

**К. Ю. Кононова**

кандидат економічних наук, доцент

Харківський національний університет ім. В.Н. Каразіна  
kateryna.kononova@gmail.com

**А. О. Дек**

аспірант

Харківський національний університет ім. В.Н. Каразіна  
dektox@gmail.com

**В. В. Марков**

старший викладач

Харківський національний університет ім. В.Н. Каразіна  
Markov.vadim@gmail.com

**М. О. Шпакович**

студент

Харківський національний університет радіоелектроніки  
Maksym.shpakovych@gmail.com

## ПРОГНОЗУВАННЯ ЕКОНОМІЧНИХ РЯДІВ НА ОСНОВІ АНАЛІЗУ НАСТРОЇВ КОРИСТУВАЧІВ ІНТЕРНЕТ

У роботі запропоновано комплекс моделей прогнозування економічних часових рядів з урахуванням об'єктивного та суб'єктивного контенту мережі Інтернет. Для аналізу обрана система Bitcoin (BTC) – перша і найбільш популярна на сьогодні криптовалюта. Експоненціальне зростання ринку криптовалют в останні роки обумовлює актуальність прогнозування курсу BTC. Теоретичну основу дослідження склали положення концепції поведінкових фінансів, в рамках якої передбачається, що поведінка трейдерів ірраціональна, а характер їхніх рішень в істотній мірі залежить від психологічних факторів. Метою дослідження є розробка методології та комплексу моделей прогнозування валютного курсу на основі аналізу фактуального (об'єктивного) та концептуального (суб'єктивного) контенту інтернет. Джерелом фактуальної інформації обрано тематичні статті спеціалізованих новинних порталів, джерелом концептуальної – записи користувачів у мікроблогах. Основу модельного комплексу склали: 1) скрипти парсингу новинних порталів та мікроблогів; 2) алгоритм формування фактуальних та концептуальних екзогенних змінних на основі латентно-семантичного аналізу, аналізу тональності текстів, оцінки казуальності за Гренджером; 3) математичним інструментарієм прогнозування обрано нейронні мережі прямого розповсюдження сигналу, а також рекурентні мережі з тривалою та короткостроковою пам'яттю (LSTM), множина входів яких формувалася з використанням генетичних алгоритмів. В результаті обробки бази даних новинних стрічок та твітів було сформовано множину екзогенних факторів, до якої увійшли чотири з чотирнадцяти фактуальних змінних – *infrastructure*, *activity*, *dissemination* і *expect*, а також дві з восьми концептуальних – *calm* і *confusion*. Автоматизація пошуку оптимальної архітектури нейронних мереж виконувалася з використанням генетичних алгоритмів: довжина хромосоми дорівнювала числу змінних; особини піддавалися схрещуванню, мутації та відбору на основі якості передбачення моделі. Порівняльний аналіз різних архітектур нейронних мереж дозволив обґрунтувати доцільність використання інтернет-контенту для прогнозування економічних рядів та продемонстрував високу адекватність побудованих моделей.

**Ключові слова:** прогнозування курсу Bitcoin, записи у мікроблогах, стрічки фінансових новин, латентно-семантичний аналіз, аналіз тональності текстів, нейронні мережі.

**JEL Classification:** C88, E44, E47, G12

**К. Ю. Kononova**

Ph.D. (Economics), Associate Professor  
V. N. Karazin Kharkiv National University  
kateryna.kononova@gmail.com

**А. О. Dek**

Ph.D. Student  
V. N. Karazin Kharkiv National University  
dektox@gmail.com

**V. V. Markov**

Senior Lecturer

V. N. Karazin Kharkiv National University

Markov.vadim@gmail.com

**M. O. Shpakovych**

Student

Kharkiv National University of Radioelectronics

Maksym.shpakovych@gmail.com

## **ECONOMIC SERIES PREDICTION BASING ON INTERNET USERS SENTIMENT ANALYSIS**

A set of economic time series forecasting models (based on objective and subjective internet content analysis) is proposed in the article. The Bitcoin (BTC) system, the first and most popular cryptocurrency today, was chosen for the analysis. Exponential increase of the cryptocurrency market stipulates relevance of BTC rate forecasting in recent years. Theoretical framework of the research is based on the behavioral finance concept supposing that traders' behavior is irrational, and the character of their decisions largely depends on psychological factors. The aim of the research is to develop currency rate forecasting methodology and a set of models based on the analysis of factual (objective) and conceptual (subjective) internet content. The source of factual information is relevant newsfeed of specialized news portals, the source of conceptual information is users' records in microblogs. The basis of the models' set includes: 1) parsing scripts of news portals and microblogs; 2) algorithm of factual and conceptual exogenous variables generation on the basis of latent-semantic analysis, sentiment analysis, Granger causality analysis; 3) chosen mathematic forecasting tools such as feedforward neural networks and recurrent networks with long short-term memory (LSTM) the set of inputs of which was formed by applying genetic algorithms. As a result of news feeds and tweets database processing, the set of exogenous factors including four out of fourteen factual variables (infrastructure, activity, dissemination and expect) and two out of eight conceptual ones (calm and confusion) was worked out. Automation of neural networks architecture optimization was conducted with the use of genetic algorithms: chromosome length equaled the number of variables; species were subject to hybridization, mutation and selection based on their fit function, which is MSE for the validation dataset. Comparative analysis of different neural networks architectures allowed proving the expediency of Internet content application for economic time series forecasting and demonstrated high appropriateness of the developed models.

Keywords: bitcoin price forecasting, microblog posts, financial news feeds, latent semantic analysis, opinion mining, neural networks.

