21 дана вратила модел $ARIMA(1,0,0)$ као најбољи док је за лагове од 7, 30 и 60 дана вратила редом моделе $ARIMA(2,0,4)$, $ARIMA(3,0,2)$, $ARIMA(2,1,3)$. Подела скупа у односу 90:10 је обухватила потпуни скок цене у тренинг скуп. Претрага *auto arima* функцијом је као најбоље моделе вратила редом $ARIMA(4,0,2)$, $ARIMA(1,0,3)$, $ARIMA(1,0,0)$, $ARIMA(1,0,1)$, $ARIMA(2,0,0)$, $ARIMA(1,0,5)$ за тестиране лагове.

5.6 Експеримент 3

У трећем експерименту је тестирано предвиђање цена све 3 криптовалуте више дана унапред коришћењем АРИМА модела. Стратегија за избор параметара је била претрага комбинација параметара описана у 4.5 уз то да је овде примењено предвиђање дан унапред а затим додавање предвиђене вредности у тест скуп уместо стварне. Временске серије криптовалута су подељене за тренинг и тестирање на исти начин као у експерименту 1. Опсежи за претрагу хиперпараметара су такође исти као у експерименту 1.

Шиба Ину тест скуп обухвата податке од 6.1.2022. до 4.2.2022. што је 29 дана. Претрага хиперпараметара је показала да је најефикаснији модел $ARIMA(8,1,7)$ са процентуалном грешком од 18.35% и Акаике информационом критеријумом од 173.

Идена тест скуп је обухватио податке од 25.8.2021 до 4.2.2022. што је 163 дана. Најбоља комбинација хиперпараметара је била $ARIMA(4,0,4)$ са процентуалном грешком предвиђања од 14.63% и Акаике информационам критеријумом у вредности од 1415.

Доцкоин тест скуп је обухватао податке од 2.9.2022 до 4.2.2022. што је 155 дана. Претрага је као најбољу комбинацију хиперпараметара вратила $ARIMA(5,1,6)$ са процентуалном грешком од 40.55% и Акаике информационам критеријумом од 3399.

5.7 Експеримент 4

У четвртном експерименту је тестирано предвиђање цена све три криптовалуте дан унапред употребом АРИМАКС модела. За потребе овог експеримента коришћени су скупови података настали пре анализе корелације, а након попуњавања недостајућих података. Као предиктивне варијабле су коришћене средње вредности и стандардне девијације променљивих рачунате за периоде претходних 3 и 7 дана јер се у претходним експериментима утврдило да кратки периоди дају најбоље резултате. Избор хиперпараметара модела је вршен помоћу *auto arima* функције.

ШИБА Ину скуп података је подељен у односу 90:10 како би се у потпуности ухватио пораст цене. Претрага хиперпараметара је показала да је за периоде рачунања променљивих од 3 и 7 дана најбољи модел $ARIMA(1,0,0)$.

Идена скуп података је подељен у односу 70:30 јер Идена нема велике ценовне шокове. Најбољи модели добијени претрагом су $ARIMA(1,0,0)$ и $ARIMA(2,0,1)$ за периоде рачунања од 3 и 7 дана.

Доцкоин скуп је подељен у односу 90:10 за тренинг и тест како би тренинг скуп обухватио ценовни скок из фебруара и априла 2021. Претрага је као најбоље вратила моделе $ARIMA(1,0,0)$ и $ARIMA(2,0,3)$.

Након идентификације параметара модела извршено је предвиђање за дан унапред и израчуната процентуална грешка.

5.8 Евалуација модела

Модели ће бити евалуирани рачунањем грешке у апсолутном проценту (енг. *MAPE*) која је дата формулом 8 и Акаике информационом критеријумом (енг. *AIC*) датим формулом 9. Циљ је да грешка предвиђања буде што мања, идеално испод 3% како би се могло рећи да потенцијални инвеститори могу бити сигурни да им губици неће бити велики. Код Акаике информационог критеријума мања вредност индикује да је у питању бољи модел, међутим пошто ова метрика даје релативне резултате у односу на тестиране моделе мора се комбиновати са још неком метриком како би се заиста евалуирала тачност модела. Сходно формули, модел са више варијабли имаће већу вредност критеријума и тиме бити лошије рангиран јер се сматра да има тенденцију да постане претрениран.

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (8)$$

где је:

n – Број момената за које се рачуна грешка.

A_t – Стварна вредност у моменту t .

F_t – Предвиђена вредност у моменту t

$$AIC = 2K - 2 \ln L \quad (9)$$

где је:

K – Број независних променљивих.

L – Вероватноћа да модел предвиди стварне вредности из скупа.

6. РЕЗУЛТАТИ И ДИСКУСИЈА

У овом поглављу биће представљени резултати експеримената спроведених у овом раду, њихово поређење као и закључци изведени из њих. У наредна 4 потпоглавља су дати резултати сва 4 експеримента редом којим су експерименти и описани у поглављу 5.

6.1 Резултати експеримента 1

Анализа графика аутокорељације и парцијалне аутокорељације није дала значајне резултате у одређивању хиперпараметара модела за Шиба Ину и Доцкоин јер ниједан график није показао податке на основу којих би се могло са неком сигурношћу рећи који је опсег параметара. Анализа истих графика за случај Идена криптовалуте показала је јасне индикаторе да се ради о чисто ауторегресивном моделу као што је описано у хеуристици рада [16]. Претрага комбинација хиперпараметара се, иако много спорија, показала као доста поузданија метода у поређењу са анализом графика аутокорељације и парцијалне аутокорељације. Међутим ако се упореди тачност модела, и модел добијен естимацијом параметара преко графика као и онај добијен претрагом имају врло сличну грешку с тим да је модел добијен претрагом параметара за нијансу тачнији (табела 6.1.1). Такође, претрага параметара је показала да је најбољи модел суштински само диференцирана верзија модела добијеног анализом графика, што говори да је претпоставка о предиференцирању временске серије била погрешна.

Начин естимације параметара	p	q	d	AIC	MAPE
Графици	1	0	0	1350	3.74%
Претрага комбинација параметара	1	1	0	1323	3.63%

табела 6.1.1 Поређење модела за предвиђање Идене.

Поређењем АРИМА модела за предвиђање дан унапред добијених претрагом комбинација (табела 6.1.2) уочава се да је тачност модела за предвиђање дан унапред тренираних овом меотодологијом врло слична.

Криптовалута	p	q	d	AIC	MAPE
Идена	1	1	0	1323	3.63%
Доцкоин	5	1	2	3350	3.34%
Шиба Ину	3	1	0	152	3.59%

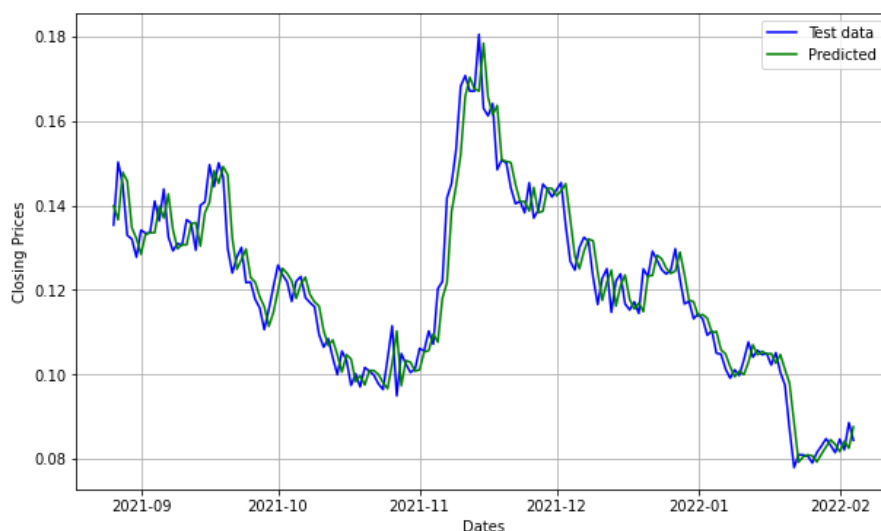
табела 6.1.2 Поређење модела за предвиђање криптовалута дан унапред

Разлика у вредности Акаике информационог критеријума код модела је последица различитог броја узорака над којима је трениран и тестиран модел. Доцкоин скуп је највећи, око три пута већи од друга два скупа те је логично да му је Акаике информациони критеријум приближно три пута већи. Врло мала вредност ове метрике за Шиба Ину је вероватно последица тога што је приликом тестирања стационарности из временске серије Шиба Ину уклоњено приближно трећина података (сви подаци до априла 2021) јер је цена у тим данима била 0.

