

НАСТАВНО-НАУЧНОМ ВЕЋУ
ЕКОНОМСКОГ ФАКУЛТЕТА
УНИВЕРЗИТЕТА У БЕОГРАДУ

На основу одлуке Наставно-научног већа бр. 1456/1 од 10.04.2024. именовани смо у Комисију за оцену докторске дисертације под називом „Рударење текста – Утицај јавно доступних текстова на приносе крипто валута“ кандидата Александра Дамјановића. Пошто смо проучили завршену докторску дисертацију, подносимо Наставно-научном већу следећи

ИЗВЕШТАЈ О ОЦЕНИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

1. Основни подаци о кандидату и дисертацији

Кандидат Александар Дамјановић рођен је у Приштини 1996. године. Завршио је основне студије на Економском факултету Универзитета у Београду, на студијском програму Статистика, информатика и квантитативне финансије (модул Статистика), као и мастер студије на Међународном мастер програму из квантитативних финансија (IMQF). Био је запослен као демонстратор на Економском факултету, а тренутно је ангажован као асистент на Рачунарском факултету Универзитета Унион и на Београдској банкарској академији.

Као студент докторских студија Економског факултета Универзитета у Београду, у оквиру студијског програма за статистику, положио је следеће испите: Методологија научног истраживања 1-Д, Методе и технике научног истраживања и анализе, Моделирање и оптимизација, Статистика – економска статистика, Демографија, Економетрија 1-Д, Економетрија панела, Анализа временских серија 1-Д и Мултиваријациона анализа 1-Д.

Кандидат је до сада објавио један научни рад:

- **Дамјановић, А.** и Дреновак, М. (2023). Да ли су све текстуалне вести само шум за инвеститоре? – Утицај онлајн текстова на приносе Биткоина. *Економске теме*, 61(2): 121-144. DOI: 10.2478/ethemes-2023-0007

Рад је објављен у часопису који се у тренутку подношења овог извештаја налази у категорији M51.

Наставно-научно веће Економског факултета одобрило је кандидату израду докторске дисертације под називом „Рударење текста – Утицај јавно доступних текстова на приносе крипто валута“ на седници одржаној 8. фебруара 2023. године. Докторска дисертација Александра Дамјановића написана је на 162 стране, што укључује списак литературе на осам страна, као и прилоге и биографију на осам страна. Поред увода и закључка, дисертација садржи четири централна поглавља: Крипто-валуте и

информатичке иновације у финансијској технологији које су их створиле (стр. 23–51), Методологија (стр. 53–95), Модификација оцена Џагадиша и Вуа (стр. 97–110) и Анализа добијених резултата (стр. 111–141). Дисертација има 19 слика и 28 табела, који пружају детаљније информације о резултатима и закључцима изнетих у тексту, као и њихову одговарајућу илустрацију. У списку литературе наведена су 134 релевантна извора, од чега се 125 извора односи на библиографске јединице које обухватају чланке из међународних научних часописа, радне папире и књиге из научних области које су предмет дисертације у ужем или ширем смислу, пет интернет извора и четири техничка документа.

2. Предмет и циљ дисертације

Предмет ове докторске дисертације је проучавање утицаја алтернативних извора података на понашање цена криптовалута. Кандидат користи текстове о криптовалутама објављене на популарним порталима које посећују крипто-мајнери и крипто-трејдери као основни овакав извор. Истраживање се ослања на савремене технике машинског учења како би прикупио текстове са релевантних портала који ће ући у процес „рударења текста“. Резултати тог процеса се потом користе за анализу утицаја јавно доступних информација на принос криптовалута. У дисертацији је разрађен нови приступ оцењивању пондера сентимента, са циљем да унапреди модел који су развили Џагадиш и Ву (2019).

Примарни циљ дисертације је да испита могућност употребе рударења текстова за предвиђање приноса на криптовалуте. Реч је о актуелном питању, будући да су криптовалуте, као једна од најмлађих врста финансијске активе на свету, још увек недовољно истражене. Са друге стране, истраживање утицаја текстова на цене утолико је интересантније због децентрализованости криптовалута. Секундарни циљ дисертације је развој модела за оцену пондера сентимента који би представљао побољшање у односу на постојећу литературу. У дисертацији је приказан теоријски оквир новог модела који је затим практично имплементиран.