**JEL Classification:** C88, E44, E47, G12.

**Е. Ю. Кононова**

кандидат экономических наук, доцент

Харьковский национальный университет им. В.Н. Каразина

kateryna.kononova@gmail.com

**А. О. Дек**

аспирант

Харьковский национальный университет им. В.Н. Каразина

dektox@gmail.com

**В. В. Марков**

старший преподаватель

Харьковский национальный университет им. В.Н. Каразина

Markov.vadim@gmail.com

**М. А. Шпакович**

студент

Харьковский национальный университет радиозлектроники

Maksym.shpakovych@gmail.com

## **ПРОГНОЗИРОВАНИЕ ЭКОНОМИЧЕСКИХ РЯДОВ НА ОСНОВЕ АНАЛИЗА НАСТРОЕНИЙ ПОЛЬЗОВАТЕЛЕЙ ИНТЕРНЕТ**

В работе предложен комплекс моделей прогнозирования экономических временных рядов с учетом объективного и субъективного контента в сети Интернет. Для анализа выбрана система Bitcoin (BTC) – первая и наиболее популярная на сегодняшний день криптовалюта. Экспоненциальный рост рынка криптовалют в последние годы обуславливает актуальность прогнозирования курса BTC. Теоретическую

основу исследования составили положения концепции поведенческих финансов, в рамках которой предполагается, что поведение трейдеров иррационально, а характер их решений в значительной мере зависит от психологических факторов. Целью исследования является разработка методологии и комплекса моделей прогнозирования валютного курса на основе анализа фактуального (объективного) контента и концептуального (субъективного) контента сети Интернет. Источником фактуальной информации были выбраны тематические статьи специализированных новостных порталов, источником концептуальной – записи пользователей в микроблогах. Основу модельного комплекса составили: 1) скрипты парсинга новостных порталов и микроблогов; 2) алгоритм формирования фактуальных и концептуальных экзогенных переменных на основе латентно-семантического анализа, анализа тональности текстов, оценки причинности по Грэнджеру; 3) математическим инструментарием прогнозирования выбраны нейронные сети прямого распространения сигнала, а также рекуррентные сети с длительной и краткосрочной памятью (LSTM), множество входов которых формировалось с использованием генетических алгоритмов. В результате обработки базы данных новостных лент и твитов было сформировано множество экзогенных факторов, в которое вошли четыре из четырнадцати фактуальных переменных – *infrastructure*, *activity*, *dissemination* и *expect*, а также две из восьми концептуальных – *calm* и *confusion*. Автоматизация поиска оптимальной архитектуры нейронных сетей выполнялась с использованием генетических алгоритмов: длина хромосомы равнялась числу переменных; особи подвергались скрещиванию, мутации и отбору на основе качества предсказания модели. Сравнительный анализ разных архитектур нейронных сетей позволил обосновать целесообразность использования интернет-контента для прогнозирования экономических рядов и продемонстрировал высокую адекватность построенных моделей.

**Ключевые слова:** прогнозирование курса Bitcoin, записи в микроблогах, ленты финансовых новостей, латентно-семантический анализ, анализ тональности текстов, нейронные сети.

**JEL Classification:** C88, E44, E47, G12

**Постановка проблеми.** З розвитком інформаційної економіки зростає число досліджень властивих їй феноменів, одним з яких є ринок криптовалют<sup>1</sup>. Для аналізу нами була обрана система Bitcoin (BTC) – перша та найбільш популярна на сьогодні криптовалюта, історія якої розпочалася у 2009 р., в травні 2010 р. стався перший обмін BTC на реальний товар, в серпні того ж року почалися торги по парі BTC / USD . Експоненціальне зростання ринку в останні роки обумовлює актуальність прогнозування курсу BTC.

**Аналіз останніх досліджень і публікацій.** Теоретичну основу дослідження склали положення концепції поведінкових фінансів (Blockchain), в рамках якої передбачається, що поведінка інвесторів і трейдерів ірраціональна, а характер інвестиційних рішень в істотній мірі залежить від психологічних факторів<sup>2</sup>, найбільш важливими з яких для цього дослідження є ефекти інформаційного каскаду<sup>3</sup>, консерватизму<sup>4</sup> та зверхреакції<sup>5</sup>.

Розвиток соціальних мереж і микроблогів (ще одного феномену інформаційної економіки) наштовхнув дослідників на ідею використання інтернет-контенту для прогнозування споживчих очікувань. Проблемам прогнозування фінансових показників на основі аналізу думок користувачів присвячені роботи (Zheludev, Smith, & Aste, 2014), (Zhang, Fuehres, & Gloor, 2011), (Lachanski, 2015), (Mao, Wei, & Wang, 2013), (Preis, Moat, & Stanley, 2013), (Ruiz, Hristidis, Castillo, Gionis, & Jaimes, 2013), (Challet, D., Ayed, 2014) та інших.

Зокрема, в статті (“Метод ієрархічної агломеративної кластеризації”, 2016) доводиться можливість оцінки настроїв користувачів за записами в микроблогах (в основному Twitter) для побудови моделі часового ряду індексу Dow Jones Industrial Average. Однак отримані Болленом та ін. багатообіцяючі результати були небезпідставно розкритиковані в (Lachanski, 2015) та (Sloot, 2012) і не підтвердилися на іншому часовому проміжку (Sharma, Vyas, 2010).

Можливість прогнозування біржових індексів за частотами появи в твіті<sup>6</sup> слів, що виражають емоції, аналізувалася в роботах (Ruiz, Hristidis, Castillo, Gionis, Jaimes, 2013) та (Zheludev, Smith, Aste, 2014). Згідно з їхніми дослідженнями, найбільш значущими виявилися слова-маркери надії та страху. У роботі Руйца розраховувалися кореляції між цінами акцій,

<sup>1</sup> Криптовалюта – цифровий актив з децентралізованою системою емісії та обліку, що функціонує в рамках розподіленої комп'ютерної мережі.

