

ПРОГНОЗУВАННЯ КУРСІВ АКЦІЙ НА ОСНОВІ ДАНИХ ІЗ СОЦІАЛЬНИХ МЕРЕЖ

© 2016 ЧОРНОУС Г. О., ЯРМОЛЕНКО Ю. А.

УДК 661.65

Чорноус Г. О., Ярмоленко Ю. А. Прогнозування курсів акцій на основі даних із соціальних мереж

У статті досліджено можливість реалізації моделей машинного навчання для прогнозування курсів акцій іноземних компаній на основі даних із соціальних мереж. Оскільки традиційні засоби прогнозування є малоєфективними в короткостроковому періоді, метою цієї роботи стало доведення зв'язку між текстовим контентом соціальних мереж стосовно діяльності компаній та динамікою відповідних їм фінансових інструментів на біржі. За допомогою інструментів обробки людської мови була досліджена емоційна складова повідомлень у мережі Twitter стосовно діяльності компанії Google впродовж одного тижня. Було показано, яким чином емоції можна перевести до кількісної ознаки та використати дану величину в динаміці для побудови математичної моделі коливань курсу акцій. Дана модель реалізована методом штучних нейронних мереж. Авторами встановлено, що на даний момент питання використання запропонованої методології для прогнозування курсів акцій українських компаній є неактуальним, проте схожий інструментарій аналізу текстових даних було б доречно використовувати для соціологічних, політичних та інших економічних досліджень.

Ключові слова: прогнозування курсів акцій, дані соціальних мереж, аналіз текстових даних, аналіз настроїв, новини в соціальних мережах
Рис.: 2. **Табл.:** 1. **Бібл.:** 24.

Чорноус Галина Олександрівна – доктор економічних наук, доцент, доцент кафедри економічної кібернетики, Київський національний університет ім. Т. Шевченка (вул. Володимирська, 60, Київ, 01601, Україна)

E-mail: chornous@univ.kiev.ua

Ярмоленко Юлія Артурівна – аспірантка, Київський національний університет ім. Т. Шевченка (вул. Володимирська, 60, Київ, 01601, Україна)

E-mail: yu.yarmolenko@gmail.com

УДК 661.65

Черноус Г. А., Ярмоленко Ю. А. Прогнозирование курсов акций на основе данных из социальных сетей

В статье исследована возможность реализации моделей машинного обучения для прогнозирования курсов акций иностранных компаний на основе данных из социальных сетей. Так как традиционные методы прогнозирования являются малоэффективными в краткосрочном периоде, целью этой работы стало доказательство связи между текстовым контентом социальных сетей касательно деятельности компаний и динамикой соответствующих им финансовых инструментов на бирже. С помощью инструментов обработки человеческой речи была исследована эмоциональная составляющая сообщений в сети Twitter касательно деятельности компании Google на протяжении одной недели. Было показано, каким образом эмоции можно перевести в количественную характеристику и использовать эту величину для построения математической модели колебаний курса акций. Данная модель реализована с помощью метода нейронных сетей. Авторы статьи пришли к выводу, что на данный момент вопрос использования предложенной методологии для прогнозирования курсов акций украинских компаний остается неактуальным, однако похожий инструментальный анализ текстовых данных было бы уместно использовать в социологических, политических и других экономических исследованиях.

Ключевые слова: прогнозирование курсов акций, данные социальных сетей, анализ текстовых данных, анализ настроений, новости в социальных сетях
Рис.: 2. **Табл.:** 1. **Библ.:** 24.

Черноус Галина Александровна – доктор экономических наук, доцент, доцент кафедры экономической кибернетики, Киевский национальный университет им. Т. Шевченко (ул. Владимирская, 60, Киев, 01601, Украина)

E-mail: chornous@univ.kiev.ua

Ярмоленко Юлия Артуровна – аспирантка, Киевский национальный университет им. Т. Шевченко (ул. Владимирская, 60, Киев, 01601, Украина)

E-mail: yu.yarmolenko@gmail.com

UDC 661.65

Chornous G. O., Iarmolenko Ju. A. Forecasting Stock Prices Based on the Data from Social Networking

The article explores the possibility of implementing models of computer-assisted learning for the forecasting stock prices of foreign companies on the basis of data from social networking. Because the traditional forecasting methods are ineffective in the short term, the article is aimed at proving the available connection between the text-based content of social networking concerning activities of companies and dynamics of their respective financial tools at the exchange. Using the tools for processing human speech, the emotional component of messages on Twitter about Google company's activities for one week has been explored. It has been displayed how emotions can be transferred to the quantitative characteristics and how further this value can be used to build a mathematical model of fluctuations of the shares. This model has been implemented using the method of neural networks. The authors of the article have concluded that for the present moment the question about use of the proposed methodology to forecast stock prices of Ukrainian companies remains irrelevant, but similar text data analysis would be appropriate to carry out sociological, political and other economic studies.

Keywords: forecasting stock prices, data from social networking, analysis of textual data, analysis of sentiments, news on social networking
Fig.: 2. **Tbl.:** 1. **Bibl.:** 24.