3. Основне хипотезе од којих се полазило у истраживању

У складу са дефинисаним предметом и циљем истраживања, у овом докторском истраживању постављена су три основна истраживачка питања и тестиране су три повезане хипотезе.

Прво истраживачко питање постављено у дисертацији је да ли информације које долазе из вести са интернет портала могу утицати на инвестиционе одлуке крипто-трговаца и одлуке о рударењу криптовалута. Позитиван одговор на ово питање би значео да информације из вести директно утичу на понуду и тражњу за криптовалутама, а самим тим и на њихове цене и стопе приноса. Да би се добио одговор на ово истраживачко питање, тестирана је следећа хипотеза:

Хипотеза 1: Постоји веза између сентимента вести о криптовалутама објављених на популарним интернет порталима и приноса на криптовалуте.

Постојање овакве везе има своје економске последице. Према хипотези ефикасности тржишта у полу-јаком облику, тржишна цена одражава све јавно доступне информације. Уколико је она валидна, онда нико не може да искористи јавно доступне информације да предвиди будуће кретање цена, односно приноса. С тим у вези, постојање статистички значајне везе између сентимента вести о криптовалутама и њихових приноса био би сигнал потенцијалног нарушења хипотезе ефикасности тржишта у полу-јаком облику.

Друго истраживачко питање у разматрање уводи две додатне променљиве добијене рударењем текста. Претпоставка о утицају читљивости изводи се из прве хипотезе: уколико је текст слабо читљив, његова порука читаоцу неће бити јасна, а самим тим неће ни имати утицај на њега. Да би се добио одговор на друго истраживачко питање, тестирана је следећа хипотеза:

Хипотеза 2: Постоји веза између приноса криптовалута и читљивости вести написаних о њима, као и веза између приноса криптовалута и сентимента вести о Биткоину.

Неодбацивање ове хипотезе сугерисало би постојање везе између кретања приноса и показатеља добијених рударењем текста. То би поново било у супротности са хипотезом ефикасности тржишта у полу-јаком облику.

Џагадиш и Ву (2019) су дефинисали модел за мерење нивоа, односно, магнитуде сентимента појединачних речи директно из података добијених рударењем текста. Ниво сентимента речи ови аутори су назвали пондерима сентимента. Оцене пондера које су предложили садржале би грешку мерења, те процена сентимента текста на бази њих не би била адекватна. Отклањањем или умањивањем грешке мерења добила би се потенцијално прецизнија процена сентимента. Дисертација поставља ову могућност као треће истраживачко питање, које се формализује следећом хипотезом:

Хипотеза 3: Грешке у прогнози приноса на бази сентимента текстова у вестима биће мање када се користи иновирана методологија оцењивања сентимента уместо оригиналне методологије Џагадиша и Вуа (2019).

Сентимент процењен на бази предложене методологије искоришћен је потом за предвиђање кретања приноса на крипто-валуте. Потенцијална предвидивост ових приноса била би, такође, у супротности са хипотезом ефикасности тржишта у полу-јаком облику.

4. Кратак опис садржаја дисертације

Докторска дисертација Александра Дамјановића структурирана је тако да након увода, који уједно поставља истраживачка питања (глава 1), даје опис криптовалута са техничког аспекта (глава 2), затим описује методологију истраживања (глава 3), потом (у глави 4) уводи модификацију приступа Џагадиша и Вуа (2019), презентује и анализира кључне резултате истраживања (глава 5) и даје закључне напомене (глава 6).

Прво поглавље носи наслов „Увод“. Описује предмет истраживања, да би потом тема била постављена у шири академски контекст, кроз преглед литературе. Затим се

постављају циљеви дисертације и истиче научни допринос у односу на препознати истраживачки јаз у литератури. Поглавље потом формализује три хипотезе и описује изворе података.