<sup>2</sup> Дослідження в цій області в 2002 р. відзначені Нобелівською премією (Сміт і Канеман).

<sup>3</sup> Схильність економічних суб'єктів до впливу сторонніх думок.

<sup>4</sup> Уповільнена зміна суб'єктами своїх переконань під впливом нової інформації.

<sup>5</sup> Гостра реакція на нову інформацію, незалежно від того, погана вона чи хороша.

<sup>6</sup> Твіт – запис у микроблозі Twitter

об'ємом торгів і активністю в сервісі Twitter. Для аналізу використовувалися графи «автор – твіт – посилання», які будувалися для денної вибірки. Желудев та ін. аналізували думки з годинною дискретністю. Ними показано, що зміні емоційного забарвлення твітів передуює зміна вартості акцій відповідної компанії.

**Мета статті, завдання і методологія дослідження.** Метою дослідження є розробка методики та комплексу моделей прогнозування валютного курсу на основі аналізу фактуального (об'єктивного) та концептуального (суб'єктивного) контенту мережі Інтернет.

Детальна схема дослідження наведена на рис. 1.

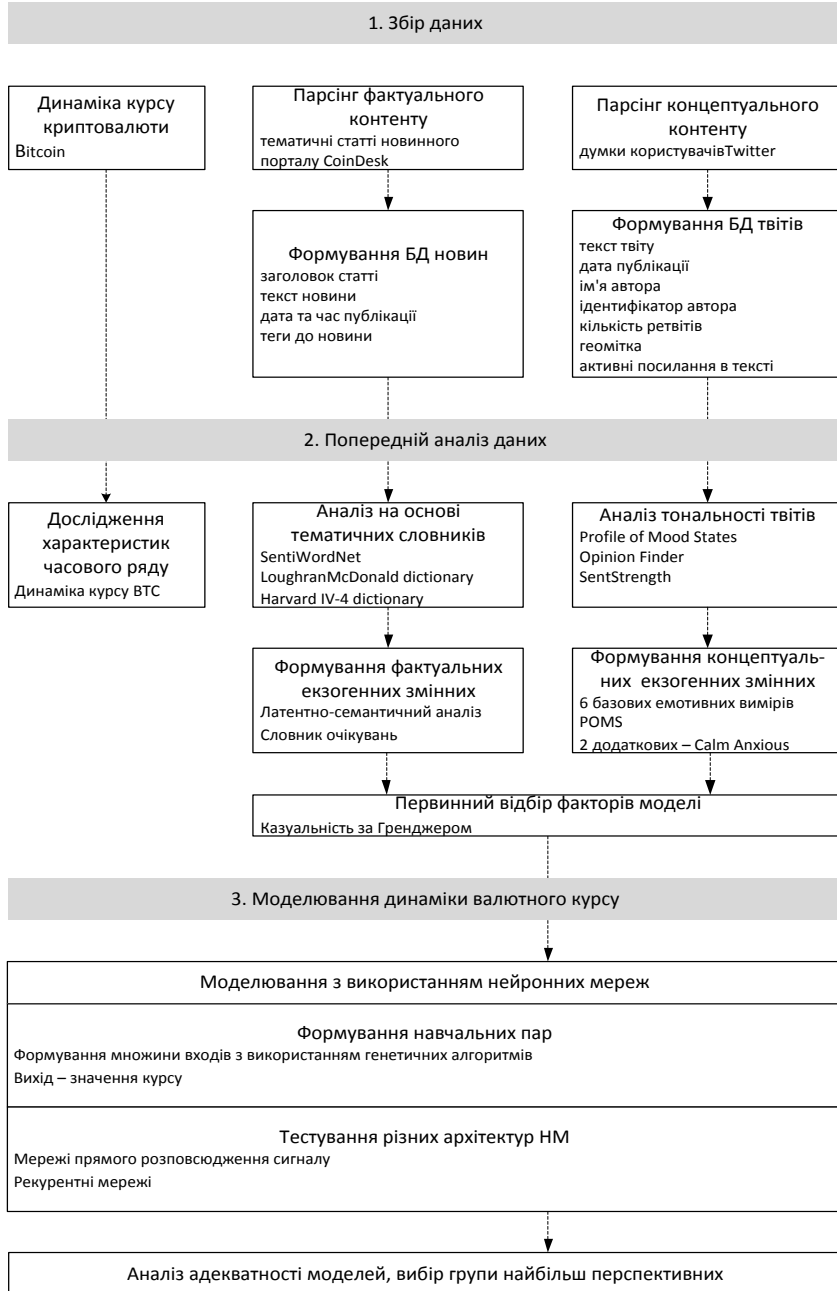


Рис. 1 – Задачі та інструменти дослідження

*Джерело: авторська розробка*

**Основні результати дослідження.** Інформація про динаміку курсу BTC / USD доступна за всю недовгу історію існування криптовалюти (рис. 2).