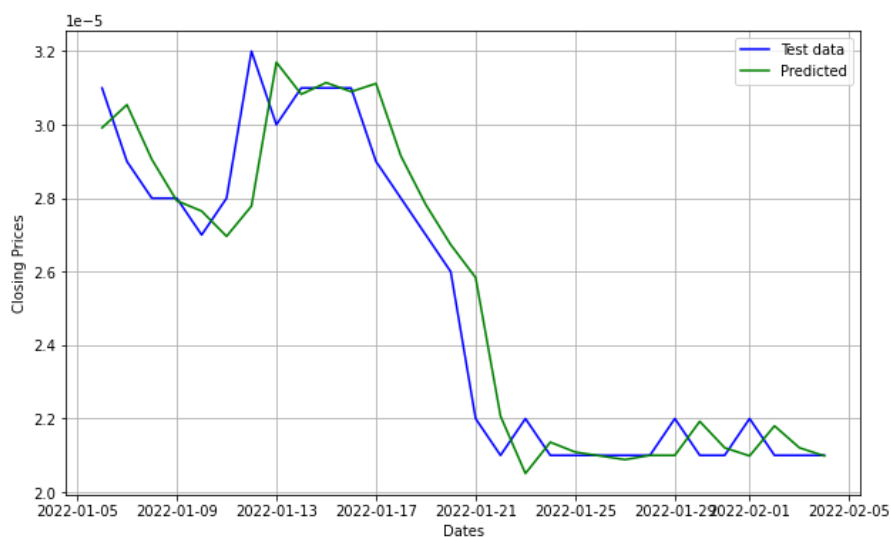
Увидом у графике предвиђања над тест скуповима за све 3 криптовалуте (слике 6.1.1, 6.1.2, 6.1.3) примећује се да је предвиђање врло приближно стварним вредностима. Међутим изгледа да је график предвиђања (зелена боја) померен удесно у односу на график стварних вредности, то јест да касни за њима. Ово би се могло тумачити као тенденција АРИМА модела да на основу претходних вредности предвиди наредну само пратећи њен тренд без могућности да заиста научи процесе који доводе до промена.



слика 6.1.1. График предвиђања цене Доцкоина дан унапред.



слика 6.1.2. Предвиђање цене Идене дан унапред.



слика 6.1.3. Предвиђање цене Шиба Ину дан унапред.

Приликом претраге комбинација хиперпараметара примећено је да подела посла по језгрима није у потпуности паралелизовала цео процес. Тестирање хиперпараметара реда 7 и већег је трајало знатно дуже од остатка, а пошто су комбинације вишег реда додељене језгрима 5 и 6 они су у овом случају били уска грла.

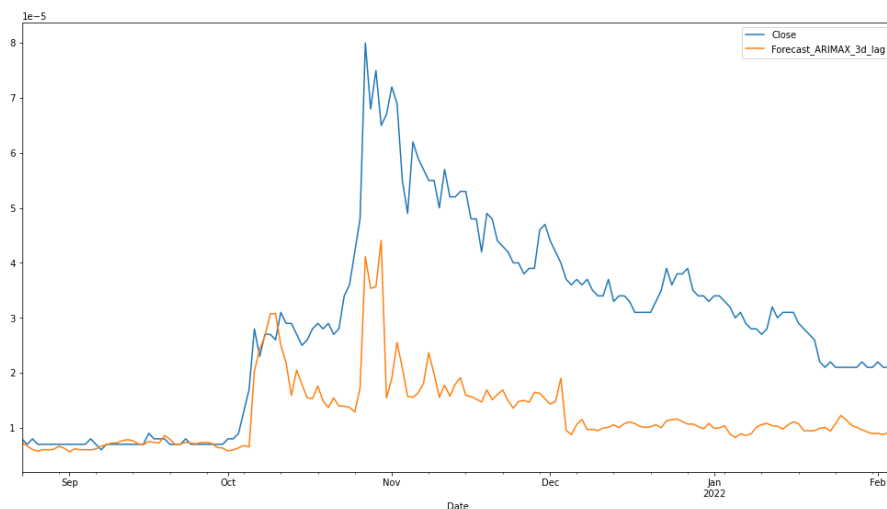
Резултати тестирања модела за краткорочно предвиђање потврђују закључке Амина Азарија из рада [12] о малој грешци при краткорочним предвиђањима цене криптовалута. Грешка при предвиђању која се креће око 3.5% је дупло већа од оне из рада [17] у ком се предвиђају вредности из макроекономских серија где је око 1.7%. У случају Идена криптовалуте показало се да анализа графика према хеуристици из рада [16] даје скоро исте резултате као и претрага комбинација хиперпараметара.

6.2 Резултати експеримента 2

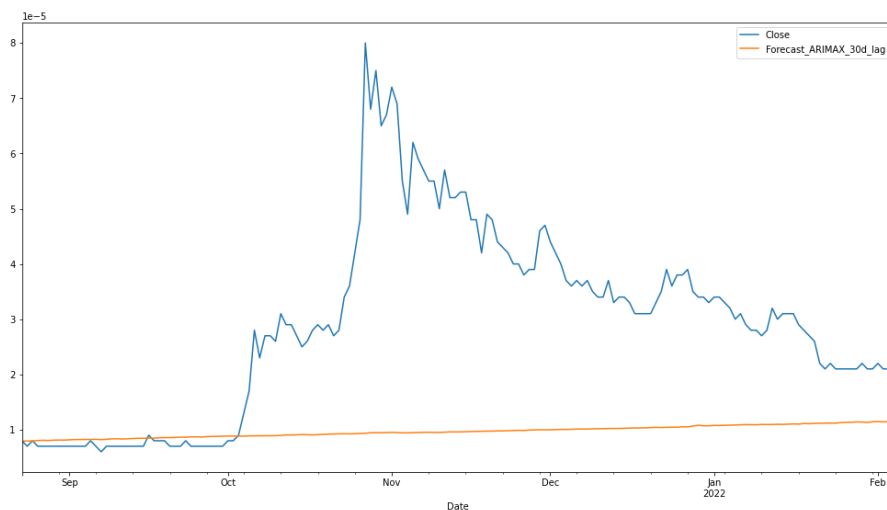
Подела Шиба Ину скупа у односу 70-30 и предвиђање са вредностима променљивих од пре 3, 7, 14, 21, 30 и 60 дана помоћу АРИМАКС модела дала је резултате приказане у табели 6.2.1. За вредности променљивих за прва 4 тестирана лага *auto arima* је као најбоље пронашла чисто ауторегресивне моделе, док је за 2 највећа лага добијен модел насумичне шетње. Велике грешке указују на непрецизност АРИМАКС модела при подели скупа у односу 70-30 чак и за вредности променљивих од пре 3 дана. Да би се боље разумело зашто је грешка толика потребно је погледати график предвиђања за лаг од 3 дана (слика 6.2.1) који се показао најтачнијим од свих. Са графика се види да је АРИМАКС модел успео да ухвати генерални тренд кретања цене али није предвидео тачну амплитуду скокова, вероватно зато што се нигде у тренинг скупу није ни срео са овако великим скоковима цене. Иако Шиба Ину има један велики скок цене у мају 2021, све до октобра цена је приближна нули и има мале осцилације па модел успева да је предвиди довољно тачно. Међутим не успева да предвиди амплитуду наглог скока у средини октобра, већ предвиђа амплитуду сличну оној из маја 2021. Модели насумичне шетње, као онај за лаг од 30 дана, дају предвиђање налик на линеарну регресију (слика 6.2.2) и не очекује се да ухвате тренд кретања цене.