Друго поглавље носи наслов „Крипто-валуте и информатичке иновације у финансијској технологији које су их створиле“ и објашњава техничке и институционалне аспекте ове класе инструмената – њихов настанак, децентрализоване мреже, технологију блок-ланаца и „рударење“ криптовалута, као и њихове предности, недостатке и примену. Поглавље се завршава прегледом криптовалута које су коришћене у истраживању у дисертацији.

Треће поглавље, под називом „Методологија“, упознаје читаоца са дизајном истраживања и приказује коришћену методологију. Конкретно, описује софтверске алате посебно израђене за прикупљање и рударење текстова, након чега даје свеобухватан преглед метода коришћених за обрачун приноса и читљивости текстова. Потом следи представљање начина на који ће хипотезе бити тестиране, опис алгоритма за предвиђање приноса, опис коришћених тестова предвиђања и тестова хипотезе ефикасности тржишта у слабом облику. Поглавље такође приказује вишеузорачки тест квалитета предвиђања.

Четврто поглавље носи наслов „Модификација оцена Џагадиша и Вуа“. У њему се описује приступ мерењу сентимента. Приказује иновирану методологију оцењивања пондера сентимента коју предлаже ова дисертација. У оквиру истог поглавља су представљене и финалне оцене сентимента коришћене у емпиријском делу истраживања.

Пето поглавље је „Анализа добијених резултата“. У њему су најпре изложене оцене сентимента и читљивости вести. Затим су анализирани везе између показатеља добијених рударењем текста и приноса одабраних криптовалута. Потом се разматра успешност конструисаног ансамбл-модела у предвиђању приноса и анализирају алтернативни облици модела и резултати добијени на основу њих. На крају овог поглавља дати су и резултати тестова хипотезе ефикасности тржишта у слабом облику.

Најзначајнији закључци и теоријске и практичне импликације резултата из све четири централне главе резимирани су и представљени у шестом поглављу дисертације, које носи наслов „Закључак“.

5. Методе које су примењене у истраживању

У дисертацији је развијен алгоритам за „скрејповање“ и рударење текста. Циљ овог алгоритма је да оспособи машину (рачунар) да генерише вектор „концепата“, односно речи и фраза чији се садржај потом анализира. Тако добијене векторе потребно је мапирати у одговарајуће нумеричке векторе. У дисертацији је за ово мапирање коришћен $w2v$ (*word-to-vector*) приступ заснован на TF-IDF (*term frequency – inverse document frequency*) методу.

Овако добијени вектори користе се потом за учење (тј. тренинг) модела за оцену пондера сентимента. У дисертацији се за ту сврху пореде модели Џагадиша и Вуа (2019) и

модификована верзија овог модела развијена у оквиру истраживања Кандидата. Добијени резултати сировог рударења текста (вектори концепата) користе се и за праћење лексичке читљивости текста, јер се на бази њих рачунају вредности „индекса замагљености“, FI (енгл. *fog index*).

Од статистичких метода коришћена је кластер анализа К-средина, која је стандардна техника машинског учења и мултиваријационе анализе, као и обична линеарна регресија са пратећим тестовима. Кластер анализа користи се како би рачунар груписао текстове према њиховој тематици у сродне групе, тј. кластере. Ово је непоходно како би се испитао утицај текстова са одређеном темом на приносе криптовалута. Обична линеарна регресија служи за испитивање оправданости употребе потенцијалних текстуалних предиктора у финалном моделу. Оценом статистичке значајности коефицијената у регресионим једначинама испитује се да ли постоји веза између потенцијалних текстуалних предиктора и приноса криптовалута. Постојање статистички значајне везе дало би оправдање да се предиктор употреби за предвиђање приноса на основу сентимента.

Квалитет предвиђања приноса које даје алгоритам машинског учења анализира се у зависности од тога да ли су текстуални предиктори конструисани уз помоћ модела Џагадиша и Вуа (2019) или модификованог модела који се предлаже у дисертацији. Ова анализа спроведена је поређењем средњег квадратног одступања предвиђања приноса добијених за сваку од изабраних криптовалута, али и кроз Диболд-Маријанов тест, како би се постојање статистички значајне разлике између нивоа грешака и формално проверило.