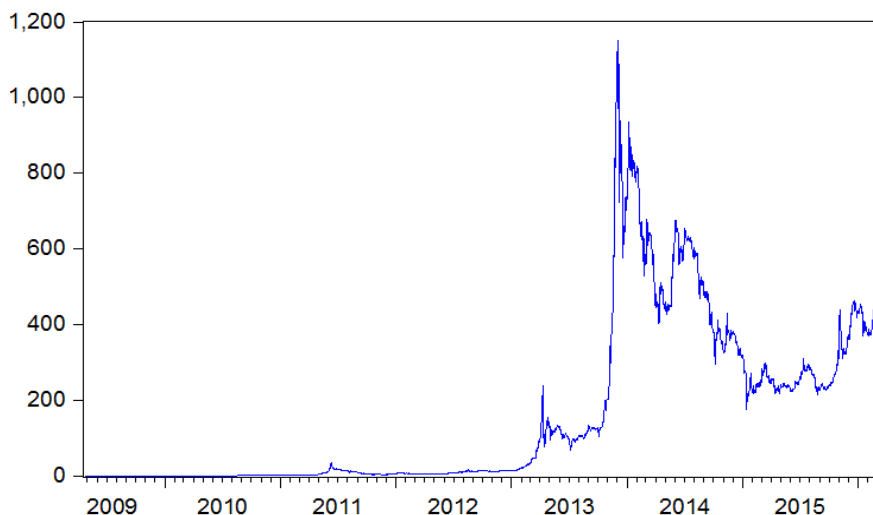


Рис. 2 – Динаміка курсу BTC / USD за 2009–2016 рр.

*Джерело: (Blockchain)*

Для побудови та верифікації комплексу моделей обрано проміжок часу тривалістю в два роки з 21.06.2014 по 21.06.2016. Цей інтервал, з одного боку, включає проміжки з тривалими трендами обох напрямків, з іншого боку, він відноситься до часу, коли кількість гравців на ринку, а також їх географія стали настільки великі, що роль спекуляцій значно послабилася. Вихідний ряд курсу BTC (*price*) згідно з ADF-тестом є нестационарним з довірчою ймовірністю  $p = 0.7466$  ( $t\text{-stat} = 1.0232$ ).

Для прогнозування динаміки BTC на основі аналізу думок користувачів використовувалася інформація двох типів: фактуальна і концептуальна. Джерелом фактуальної інформації обрано тематичні статті спеціалізованих новинних порталів, джерелом концептуальної – записи користувачів у мікроблогах.

Розглянемо процес збору та попереднього дослідження об'єктивного контенту. Вихідна база новинних статей була зібрана на основі інформації, що агрегована порталом Coindesk (Coindesk). Зберігалися: заголовок статті, текст новини, дата та час публікації, теги до новини (при їх наявності). На ресурсі Coindesk за період з 1.04.2013 по 24.06.2016 було опубліковано 6186 новин, які було зібрано у таблицю наступного вигляду (рис. 3).

Id	Content	Tidy_content	Tags
...	Фільтр	Фільтр	Фільтр
1	<div class="si... San Francisco-based digital currency exchange Kraken is making yet another acquisition, scoop...	San Francisco-based digital currency exchange Kraken is making yet another acquisition, scoop...	Acquisitions, ...
2	<div class="si... ItBit is considering whether to add support for ether, the native cryptocurrency of ethereum, to	ItBit is considering whether to add support for ether, the native cryptocurrency of ethereum, to	Ether, Ethere...
3	<div class="si... Friendly regulatory policies were rapidly turning London into a capital for blockchain innovation...	Friendly regulatory policies were rapidly turning London into a capital for blockchain innovation...	Brexit, Europe...
4	<div class="si... A bitcoin investment firm led by two former JPMorgan traders has published a note which spec...	A bitcoin investment firm led by two former JPMorgan traders has published a note which spec...	Brexit, Prices
5	<div class="si... A bitcoin exchange startup based in Singapore has raised \$16m as part of an ongoing Series A ...	A bitcoin exchange startup based in Singapore has raised \$16m as part of an ongoing Series A ...	Asia, Funding,...
6	<div class="si... Martin Hagelstrom is a bitcoin enthusiast, project executive and consultant working on IT	Martin Hagelstrom is a bitcoin enthusiast, project executive and consultant working on IT	Digital Wallet...

Рис. 3 – Фрагмент бази новин порталу Coindesk

*Джерело: авторські розрахунки*

Далі була складена таблиця частот появи слів протягом кожного дня. Слова піддавалися стемінгу за алгоритмом Портера (Алгоритм Портера) – обрізці закінчень для того, щоб об'єднати в одну групу однокореневі слова та форми однини та множини. Для формування

фактуальних екзогенних змінних були досліджені групи слів, що використовувались спільно в багатьох новинах. Для цього отримана вибірка була досліджена з використанням латентно-семантичного аналізу – таблиця частот піддавалася сингулярному розкладанню. Два перші сингулярні вектори розглядалися як координати слова на площині. Отримана карта (рис. 4) була проаналізована методом ієрархічної агломеративної кластеризації (“Метод ієрархічної агломеративної кластеризації”, 2016), в результаті якої виділено 14 однорідних груп.

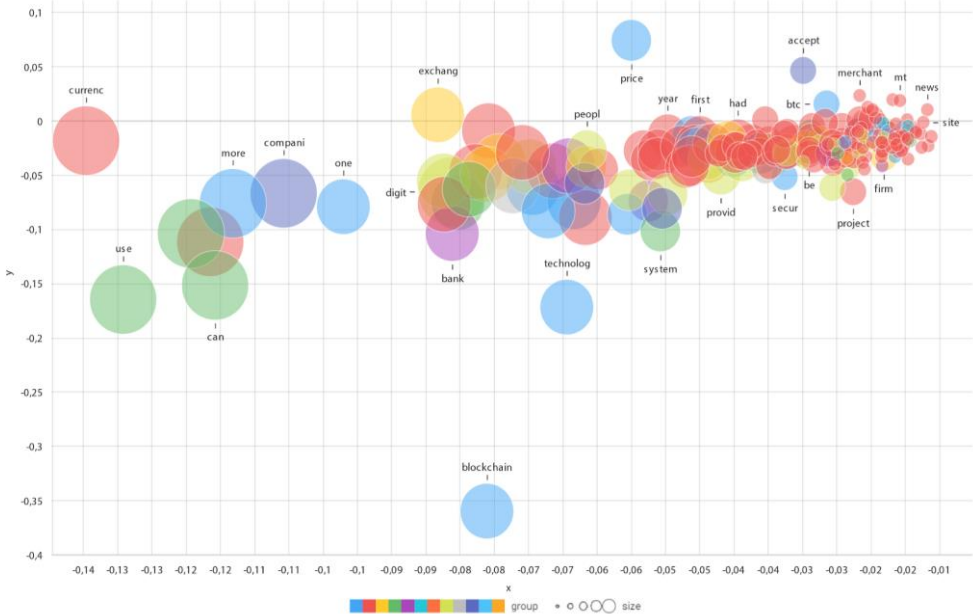


Рис. 4 – Результати латентно-семантичного аналізу (площа, що відповідає слову на карті, пропорційна частоті його появи в текстах)  
Джерело: авторські розрахунки