Подела 70%-30% Шиба Ину					
Период лага	p	q	d	AIC	MAPE
3 дана	1	0	0	-8166	47.29%
7 дана	1	0	0	-8170	64.90%
14 дана	1	0	0	-8174	83.14%
21 дан	1	0	0	-8176	89.99%
30 дана	0	1	0	-8138	55.47%
60 дана	0	1	0	-8142	55.74%

табела 6.2.1. Поређење модела Шиба Ину за 70-30 поделу скупа.



слика 6.2.1. Предвиђање цене Шиба Ину са варијаблима од пре 3 дана.



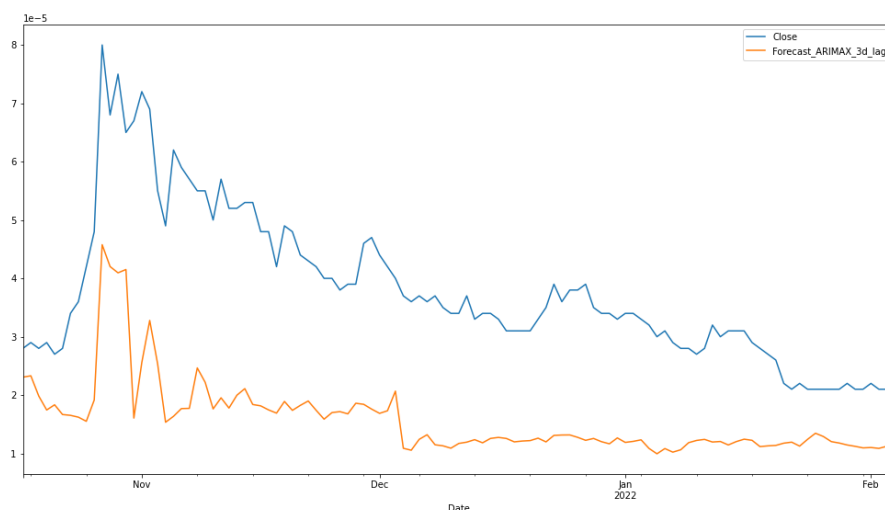
слика 6.2.2. Предвиђање цене Шиба Ину са варијаблима од пре 30 дана.

Подела Шиба Ину у односу 80:20 дала је резултате предвиђања и моделе приказане у табели 6.2.2. Грешке су поново велике из разлога истог као код поделе 70-30. У овом случају су чак још и веће јер је подела 80-20 обухватила тренинг скупом део података из септембра и новембра када је цена валуте скоро без осцилација, а где је модел трениран над 70% података вршио најтачнија предвиђања. Модел трениран над 80% података и даље у тренинг скупу нема цео скок цене

из периода октобар-ноovemбар те прави исту врсту грешке у процени амплитуде скока (слика 6.2.3) као и модел трениран над 70% података.

Подела 80%-20% Шиба Ину					
Период лага	p	q	d	AIC	MAPE
3 дана	1	0	0	-9328	57.41%
7 дана	1	0	0	-9326	62.44%
14 дана	1	0	0	-9335	88.14%
21 дан	1	0	0	-9335	102.43%
30 дана	1	0	0	-9337	93.66%
60 дана	0	1	0	-9305	28.61%

табела 6.2.2. Поређење модела Шиба Ину за 80-20 поделу скупа.



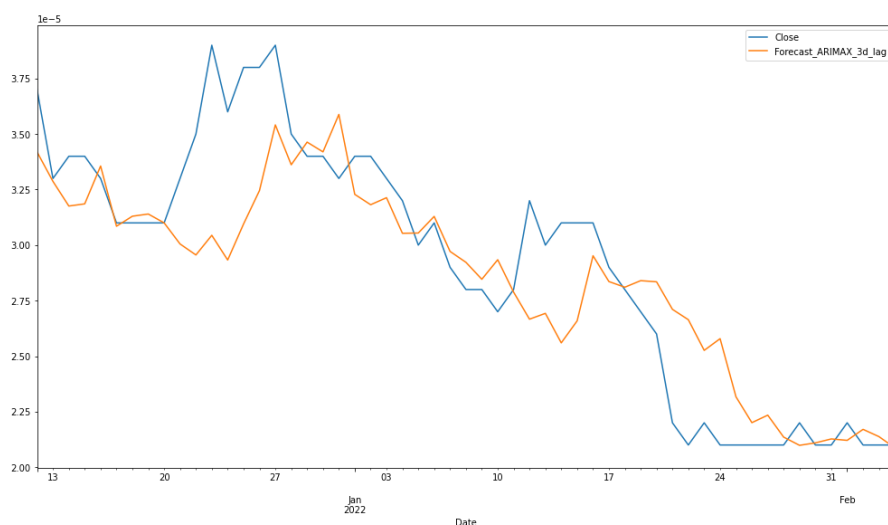
слика 6.2.3. Предвиђање цене Шиба Ину са варијаблима од пре 3 дана и тренинг скупом од 80%.

Подела скупа Шиба Ину у односу 90% тренинг 10% тест дала је резултате приказане у табели 6.2.3. За чисто ауторегресивне моделе је при овој подели значајно побољшана тачност, док је за моделе насумичне шетње остала слична јер се ту и даље ради о предвиђању налик на линеарну регресију. Пошто при обухвату тренинг скупа од 90% података у тест скупу не постоје значајне осцилације у цени модел даје прилично тачна предвиђања, чија тачност опада са повећањем броја дана од пре којих се узимају вредности варијабли. График предвиђања је померен као и у случају експеримента 1 (слика

6.2.4), само овај пут за онолико дана за од пре колико се узимала вредност предиктивних варијабли.

Подела 90%-10% Шиба Ину					
Период лага	p	q	d	AIC	MAPE
3 дана	1	0	0	-10497	7.15%
7 дана	1	0	0	-10463	14.89%
14 дана	1	0	0	-10489	19.23%
21 дан	1	0	0	-10490	17.15%
30 дана	0	1	0	-8780	29.97%
60 дана	0	1	0	-8782	28.90%

табела 6.2.3. Поређење модела Шиба Ину за 90-10 поделу скупа.



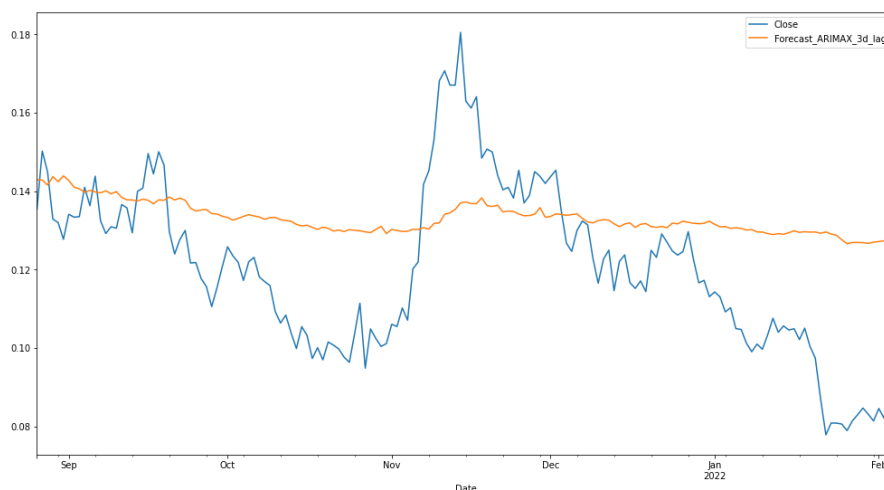
слика 6.2.4. Предвиђање цене Шиба Ину са варијаблама од пре 3 дана и тренинг скупом од 90%.