6. Остварени резултати и научни допринос

Модификација модела Џагадиша и Вуа (2019) делимично отклања непрецизности у мерењу сентимента. Прецизност се може побољшати увођењем додатних претпоставки, које се математички задају као ограничења у проблему оптимизације. Дисертација је понудила једно такво решење, претпоставивши да се сентимент речи креће у стандардним границама које се користе у литератури, тј. у интервалу $[-1, 1]$.

Како би се испитала употребна моћ нових оцена, спроведена је троетапна процедура. Целокупан узорак је подељен у три подузорка, по један за сваку етапу. Први подузорак намењен је калибрацији пондера и највећи је од три подузорка. Његова величина обезбеђује да алгоритам има довољно информација за учење приликом оцене пондера сентимента. Након ове фазе, оцене пондера се користе за мерење сентимента било ког текста, на основу кога се препознају потенцијални предиктори приноса. Други подузорак намењен је идентификовању релевантних предиктора приноса. Ови предиктори се потом користе за тестирање способности предвиђања приноса у трећем подузорку. На тај начин предвиђања се врше над подацима ван узорка за анализу предиктора.

Резултати добијени на бази другог подузорка омогућили су проверу две од три постављене хипотезе. Прва хипотеза претпоставља значајан утицај сентимента вести о одређеној криптовалуту на њене приносе. Друга хипотеза састојала се од два дела. Први

део претпоставља значајан утицај читљивости објављених вести на приносе, док други део претпоставља да на приносе утиче и унакрсни сентимент. Резултати су разматрани посебно и за оцене добијене уз претпоставку о стандардним границама пондера (егзактне нормиране оцене) и за оцене са апроксимацијом остатка грешке (апроксимативне оцене).

У случају егзактних нормираних оцена прва хипотеза није одбачена за четири од осам анализираних криптовалута. Први део друге хипотезе је у највећој мери одбачен. То није био случај са другим делом ове хипотезе, који није одбачен у седам од осам случајева.

Примена апроксимативних оцена дала је сличне резултате. Промене су примећене у три од осам анализираних случајева. Овога пута прва хипотеза није одбачена у шест случајева. Утицај сентимента о вестима директно повезаним са посматраном криптовалutom сада је препознат и код ETH и код BTC. Промене су примећене и у знаку веза. У два случаја добијен је нелогичан предзнак, што кандидат приписује периоду за који је анализа спроведена. Закључци су остали непромењени и за оба дела друге хипотезе.

На основу претходно изложених резултата о обе анализиране врсте сентимента (сопствени сентимент и унакрсни сентимент), кандидат изводи закључак да је сентимент вести користан индикатор за анализу криптовалута. То нарочито важи за унакрсни сентимент, на који се ова дисертација посебно усредсређује. Резултати указују и на то да је анализа јако осетљива на избор мере сентимента, тј. методе оцене пондера. Приметно је да се закључци у појединим случајевима могу значајно разликовати уколико се на њих примене различите мере сентимента. Сходно томе, избор адекватне мере сентимента јесте веома важно питање у анализи садржаја текста, што је један од основних закључака ове дисертације.

Приказани резултати су сигнали потенцијалне неефикасности тржишта криптовалута. Резултати дисертације показују да је у анализираном периоду постојала веза између информација добијених анализом садржаја текста и приноса криптовалута. Поред тога ауторегресивна компонента је била значајна за пет од осам криптовалута. На основу ових налаза кандидат закључује да нису све јавно доступне информације укључене у цене криптовалута. Са жељом да дубље истражи ово питање, у дисертацији су спроведени ауторегресивни тестови хипотезе ефикасности тржишта у слабом облику. Ови тестови су показали да се цене криптовалута AVAX, BTC, DOGE и ETH могу описати ARMA процесом, док за криптовалуте ADA, DOT и LUNA цене имају јединични корен, али се њихови приноси не понашају као бели шум, већ се такође могу описати ARMA процесом. За криптовалюту SOL није било могуће одбацивати хипотезу ефикасности тржишта у слабом облику.