Згідно з тестом каузальності Гренджера, лише три (назвемо їх *infrastructure*, *activity*, *dissemination*) з чотирнадцяти однорідних груп слів-маркерів виявилися значущими (рис. 5). На їх основі були побудовані динамічні ряди частотності відповідних слів.

Hypothesis	lag 01	lag 02	lag 03	lag 04	lag 05	lag 06	lag 07
<b>DNUM10Z does not Granger Cause DPRICEZ</b>	0.0363	0.0589	0.1404	0.1468	0.2191	0.1279	0.0826
<b>DNUM1Z does not Granger Cause DPRICEZ</b>	0.318	0.0391	0.0686	0.1034	0.1214	0.1706	0.2381
<b>DNUM7Z does not Granger Cause DPRICEZ</b>	0.9233	0.0414	0.0761	0.01	0.0136	0.0222	0.0089
DNUM0Z does not Granger Cause DPRICEZ	0.1789	0.1526	0.2894	0.1526	0.2521	0.1923	0.0938
DNUM4Z does not Granger Cause DPRICEZ	0.2211	0.121	0.2596	0.1922	0.3157	0.2783	0.1395
DNUM12Z does not Granger Cause DPRICEZ	0.8673	0.346	0.3169	0.0855	0.0977	0.1179	0.0324
DNUM5Z does not Granger Cause DPRICEZ	0.1499	0.3687	0.5808	0.7058	0.1199	0.1412	0.1626
DNUM8Z does not Granger Cause DPRICEZ	0.8594	0.2399	0.4003	0.299	0.4074	0.472	0.3112
DNUM6Z does not Granger Cause DPRICEZ	0.2282	0.2614	0.4437	0.523	0.6894	0.7189	0.7623
DNUM2Z does not Granger Cause DPRICEZ	0.4925	0.4784	0.6784	0.542	0.6442	0.7558	0.7314
DNUM3Z does not Granger Cause DPRICEZ	0.6243	0.8702	0.946	0.8082	0.7443	0.636	0.7041
DNUM13Z does not Granger Cause DPRICEZ	0.5234	0.8207	0.928	0.7831	0.876	0.7829	0.7108
DNUM9Z does not Granger Cause DPRICEZ	0.858	0.9301	0.9452	0.9336	0.9468	0.9012	0.9089

Рис. 5 – Результати тесту Гренджера для виявлених груп слів-маркерів (жирним виділені значущі показники, відповідні змінним *infrastructure*, *activity*, *dissemination*)  
Джерело: авторські розрахунки

Для перевірки гіпотези «неважливо, про що говорять новинні статті, головне – наскільки очікувано ту чи іншу подію», був складений словник слів, що використовуються для

висловлення надій та очікувань<sup>1</sup>. В результаті тестування гіпотези за критерієм Гренджера, її було визнано значущою. Таким чином, список фактуальних екзогенних змінних було розширено за рахунок змінної *expect*.

Для аналізу суб'єктивного контенту збиралася вибірка твітів користувачів з 21.06.2014 по 21.06.2016 – було проімітовано запит анонімного користувача до служби розширеного пошуку Twitter<sup>2</sup>. Пошук проводився тільки серед англомовних твітів за пошуковим запитом «BTC/bitcoin». У результатах пошуку будуть записи, що містять хоча б одне з заданих слів. За вказаний період в базі було збережено близько 15 млн твітів<sup>3</sup>, структура бази така: текст твіта, дата публікації, ім'я автора, ідентифікатор автора, кількість ретвітів, геометка (при її наявності), активні посилання в тексті твіта (при їх наявності).

На початковому етапі твіти досліджувалися на основі одновимірної шкали «позитивний – негативний». Для цього були використані два інструменти, що дозволяють оцінювати полярність текстів: Opinion Finder (Wilson, Hoffmann, Somasundaran, Kessler, Wiebe, 2005) та SentiStrength (Thelwall, Buckley, Paltoglou, Cai, 2010). Однак аналіз причинності за Гренджером отриманих показників показав, що їх недоцільно використовувати для побудови моделей. Тому для подальшого аналізу був застосований психометричний інструмент POMS (Pollock, Cho, Reker, Volavka, 1979), що дозволяє оцінити настрій на основі шести емотивних вимірювань: *tension – anxiety* (напруга – тривожність), *depression – dejection* (депресія – пригніченість), *anger – hostility* (гнів – ворожість), *fatigue – inertia* (втома – інерція), *vigor – activity* (бадьорість – діяльність), *confusion – bewilderment* (сплутаність – розгубленість).

У статті Болена (Bollen, Mao, Zeng, 2010) описується метод розширення оригінального списку з 65 прикметників до більше ніж 900 слів за допомогою n-грам Google (Google Books. Ngram Viewer), однак відтворити досвід Болена не вдалося, тому словники POMS були розширені синонімами зі словника WordNet (WordNet). На основі розширеного словника зібрана база твітів була розкладена на емотивні вимірювання та для кожного була перевірена гіпотеза про його вплив на курс BTC (рис. 6). У результаті аналізу каузальності за Гренджером тільки дві концептуальні змінні – *calm* та *confusion* – виявилися значущими.