Резултати предвиђања за Идену над поделом скупа у односима 70-30, 80-20 и 90-10 су дати у табели 6.2.4 али само за варијабле од пре 3 и 7 дана које су тестиране. Иако је грешка од 17.72% при предвиђању са варијаблама од пре 3 дана много боља од оне коју прави Шиба Ину модел за исту поделу скупа не може се рећи да је овај модел бољи. Закључак да није бољи може се извести тек ако се погледа график предвиђања за променљиве од пре 3 дана и поделу 70% (слика 6.2.5) где се види да модел не успева да ухвати тренд кретања цене већ само повлачи хоризонтални лук са мањим

осцилацијама кроз центар графика. Графици предвиђања за остале моделе изгледају слично и такође не успевају да ухвате тренд кретања цене већ дају предвиђање налик на линерану регресију. Овакви резултати се могу потенцијално објаснити тиме да су у фази анализе корелација избачене многе променљиве и тиме је АРИМАКС модел осиромашен. Такође, популарност Идене у заједници нема тако вртоглаве скокове као што има Шиба Ину и Доцкоин што би такође могло утицати на то да варијабле везане за заједницу погоршавају тачност модела уколико су остале у скупу.

Период лага	p	q	d	AIC	MAPE
Подела 70%-30% Идена					
3 дана	1	0	0	-2265	17.72%
7 дана	2	0	1	-2258	19.93%
Подела 80%-20% Идена					
3 дана	1	0	0	-2631	17.37%
7 дана	2	0	1	-2624	18.84%
Подела 90%-10% Идена					
3 дана	1	0	2	-2994	23.50%
7 дана	2	0	1	-2994	19.47%

табела 6.2.4. Поређење модела Идене за све односе подела скупа.



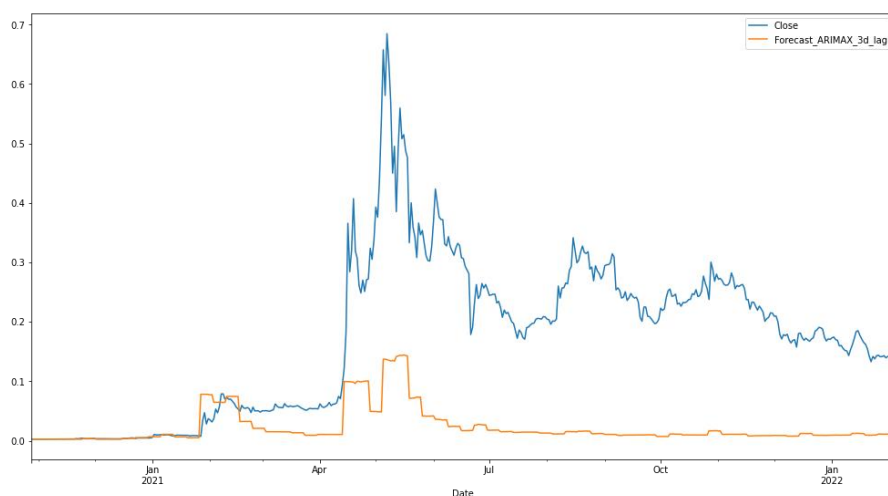
слика 6.2.5. Предвиђање цене Идене са лагом од 3 дана и поделом 70-30.

Дељењем Доцкоин скупа у односу 70-30 за све тестиране лагове добијају се АРИМАКС модели из табеле 6.2.5. Грешка је веома

висока као и код Шиба Ину, с тим да је овде још и већа. Разлог за овако велику грешку је веоватно исти као и у случају Шиба Ину, с тим да Доцкоин има само један велики скок цене који овде није обухваћен тренинг скупом тако да амплитуда пораста цене која је овде предвиђена (слика 6.2.6) није ни близу стварној. Доцкоин тренинг скуп при подели од 70% обухвата искључиво период када је његова цена била испод 0.1 долара док је скок који се нашао у тест скупу ишао до 0.7 долара.

Подела 70%-30% Доцкоин					
Период лага	p	q	d	AIC	MAPE
3 дана	2	0	0	-13606	74.63%
7 дана	3	0	0	-13611	79.49%
14 дана	0	1	1	-14424	85.83%
21 дан	0	1	1	-14427	84.10%
30 дана	1	1	2	-14419	85.38%
60 дана	1	1	0	-14462	82.56%

табела 6.2.5. Поређење модела Доцкоина за однос поделе скупа 70-30.



слика 6.2.6. Предвиђање цене Доцкоина са лагом од 3 дана и поделом 70-30.

Проширењем тренинг скупа на 80% добијени су модели приказани у табели 6.2.6. Поново су се као најбољи показали модели који користе најскорије променљиве. Проширењем тренинг скупа резултати за предвиђање на основу варијабли од пре 3 и 7 дана су се значајно побољшали, док је побољшање за остале периоде доста мање.

Увидом у график предвиђања за варијабле од пре 3 дана (слика 6.2.7) уочава се да модел поново не може у потпуности да предвиди амплитуду скока, али овај пут предвиђање прилази много ближе. Иако је тренинг скуп поделом 80% ухватио први део скока цене из јуна 2021 модел је овде много прецизнији него што је модел Шиба Ину био у сличној ситуацији. И у овом случају график предвиђања је померен удесно и касни за стварним вредностима цена.

Подела 80%-20% Доцкоин					
Период лага	p	q	d	AIC	MAPE
3 дана	1	0	0	-11402	20.72%
7 дана	2	0	4	-11462	36.83%
14 дана	1	0	0	-11707	61.35%
21 дан	1	0	0	-12085	79.37%
30 дана	3	0	2	-12149	78.67%
60 дана	2	1	3	-12587	72.38%

табела 6.2.6. Поређење модела Доцкоина за однос поделе скупа 80-20.



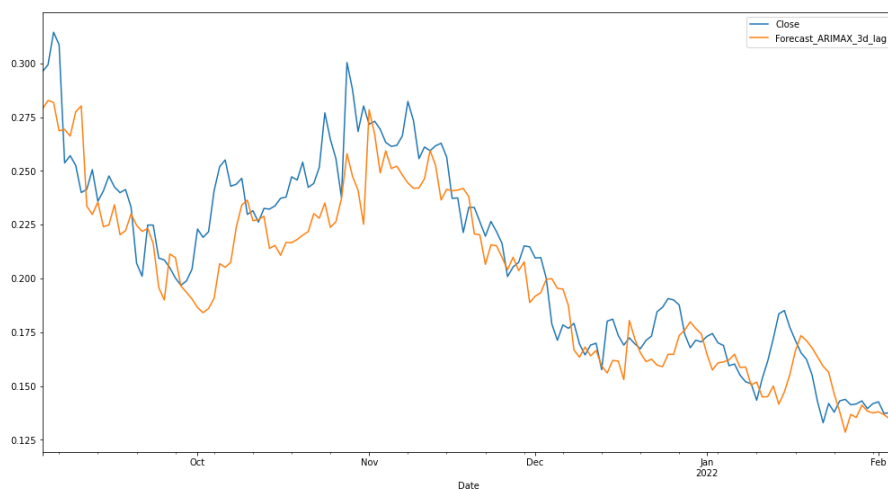
слика 6.2.7. Предвиђање цене Доцкоина са лагом од 3 дана и поделом 80-20.