За формалну проверу треће хипотезе, чији је циљ био да испита да ли ће нове мере сентимента генерисати побољшане предикторе приноса, конструисан је ансамбл алгоритама који обрачунава предикторе по две методологије – оригиналној методологији Џагадиша и Вуа (2019) и модификацији предложеној у овој дисертацији. Алгоритам формира предикције приноса за наредни дан на бази свих доступних вести из текућег дана. Добијене грешке су најпре упоређене тачкастим оценама квадратног корена

средње квадратне грешке предвиђања. И егзактне и апроксимативне оцене имају мање грешке предвиђања од оцена Џагадиша и Вуа (2019). У циљу уопштавања закључака спроведена су два статистичка теста: омнибус Диболд-Маријанов тест и МекКракенов (2000) тест. Први тест, који је такође развијен у склопу дисертације, у оба случаја сугерише да модификоване оцене сентимента дају боља предвиђања приноса од оригиналних оцена по методологији Џагадиша и Вуа (2019). Насупрот томе, МекКракенов (2000) тест даје статистички значајну разлику само код апроксимативних оцена. Иако статистичка значајност није потврђена код егзактних нормираних оцена, тачкасте мере квалитета прогнозе су мање у свим анализираним случајевима. Кандидат закључује да су се апроксимативне оцене показале бољим од егзактних нормираних оцена, будући да су оба статистичка теста препознала њихову значајност.

Ова дисертација има емпиријски и методолошки допринос.

Са емпиријске стране, допринос се састоји у утврђивању природе утицаја онлајн текстова на кретање приноса криптовалута. Испитан је утицај показатеља добијених на основу информација садржаних у текстовима на приносе криптовалута. Ту се посебно издваја испитивање утицаја унакрсног сентимента, који је још увек недовољно истражен у постојећој литератури, а показао се као потенцијално значајан предиктор приноса криптовалута. Постојање мерљивог утицаја доводи до одбацивања хипотезе ефикасности тржишта у слабом облику код појединих криптовалута из узорка.

Са методолошког аспекта, дисертација је понудила алтернативан начин оцене пондера сентимента, чиме су остварена одређена побољшања преовлађујућих приступа у литератури. За иновираних оцена сентимента речи се испоставило да помажу у предвиђању будућег кретања приноса. У дисертацији је развијен и сопствени лексикон потребан за „рударење“ текста, чиме су унапређени јавно доступни алати који се користе у пракси. Коначно, у дисертацији је представљена и модификована верзија Диболд-Маријановог теста коју је кандидат развио у склопу свог истраживања. Овај „омнибус“ Диболд-Маријанов тест погодан је за вишеузорачко тестирање оцена способности предвиђања коришћених модела.

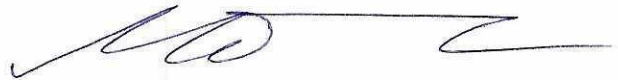
7. Закључак

Након анализе докторске дисертације Александра Дамјановића под називом „Рударење текста – Утицај јавно доступних текстова на приносе крипто валута“, Комисија оцењује да је кандидат, користећи релевантну научну методологију и опсежну литературу, успешно обрадио предложену тему. Докторска дисертација је урађена у складу са важећим стандардима и одобреном пријавом од стране Наставног-научног већа Економског факултета Универзитета у Београду и Већа научних области правно-економских наука Универзитета у Београду, и у погледу предмета, циља и метода истраживања, и у погледу садржаја. Комисија оцењује да је остварен циљ истраживања, те да је докторска дисертација резултат оригиналног и самосталног научног рада кандидата. Сматрамо да резултати дисертације могу дати допринос у научној области финансијске економије.

На основу свега изложеног, предлажемо Наставно-научном већу Економског факултета Универзитета у Београду да прихвати позитивну оцену докторске дисертације Александра Дамјановића „Рударење текста – Утицај јавно доступних текстова на приносе крипто валута“ и одобри њену јавну одбрану.

У Београду, 20.05.2024. године

Чланови Комисије:



проф. др Милош Божовић
Универзитет у Београду, Економски факултет



др Драгана Радојичић, доцент
Универзитет у Београду, Економски факултет



проф. др Микица Дреновак
Универзитет у Крагујевцу, Економски факултет