Hypothesis	lag 01	lag 02	lag 03	lag 04	lag 05	lag 06	lag 07
VIGOURZ does not G	0.1652	0.355	0.5175	0.6824	0.7355	0.805	0.8476
TENSIONZ does not	0.4405	0.4845	0.599	0.6808	0.81	0.8725	0.8777
FATIGUEZ does not	0.1795	0.4358	0.2624	0.4252	0.4573	0.621	0.7789
DEPRESSIONZ does	0.5994	0.2935	0.2426	0.4238	0.5392	0.6395	0.4791
CONFUSIONZ does	0.1875	0.0993	0.2004	0.2949	0.3926	0.382	0.26
<b>CALMZ does not G</b>	<b>0.7036</b>	<b>0.009</b>	<b>0.0197</b>	<b>0.0307</b>	<b>0.0111</b>	<b>0.0123</b>	<b>0.007</b>
ANXIOUSZ does not	0.7718	0.6057	0.4489	0.5414	0.5705	0.6709	0.8143
ANGERZ does not G	0.4172	0.6116	0.6008	0.8592	0.8303	0.8669	0.9084

Рис. 6 – Результати тесту Грейнджера для емотивних змінних (*vigour/tension/.../anger*)  
Джерело: авторські розрахунки

Таким чином, на основі аналізу новинних стрічок та твітів було сформовано множину екзогенних змінних моделі: чотири фактуальні – *infrastructure*, *activity*, *dissemination* та *expect*, а також дві концептуальні – *calm* та *confusion*. На їх основі було поставлено завдання прогнозування курсу BTC з використанням об'єктивного та суб'єктивного контенту.

Для моделювання були використані нейронні мережі прямої передачі сигналу та рекурентні мережі LSTM (рис. 7), які функціонують з короткою і довгостроковою пам'яттю: на кожному етапі навчання нейромережа визначає, наскільки можна забути попередню інформацію та наскільки важлива нова (Olah, 2015).

<sup>1</sup> Наприклад, future, next, following, tomorrow, shall, should, could, might, must, will, going to, expect\*, anticipate\*, predict\*, look/s/ing for, look/s/ing to, wait/s/ing, look/s/ing forward, hope/s, go/s/ing for. Стемовані слова позначені "\*", різні варіанти закінчень вказані через "/".

<sup>2</sup> Ця служба дозволяє задати пошуковий запит, діапазон дат і мову твітів. Інші можливості, такі як пошук твітів певних авторів, не використовувалися.

<sup>3</sup> Твіти, які є ретвітами, зберігалися в базі повторно.

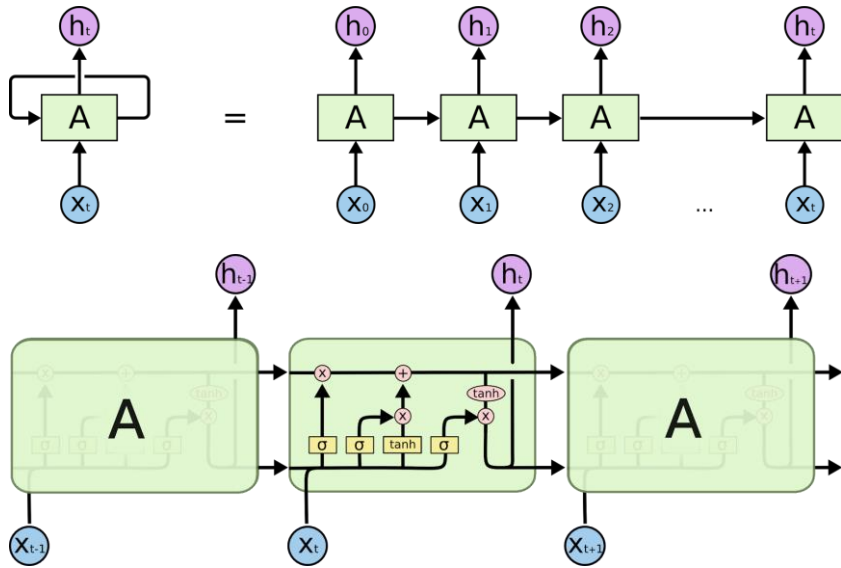


Рис. 7 – Архітектура LSTM-RNN

Джерело: (Olah, 2015).