Обухват тренинг скупа од 90% даје моделе чија је метрика у табели 6.2.7. Проширење тренинг скупа на 90% узорка довело је до много већег побољшања за предвиђања на основу варијабле са лагом од 3 и 7 дана, док је тачност за остале лагове негде боља негде лошија. Поређењем ових резултата са резултатима код поделе 70-30 и 80-20 види се да повећање тренинг скупа конзистентно побољшава тачност

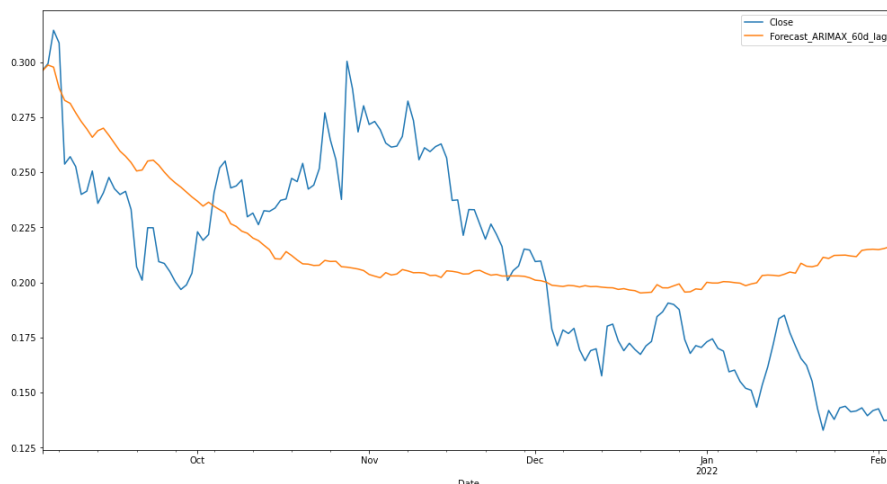
код предвиђања над варијаблама од пре 3 (слика 6.2.8) и 7 дана док за остале лагове даје мешовите резултате. Значајно побошање које се у овом случају види од предвиђања са лаговима 30 и 60 дана није последица доброг модела, већ предвиђања које је слично оном код Идене где се повлачи лук кроз центар графика са малим осцилацијама (слика 6.2.9).

Подела 90%-10% Доцкоин					
Период лага	p	q	d	AIC	MAPE
3 дана	4	0	2	-7955	7.12%
7 дана	1	0	3	-7492	8.75%
14 дана	1	0	0	-8000	56.82%
21 дан	1	0	1	-7957	100.18%
30 дана	2	0	0	-7780	33.50%
60 дана	1	0	5	-8217	18.33%

табела 6.2.7. Поређење модела Доцкоина за тренинг скуп од 90%.



слика 6.2.8. Предвиђање цене Доцкоина са лагом од 3 дана и поделом 90-10.



слика 6.2.9. Предвиђање цене Доцкоина са лагом од 60 дана и поделом 90-10.

У овом експерименту најбоље су се показали АРИМАКС модели који користе варијабле од пре 3 и 7 дана, док остали дају неконзистентне резултате и предвиђања у која се не може поуздати. Као најбољи однос поделе скупа за тренирање и тестирање показао се онај који у тренинг скупу обухвата све велике скокове и падове. У случају Шиба Ину и Доцкоина то је однос 90-10, док су у случају Идене сви модели лоши, али најмања грешка се прави поделом 80-20. Поређењем модела са варијаблама од пре 3 и 7 дана за најбоље односе поделе скупа за све три криптовалуте (табела 6.2.8) најбоље се показао модел Доцкоина са грешкама од 7.12% и 8.75%.

АРИМАКС модели за краткорочно предвиђање						
Валута	Период лага	p	q	d	AIC	MARE
Доцкоин	3 дана	4	0	2	-7955	7.12%
Доцкоин	7 дана	1	0	3	-7492	8.75%
Идена	3 дана	1	0	0	-2631	17.37%
Идена	7 дана	2	0	1	-2624	18.84%
Шиба Ину	3 дана	1	0	0	-10497	7.15%
Шиба Ину	7 дана	1	0	0	-10463	14.89%

слика 6.2.8. Поређење АРИМАКС модела за варијабле од пре 3 и 7 дана.

Резултати код предвиђања цене Доцкоина и Шиба Ину не изненађују јер указују на способност АРИМАКС модела да предвиди искључиво пар дана унапред. Такође потврђују закључак рада [12] о

немогућности да предвиди нагле скокове у вредностима. Међутим модели Идене су показали да нису од користи за било каква предвиђања.

6.3 Резултати експеримента 3

При динамичком предвиђању цене више дана унапред показало се да АРИМА модел нема велики потенцијал. Модели за различите криптовалите праве и врло различите грешке у предвиђању. Поређење модела са метрикама евалуације је у табели 6.3.1. Увидом у метрику евалуације може се закључити да је најбоља генерализација и најтачније предвиђање било за Идена модел. Резултати су још убедљивији ако се има у виду да је период предвиђања за Идену био и највећи од свих а то је 163 дана унапред. Доцкоин је за приближан период предвиђања од 155 дана унапред имао грешку скоро 3 пута већу, а Шиба Ину модел је за 5 пута краћи период имао већу грешку од 18.35%. Акаике информациони критеријум је за ове моделе нешто већи него за моделе за предвиђање дан унапред, али је међусобни однос метрика модела за валуте скоро исти као и у експерименту 1.

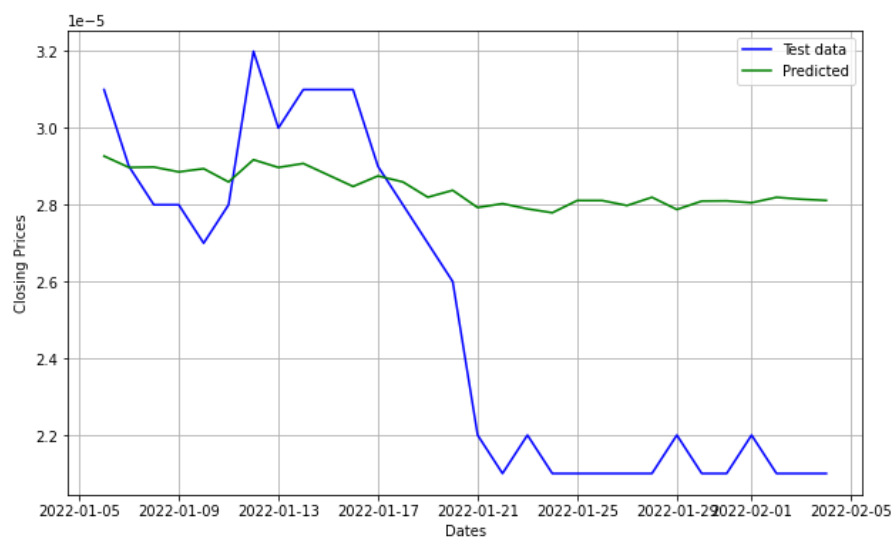
Криптовалута	Период	p	q	d	AIC	MAPE
Идена	163 дана	4	0	4	1415	14.63%
Доцкоин	155 дана	5	1	6	3399	40.55%
Шиба Ину	29 дана	8	1	7	173	18.35%

табела 6.3.1. Поређење АРИМА модела за динамичко предвиђање цена више дана унапред.