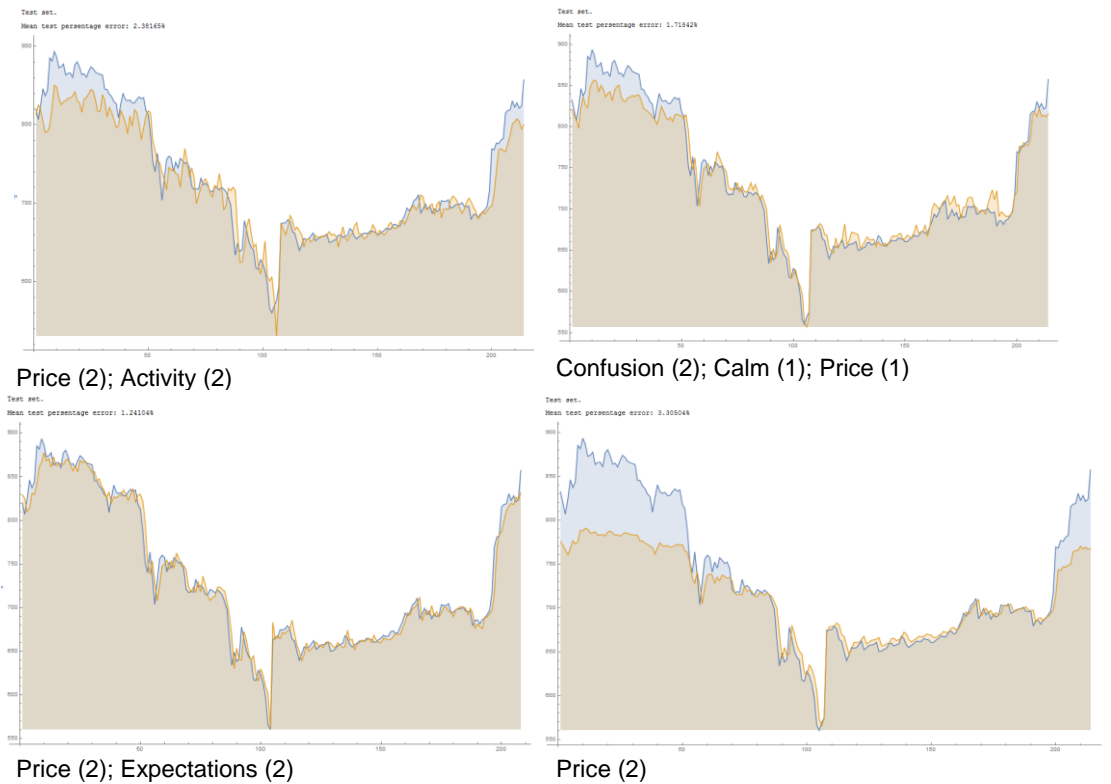


Рис. 8 – Порівняння якості мереж на різних наборах екзогенних змінних  
Джерело: авторські розрахунки



Автоматизація пошуку оптимальної архітектури нейронних мереж виконувалася з використанням генетичних алгоритмів. У якості хромосоми використовувався набір лагів по кожній змінній, яка подавалася на вхід нейромережі довжина хромосоми дорівнювала числу змінних, не враховуючи лагів. Алелі могли приймати значення від -1 до 7 (значення -1 повністю виключало даний параметр з розгляду, невід'ємні значення відповідали максимальному лагу параметра, який включався до розрахунку). Сформовані особини піддавалися схрещуванню, мутації та відбору на основі їх пристосованості (якості передбачення для валідаційного набору даних). У всіх експериментах мінімальні значення середньоквадратичної помилки були отримані на множині входів, що включають екзогенні змінні. Це підтверджує доцільність використання інтернет-контенту для прогнозування фінансових рядів (рис. 8).

**Висновки.** Найкраща якість прогнозування отримана на моделі, що побудована на основі словника очікувань, що підтверджує гіпотезу «неважливо, про що говорять новинні статті, головне – наскільки очікувано ту чи іншу подію», а також свідчить на користь ефекту зверхреакції. В цілому, порівняльний аналіз різних архітектур нейронних мереж, побудованих на основі актуальних та казуальних змінних, продемонстрував високу адекватність моделей і дозволив обґрунтувати доцільність використання інтернет-контенту для прогнозування валютного курсу.

### Література

1. Zheludev I., Smith R., Aste T. When Can Social Media Lead Financial Markets? [Електронний ресурс]. – Режим доступу : <http://www.nature.com/articles/srep04213>
2. Zhang X., Fuehres H., Gloor P. Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear" [Електронний ресурс]. – Режим доступу : doi:10.1016/j.sbspro.2011.10.562
3. Lachanski M. Did Twitter "Calm"-ness Really Predict the DJIA? [Електронний ресурс]. – Режим доступу : <http://files.meetup.com/7616132/DC-NLP-2015-07%20Michael%20Lachanski.pdf>
4. Mao Y., Wei W., Wang B. Twitter Volume Spikes: Analysis and Application in Stock Trading [Електронний ресурс]. – Режим доступу : <http://nlab.engr.uconn.edu/papers/SNAKDD025.pdf>
5. Preis T., Moat H., Stanley E. Quantifying Trading Behavior in Financial Markets Using Google Trends [Електронний ресурс]. – Режим доступу : <http://www.nature.com/articles/srep01684>
6. Ruiz E. J., Hristidis, V., Castillo C., Gionis A., Jaimes, A. Correlating Financial Time Series with Micro Blogging Activity [Електронний ресурс]. – Режим доступу : <http://www.cs.ucr.edu/~vangelis/publications/wsdm2012-microblog-financial.pdf>
7. Challet, D., Ayed, A. Predicting financial markets with Google Trends and not so random keyword [Електронний ресурс]. – Режим доступу : <https://arxiv.org/pdf/1307.4643v3.pdf>
8. Blockchain [Електронний ресурс]. – Режим доступу : <http://blockchain.info>
9. Бейкер К., Нofsингер Дж. Поведенческие финансы. Инвесторы, компании, рынки. Маросейка, 2016.
10. Coindesk [Електронний ресурс]. – Режим доступу : <http://www.coindesk.com/>
11. Алгоритм Портера [Електронний ресурс]. – Режим доступу : [http://www.cs.toronto.edu/~frank/csc2501/Readings/R2\\_Porter/Porter-1980.pdf](http://www.cs.toronto.edu/~frank/csc2501/Readings/R2_Porter/Porter-1980.pdf)
12. Метод ієрархічної агломеративної кластеризації [Електронний ресурс]. – Режим доступу : <http://www.mathworks.com/help/stats/hierarchical-clustering.html>
13. Bollen J., Mao H., Zeng X. Twitter mood predicts the stock market [Електронний ресурс]. – Режим доступу : [http://arxiv.org/PS\\_cache/arxiv/pdf/1010/1010.3003v1.pdf](http://arxiv.org/PS_cache/arxiv/pdf/1010/1010.3003v1.pdf)
14. Sloot D. The junk science behind the 'Twitter Hedge Fund' [Електронний ресурс]. – Режим доступу : <http://sellthenews.tumblr.com/post/21067996377/noitdoesnot>
15. Sharma J., Vyas A. Twitter Sentiment Analysis [Електронний ресурс]. – Режим доступу : <http://www.cse.iitk.ac.in/users/cs365/2012/submissions/jaysha/cs365/projects/report.pdf>
16. Wilson T., Hoffmann P., Somasundaran S., Kessler J., Wiebe J. OpinionFinder: A system for subjectivity analysis [Електронний ресурс]. – Режим доступу : <http://people.cs.pitt.edu/~swapna/papers/OpinionFinder-extendedabstract.pdf>
17. Thelwall M., Buckley K., Paltoglou G., Cai D. Sentiment Strength Detection in Short Informal Text [Електронний ресурс]. – Режим доступу : <http://www.scit.wlv.ac.uk/~cm1993/papers/SentiStrengthPreprint.doc>
18. Pollock V., Cho D., Reker D., Volavka J. Profile of mood states: the factors and their psychological correlates [Електронний ресурс]. – Режим доступу : <http://sci-hub.cc/10.1097/00005053-197910000-00004>
19. Google Books. Ngram Viewer [Електронний ресурс]. – Режим доступу : <https://books.google.com/ngrams>
20. WordNet [Електронний ресурс]. – Режим доступу : <http://wordnetweb.princeton.edu/perl/webwn>
21. Olah C. Understanding LSTM Networks [Електронний ресурс]. – Режим доступу : <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>

## References

1. Zheludev I., Smith R., & Aste T. (2014). When Can Social Media Lead Financial Markets? *Scientific reports* 4, Article number: 4213 doi:10.1038/srep04213
2. Zhang X., Fuehres H., & Gloor P. (2011). Predicting Stock Market Indicators Through Twitter "I hope it is not as bad as I fear". *Procedia Social and Behavioral Sciences*, 26, 55-62 doi:10.1016/j.sbspro.2011.10.562
3. Lachanski M. (2014). Did Twitter "Calm"-ness Really Predict the DJIA? *Journal of Undergraduate Research in Finance*, Volume 4(10). Retrieved from: <http://files.meetup.com/7616132/DC-NLP-2015-07%20Michael%20Lachanski.pdf>
4. Mao Y., Wei W., Wang B. (2013, August 11). Twitter Volume Spikes: Analysis and Application in Stock Trading. Proceedings of the 7th Workshop on Social Network Mining and Analysis. Retrieved from: <http://nlab.engr.uconn.edu/papers/SNAKDD025.pdf>
5. Preis, T., Moat, H., & Stanley, E. (2013). Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific reports* 3, Article number: 1684 doi:10.1038/srep01684
6. Ruiz E. J., Hristidis, V., Castillo C., Gionis A., & Jaimes, A. (2012) Correlating Financial Time Series with Micro Blogging Activity. *Conference: Proceedings of the Fifth International Conference on Web Search and Web Data Mining*. doi: 10.1145/2124295.2124358
7. Challet, D., & Ayed, A. (2013, July) Predicting financial markets with Google Trends and not so random keyword. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2310621
8. Blockchain. Retrieved from: <http://www.blockchain.info>
9. Baker H.K., Nofsinger J.R. (2016) Behavioral finance: Investors, Corporations, and Markets. Moscow: Maroseyka.
10. Coindesk. Retrieved from: <http://www.coindesk.com/>
11. Porter, M. (1980, July). An algorithm for suffix stripping. Program, 14 no. 3, 130-137. Retrieved from: [http://www.cs.toronto.edu/~frank/csc2501/Readings/R2\\_Porter/Porter-1980.pdf](http://www.cs.toronto.edu/~frank/csc2501/Readings/R2_Porter/Porter-1980.pdf)
12. Hierarchical Clustering. Retrieved from: <http://www.mathworks.com/help/stats/hierarchical-clustering.html>
13. Bolen, J., Mao, H., and Zen, X. (2011). Twitter moodpredicts the stock market. *Journal of Computational Science*, 2(1), 1-8. doi 10.1016/j.jocs.2010.12.007
14. Sloat D. The junk science behind the 'Twitter Hedge Fund'. Retrieved from: <http://sellthenews.tumblr.com/post/21067996377/noitdoesnot>
15. Sharma J., & Vyas A. (2012). Twitter Sentiment Analysis. Retrieved from: <http://www.cse.iitk.ac.in/users/cs365/2012/submissions/jaysha/cs365/projects/report.pdf>
16. Wilson T., Hoffmann, P., Somasundaran, S., Kessler, J., & Wiebe J. (2005). OpinionFinder: A system for subjectivity analysis. Proceedings of the Conference Human Language Technology Conference and Conference on Empirical Methods in Natural doi: 10.3115/1225733.1225751
17. Thelwall M., Buckley K., Paltoglou G., Cai D. Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, 61(12), 2544-2558 Retrieved from: <http://www.scit.wlv.ac.uk/~cm1993/papers/SentiStrengthPreprint.doc>
18. Pollock, V., Cho, D., Reker, D, Volavka, J. (1979). Profile of mood states: the factors and their psychological correlates. *The Journal of Nervous and Mental Disease* 167(10) Retrieved from: <http://sci-hub.cc/10.1097/00005053-197910000-00004>
19. Google Books. Ngram Viewer. Retrieved from: <https://books.google.com/ngrams>
20. WordNet. Retrieved from: <http://wordnetweb.princeton.edu/perl/webwn>
21. Olah, C. (2015). Understanding LSTM Networks. Retrieved from: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>