Само гледајући метрику модела могло би се закључити да је Идена модел прилично добар имајући у виду период предвиђања од 163 дана. Међутим увидом у графике предвиђања за све три криптовалите (слике 6.3.1, 6.3.2, 6.3.3) долази се до потпуно супротног закључка.



слика 6.3.1. Динамичко предвиђање цене Доџкоина АРИМА моделом.



слика 6.3.2. Динамичко предвиђање цене Шиба Ину АРИМА моделом.



слика 6.3.3. Динамичко предвиђање цене Идене АРИМА моделом.

Предвиђања на сва три графика су скоро па равна линија уз минималне осцилације које се једва виде јер су превише мале у односу на стварну осцилацију цене. Овакав метод предвиђања унапред се понаша скоро као линеарна регресија и не може да ухвати стварне ценовне осцилације већ само неки генерални тренд кретања. Мала грешка од 14.35% код предвиђања цене Идене је таква вероватно зато што је њена цена имала пораст у средини периода предвиђања што је допринело томе да линија предвиђања прође кроз центар графика. Цена Доцкоина је у периоду тестирања у поређењу са Иденем имала већи пад, у смислу да ни у једном моменту није била већа од цене на почетку периода. Због тога је у другом делу периода тестирања прављена веома велика грешка која је на крају повећала просек грешке у односу на Идену.

У поређењу са резултатима рада [20] који се такође бави дугорочним предвиђањима ово су слични резултати. Процентуалне грешке су сличне и крећу се од 14% до 41% док у наведеном раду иду од 16% до 38%.

6.4 Резултати експеримента 4

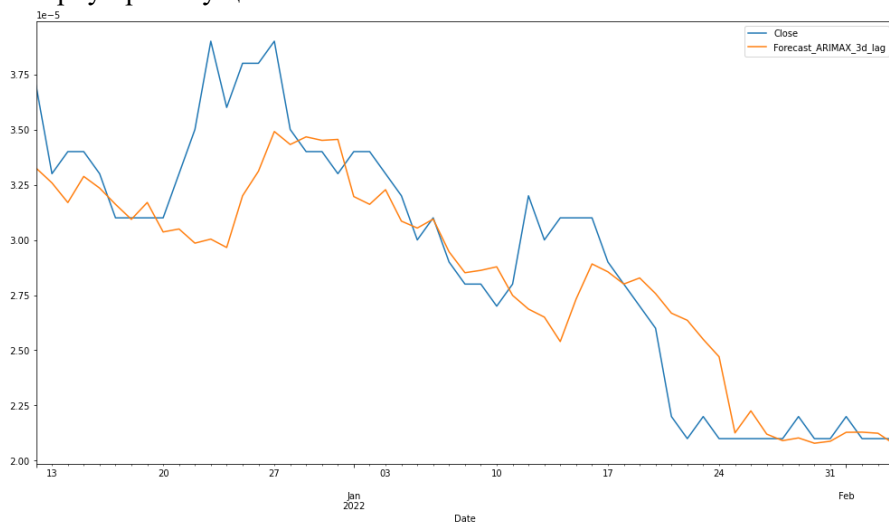
АРИМАКС модели су се при предвиђању дан унапред користећи средње вредности варијабли и њихову стандарду девијацију за 3 и 7 дана показали добро (табела 6.4.1). Увођење ових варијабли је

побољшало тачност модела, али у поређењу са методологијом коришћеном у експерименту 1 грешка је око 2 пута већа. Најбољи модел у овом случају је био онај за Доцкоин са варијаблама рачунатим за период од претходна 3 дана где је грешка 6.21%.

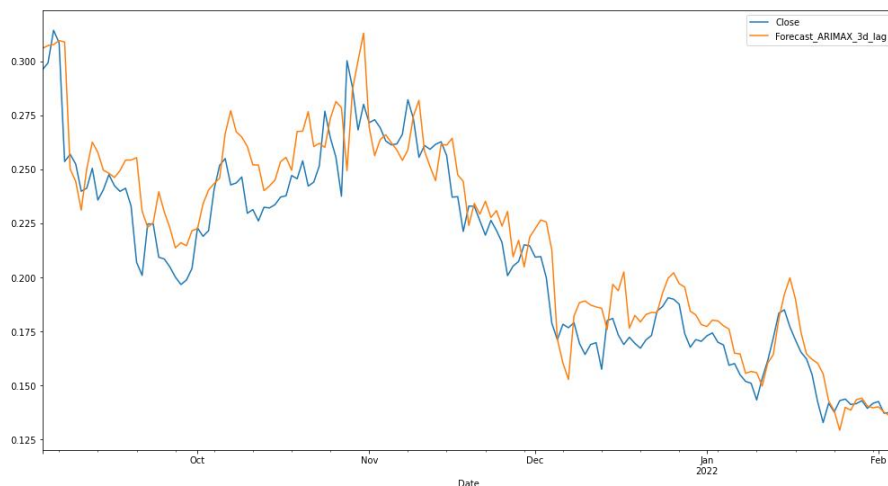
АРИМАКС модели за предвиђање дан унапред						
Валута	Период лага	p	q	d	AIC	MAPE
Доцкоин	3 дана	1	0	0	-8418	6.21%
Доцкоин	7 дана	2	0	3	-8351	12.73%
Идена	3 дана	1	0	0	-2132	7.46%
Идена	7 дана	2	0	1	-2119	12.38%
Шиба Ину	3 дана	1	0	0	-10493	6.46%
Шиба Ину	7 дана	1	0	0	-10465	10.50%

табела 6.4.1. Поређење АРИМАКС модела за предвиђање дан унапред.

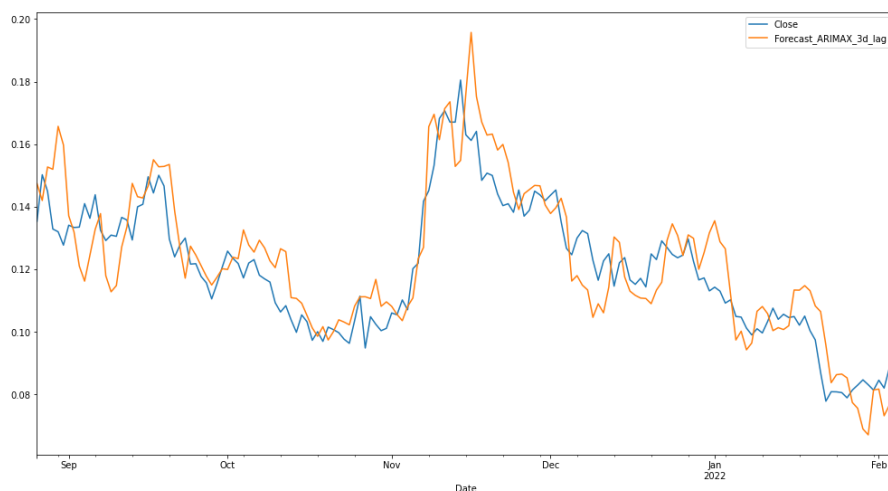
Поређењем графика предвиђања за моделе са варијаблама рачунатим над периодом од 3 дана (слике 6.4.1, 6.4.2, 6.4.3) види се да је и овде као и у претходним експериментима график померен удесно, с тим што је овде тај померај мањи и чини се да график боље осликава стварну промену цене.



слика 6.4.1. Предвиђање цене Шиба Ину дан унапред.



слика 6.4.2. Предвиђање цене Доцкоина дан унапред.



слика 6.4.3. Предвиђање цене Идене дан унапред.

У поређењу са радом [17] грешка коју прави АРИМАКС модел овде је скоро 2 пута већа и износи око 6% док је у наведеном раду око 3.7%. Мањи померај графика удесно у случају предвиђања коришћењем ове методологије је вероватно последица тога што се не користе вредности од пре 3 дана за предвиђање наредне, већ просек вредности и стандардна девијација за претходна 3 дана што доводи до ублажавања ефекта понављања.

7. ЗАКЉУЧАК

У овом раду представљено је више метода за краткорочно и дугорочно предвиђање цена 3 криптовалуте: Идена, Доцкоин и Шиб Ину помоћу АРИМА и АРИМАКС модела. Главна мотивација за ово истраживање била је нагли пораст броја нових инвеститора у свету криптовалута који се десио од почетка пандемије Коронавируса и њихови велики инвестициони губици. За оба модела тестирано је краткорочно и дугорочно предвиђање, више метода за избор хиперпараметара АРИМА и АРИМАКС модела, најбољи однос поделе скупова података на тренинг и тест скуп и утицај додатних варијабли на тачност модела.

Скупови података су прикупљени са Коингека, Јаху финансија и Гугл трендова, затим су подаци агрегирани у јединствене скупове за сваку валуту и претпроцесирани. Након претпроцесирања за временске серије је тестирано да ли су стационарне помоћу комбинације АДФ и КПСС тестова, а уколико је утврђено да нису извршено је логаритмовање и диференцирање. Хиперпараметри модела су бирани на 2 начина: помоћу *auto arima* функције која то ради аутоматски и тестирањем различитих комбинација хиперпараметара у језгрима процесора и бирањем најбоље. Предвиђање цена помоћу модела је вршено 1 и више дана унапред а грешке упоређиване како би се одабрали најбољи интервали и модели. Евалуација модела приликом избора хиперпараметара је вршена Акаике информационам критеријумом а приликом предвиђања грешком у апсолутном проценту. У процесу евалуације тестирано је више односа поделе скупа података на тренинг и тест скупове.

Методологија предвиђања дан унапред помоћу АРИМА модела коришћена у експерименту 1 показала се као супериорна у односу на предвиђање АРИМАКС моделом дан унапред и коришћењем средње вредности варијабли и стандардне девијације за период од претходних 3 дана. Грешка коју овом приликом прави АРИМА модел је око 3.5% док АРИМАКС прави грешку од око 6%. У предвиђањима више дана унапред АРИМАКС модел са грешком од око 7% је био бољи од АРИМА модела са грешком од око 14% који је правио предвиђања налик на линеарну регресију.

Анализа графика аутокорејације и парцијалне аутокорејације се показала као непоуздана код избора хиперпараметара АРИМА модела у односу на претрагу комбинација. Претрага комбинација са друге стране траје изузетно дуго, за моделе вишег реда и по неколико

сати. Паралелизовање процеса претраге на више језгара је доста убрзало процес. Међутим пошто је скуп комбинација за претрагу подељен узимајући комбинације редом из скупа декартовог производа опсега јавило се уско грло код тестирња модела вишег реда који су сви завршили у последња 2 језгра. Ауто арима функција ради много брже од имплементиране претраге, али није познато да ли је једнако поуздана.

При подели скупова на тренинг и тест подскупове потребно је обратити пажњу да велики скокови цене морају бити садржани у тренинг скупу. АРИМА и АРИМАКС модели су показали да могу да предвиде скокове цена, међутим не и њихов интензитет и амплитуду чак и ако у тренинг скупу имају сличан скок али мањег интензитета. Приликом анализе утицаја променљивих АРИМАКС модела на предвиђање потребно је узети у обзир њихову вредност од највише 7 дана раније јер су већи периоди показали слабију корелацију, а вредности променљивих из тих периода дале моделе слабије тачности.

Уочено је да оба модела имају тенденцију да приликом предвиђања чак и дан унапред дају графике предвиђања благо померене удесно у односу на график стварне цене. Ово би могло бити последица тога да АРИМА модел не може стварно да предвиди ценовне осцилације већ само да покуша предвиђање на основу тренда кретања од пара дана уназад.

Као један од могућих праваца надоградње предлаже се упоређивање АРИМА модела и линеарне регресије за предвиђање више дана унапред имајући у виду да график предвиђања АРИМА модела у том случају доста подсећа да линеарну регресију.

Може се проверити и како се модели понашају ако се прикупљени подаци на дневном нивоу пребаце на недељни, месечни и годишњи ниво рачунањем просека. Било би интересантно утврдити да ли је могуће тачније дугорочно предвиђање на овај начин, барем у смислу предвиђања тренда кретања цене.

Како би се отклонило уско грло у тестирању комбинација хиперпараметара комбинације се могу боље поделити. Модели вишег реда се могу равномерно распоредити по језгрима тако да оптерећеност буде једнака. Додатно, могу се проширити и опсези претраге хиперпараметара како би се утврдило да ли модели много већег реда показују боље резултате и евентуално проверити да ли доводе до претренираности модела њиховом евалуацијом над различитим подинтервалима.

Додатно, пошто график АРИМАКС модела при предвиђању дан унапред коришћењем средње вредности и стандардне девијације варијабли поакзује мању помереност удесно могла би се тестирати методологија коришћена у експерименту 1 како би се АРИМА и АРИМАКС модели равноправно тестирали и утврдило да ли АРИМАКС може додатно да буде тачнији.

8. ЛИТЕРАТУРА

- [1] Farrell, Ryan. "An analysis of the cryptocurrency industry." (2015).
- [2] Chohan, Usman W. "A history of Dogecoin." Discussion Series: Notes on the 21st Century (2021).
- [3] <https://shibatoken.com/> Шиба Ину валута (приступљено априла 2022.)
- [4] <https://docs.idena.io/docs/wp/summary/> Идена валута (приступљено априла 2022.)
- [5] Sakamoto, Yosiyuki, Makio Ishiguro, and Genshiro Kitagawa. "Akaike information criterion statistics." Dordrecht, The Netherlands: D. Reidel 81.10.5555 (1986): 26853
- [6] Box, George E. P, and Gwilym M. Jenkins. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1976. Print.
- [7] Schöneburg, Eberhard. "Stock price prediction using neural networks: A project report." *Neurocomputing* 2.1 (1990): 17-27.
- [8] KOHARA, K., ISHIKAWA, T., FUKUHARA, Y. and NAKAMURA, Y. (1997), Stock Price Prediction Using Prior Knowledge and Neural Networks. *Intelligent Systems in Accounting, Finance and Management*, 6: 11-22. [https://doi.org/10.1002/\(SICI\)1099-1174\(199703\)6:1<11::AID-ISAF115>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1099-1174(199703)6:1<11::AID-ISAF115>3.0.CO;2-3)
- [9] J. Rebane, I. Karlsson, P. Papapetrou, and S. Denic, ‘Seq2Seq RNNs and ARIMA models for Cryptocurrency Prediction : A Comparative Study’, in *Proceedings of SIGKDD Workshop on Fintech (SIGKDD Fintech’18)*, 2018.

- [10] Wang, Yu, and Runyu Chen. "Cryptocurrency price prediction based on multiple market sentiment." Proceedings of the 53rd Hawaii International Conference on System Sciences. 2020.
- [11] Toriola, Adebayo J. *Prediction of Bitcoin Prices Using Deep learning and Sentiment Analysis Based on Bitcoin Tweets*. Diss. Dublin, National College of Ireland, 2021.
- [12] Azari, Amin. "Bitcoin price prediction: An ARIMA approach." arXiv preprint arXiv:1904.05315 (2019).
- [13] Nakamoto, Satoshi, and A. Bitcoin. "A peer-to-peer electronic cash system." *Bitcoin*.—URL: <https://bitcoin.org/bitcoin.pdf> 4 (2008).
- [14] Gagniuc, Paul A. (2017). *Markov Chains: From Theory to Implementation and Experimentation*. USA, NJ: John Wiley & Sons. pp. 1–256. ISBN 978-1-119-38755-8.
- [15] Scalzotto, Giovanni. "Social Media Impact on Cryptocurrencies." (2021).
- [16] 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.
- [17] Peter, Ďurka, 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.
- [18] Ayodele Ariyo Adebisi, Aderemi Oluyinka Adewumi, Charles Korede Ayo, "Comparison of ARIMA and Artificial Neural Networks Models for Stock Price Prediction", *Journal of Applied Mathematics*, vol. 2014, Article ID 614342, 7 pages, 2014. <https://doi.org/10.1155/2014/614342>
- [19] Iqbal, Mahir, et al. "Time-series prediction of cryptocurrency market using machine learning techniques." EAI Endorsed Transactions on Creative Technologies (2021): e4.

- [20] Devi, B. Uma, D. Sundar, and P. Alli. "An effective time series analysis for stock trend prediction using ARIMA model for nifty midcap-50." *International Journal of Data Mining & Knowledge Management Process* 3.1 (2013): 65.
- [21] Ariyo, Adebisi A., Adewumi O. Adewumi, and Charles K. Ayo. "Stock price prediction using the ARIMA model." *2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation*. IEEE, 2014.
- [22] <https://people.duke.edu/~rnau/411arim.htm> Арима модел (приступљено априла 2022.)
- [23] Kongcharoen, Chaleampong, and Tapanee Kruangpradit. "Autoregressive integrated moving average with explanatory variable (ARIMAX) model for Thailand export." *33rd International Symposium on Forecasting, South Korea*. 2013.
- [24] <https://www.bauer.uh.edu/rsusmel/phd/ec2-5.pdf> Стационарност временске серије (приступљено априла 2022.)
- [25] Dickey, David A., and Wayne A. Fuller. "Distribution of the estimators for autoregressive time series with a unit root." *Journal of the American statistical association* 74.366a (1979): 427-431.
- [26] Dickey, David A., and Wayne A. Fuller. "Likelihood ratio statistics for autoregressive time series with a unit root." *Econometrica: journal of the Econometric Society* (1981): 1057-1072.
- [27] <https://www.statsmodels.org> Statsmodels библиотека у Пајтону (приступљено априла 2022.)
- [28] Kwiatkowski, D.; Phillips, P. C. B.; Schmidt, P.; Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics*, 54 (1-3): 159-178.
- [29] <https://finance.yahoo.com/> Јаху финансије (приступљено априла 2022.)

[30] <https://www.coingecko.com/> Коингеко (приступљено априла 2022.)

[31] <https://trends.google.com> Гугл трендови (приступљено априла 2022.)

[32] <https://alkaline-ml.com/pmdarima/> *Pmdarima* библиотека (приступљено априла 2022.)

[33] <https://www.facebook.com/> Фејсбук друштвена мрежа (приступљено априла 2022.)

[34] <https://www.reddit.com/> Редит друштвена мрежа (приступљено априла 2022.)

[35] <https://github.com/> Гитхаб платформа (приступљено априла 2022.)

[36] <https://www.youtube.com/> Јутјуб платформа (приступљено априла 2022.)

9. БИОГРАФИЈА

Предраг Главаш је рођен 3.5.1998. у Београду. Основну школу је завршио у Сремској Каменици 2013. године, а средњу електротехничку школу „Михајло Пупин“ у Новом Саду 2017. године. Одмах након средње школе уписује Факултет техничких наука у Новом Саду школске 2017/2018. године. Диплому инжењера електротехнике и рачунарства стиче 9.9.2021. на смеру Примењено софтверско инжењерство. Године 2022. положио је све испите на мастер студијама смера рачунарство и аутоматика и стекао услов за одбрану мастер завршног рада.

KLJUČNA DOKUMENTACIJSKA INFORMACIJA

Redni broj, RBR :	
Identifikacioni broj, IBR :	
Tip dokumentacije, TD :	monografska publikacija
Tip zapisa, TZ :	tekstualni štampani dokument
Vrsta rada, VR :	master rad
Autor, AU :	Predrag Glavaš
Mentor, MN :	dr Aleksandar Kovačević, redovni profesor
Naslov rada, NR :	Predviđanje cena kriptovaluta upotrebom ARIMA i ARIMAKS modela
Jezik publikacije, JP :	srpski
Jezik izvoda, JL :	srpski / engleski
Zemlja publikovanja, ZP :	Srbija
Uže geografsko područje, UGP :	Vojvodina
Godina, GO :	2022
Izdavač, IZ :	autorski reprint
Mesto i adresa, MA :	Novi Sad, Fakultet tehničkih nauka, Trg Dositeja Obradovića 6
Fizički opis rada, FO :	9 / 71 / 36 / 18 / 30 / 0 / 0
Naučna oblast, NO :	Računarstvo i automatika
Naučna disciplina, ND :	Mašinsko učenje
Predmetna odrednica / ključne reči, PO :	Kriptovalute, ARIMA, ARIMAKS, Dogecoin, Shiba Inu, Idena
UDK	
Čuva se, ČU :	Biblioteka Fakulteta tehničkih nauka, Trg Dositeja Obradovića 6, Novi Sad
Važna napomena, VN :	
Izvod, IZ :	U radu je predstavljeno predviđanje cena kriptovaluta jedan i više dana unapred upotrebom ARIMA i ARIMAKS modela. Testirane su različite metode izbora hiperparametara, podela na trening i test skupove i različiti periodi predviđanja. Na kraju su prikazani rezultati sa diskusijom i predstavljeni potencijalni pravci unapređenja.
Datum prihvatanja teme, DP :	27.6.2022.
Datum odbrane, DO :	4.7.2022.
Članovi komisije, KO :	
predsednik	dr Jelena Slivka, vanredni profesor
član	dr Lidija Krstanović, docent
mentor	dr Aleksandar Kovačević, redovni profesor
Potpis mentora	

KEY WORDS DOCUMENTATION

Accession number, ANO :	
Identification number, INO :	
Document type, DT :	monographic publication
Type of record, TR :	textual material
Contents code, CC :	master thesis
Author, AU :	Predrag Glavaš
Mentor, MN :	Aleksandar Kovačević full professor, PhD
Title, TI :	Cryptocurrency price prediction using ARIMA and ARIMAX models
Language of text, LT :	Serbian
Language of abstract, LA :	Serbian / English
Country of publication, CP :	Serbia
Locality of publication, LP :	Vojvodina
Publication year, PY :	2022
Publisher, PB :	author's reprint
Publication place, PP :	Novi Sad, Faculty of Technical Sciences, Trg Dositeja Obradovića 6
Physical description, PD :	9 / 71 / 36 / 18 / 30 / 0 / 0
Scientific field, SF :	Computing and control
Scientific discipline, SD :	Machine learning
Subject / Keywords, S/KW :	Cryptocurrencies, ARIMA, ARIMAX, Shiba Inu, Dogecoin, Idena
UDC	
Holding data, HD :	Library of the Faculty of Technical Sciences, Trg Dositeja Obradovića 6, Novi Sad
Note, N :	
Abstract, AB :	This thesis presents cryptocurrency price prediction one or several days ahead using ARIMA and ARIMAX models. Different methods were tested for hyperparameter search and train/test split. Various prediction periods were also tested. In the end results were presented with discussion and suggestions for improvement.
Accepted by sci. Board on, ASB :	27.6.2022.
Defended on, DE :	4.7.2022.
Defense board, DB :	
president	Jelena Slivka, associate professor, PhD
member	Lidija Krstanović, assistant professor, PhD
mentor	Aleksandar Kovačević, full professor, PhD
Mentor's signature	