Chornous Galyna O. – D. Sc. (Economics), Associate Professor, Associate Professor of the Department of Economic Cybernetics, Kyiv National University named after T. Shevchenko (60 Volodymyrska Str., Kyiv, 01601, Ukraine)

E-mail: chornous@univ.kiev.ua

Iarmolenko Iuliia A. – Postgraduate Student, Kyiv National University named after T. Shevchenko (60 Volodymyrska Str., Kyiv, 01601, Ukraine)

E-mail: yu.yarmolenko@gmail.com

Більшість моделей для прогнозування вартості цін фінансових інструментів використовують історичні дані для виявлення закономірностей часових рядів, виділення так званих технічних фігур на графіках і низку інших підходів, що дозволяють на основі ретроспективного аналізу визначити подальші коливання цін. Зазвичай такі моделі демонструють низьку ефективність, коли на ринку з'являється нетривіальна

інформація, що може сколихнути «передбачуваний» сценарій руху вартості інструменту. Проте, як показує практика, економічні гравці, а, відповідно, і ціни, є чутливими до важливих новин, що розповсюджуються різними медійними каналами.

Відповідно до гіпотези ефективного ринку вартість цінного папера формується на основі всієї наявної на ринку в певний момент часу інформації про нього.

Виходячи з цього, спрогнозувати ціну фінансового інструменту можливо, якщо проаналізувати всю наявну про нього інформацію на ринку. Соціальні мережі є одним із найповніших джерел, що містять як об'єктивні факти (події, що стосуються діяльності підприємств), так і суб'єктивні судження людей (оцінки цієї самої діяльності, відгуки про роботу компаній тощо). Слід зазначити, що, незалежно від того, достовірною є деяка новина в мережі чи ні, вона все одно має вплив на формування вартості, адже може бути помилково сприйнята економічними суб'єктами як правдива (особливо від впливових джерел).

Таким чином, існує можливість практичної реалізації прогностичних моделей курсів акцій компаній, що використовували б не лише логіку минулих подій, яка міститься в історичних даних, а й найактуальнішу інформацію про стан ринку. Нинішня ситуація у сфері інформаційних технологій та аналізу даних сприяє цьому. *По-перше*, глобальна інформатизація суспільства призвела до того, що постійно генерується величезна кількість користувацького контенту, який можна і необхідно аналізувати з метою виявлення суб'єктивних настроїв людей та їхніх оцінок економічної реальності. *По-друге*, найсучасніші інструменти аналізу текстових та інших типів даних дозволяють в автоматичному режимі перетворити емоційні характеристики контенту в числові показники. У даній статті буде показано, яким чином можна створити математичну модель для прогнозування курсів акцій компаній на основі текстових повідомлень у соціальних мережах.

Проблемою розробки методології та реалізації моделей для прогнозування курсів акцій займалися аналітики та науковці різних азіатських, американських та європейських шкіл: О. Alsing та О. Bahceci [1], Р. D. Azar та А. W. Lo [2], F. Beer [3], А. Buche і М. В. Chandak [4], Т. R. Burki [5], X. Li [6], Y. Lu [7], Т. Н. Nguyen [8], N. Oliveira [9], Т. Preis [10], Т. Rao і S. Srivstava [11]. В Україні практично відсутні відповідні дослідження, за пошуковими запитами можна знайти лише згадки про навчальні проекти в рамках курсів з Machine Learning. Проте слід відзначити значний вклад українських дослідників та науковців у теоретичне висвітлення інструментарію для обробки відповідних даних та супутніх алгоритмів: застосування нейронних мереж для моніторингу стану ринку акцій (О. В. Піскун, Т. Г. Міщенко [12]), покращення якості даних із соціальних мереж (О. О. Слабченко, В. М. Сидоренко [13]), інструментарій Data Mining та Text Mining (Д. В. Ланде [14], Н. Б. Паклін [15], О. І. Черняк [16], О. Ю. Чубукова [17]).

Вперше проблема була розглянута на початку 2010-х років, а за останні 5 років різними дослідниками було протестовано низку підходів до попередньої обробки даних і побудови відповідних моделей. Для кращого розуміння того, яким чином розвивалася наукова гіпотеза, слід розглянути динаміку досліджень останніх років та зазначити, до яких саме висновків приходили автори згаданих робіт. У 2013 р. у роботі [9] було визначено, що на основі динаміки твітів (повідомлень у мережі Twitter) можна спрогнозувати обсяг торгів наступного дня.

У цьому ж році у статті [3] автор доводить, що оцінки впливових інвесторів, висловлені в соціальних мережах, деякою мірою впливають на формування цін фінансових інструментів. Дослідження [10] того ж самого періоду вказує на те, що динаміка пошукових запитів сильно корелює з падінням і зростанням фінансових ринків. У статті [11] показано, що за допомогою повідомлень у соціальних мережах можна прогнозувати ціни на золото, нафту, а традиційні багатофакторна модель індексу DJIA може бути значно покращена при включенні до аналізу даних із соціальних мереж. Можна побачити, як ще у 2013 р. різні вчені приходили до нетривіальних результатів при дослідженнях соціальних мереж.

Характерним стало те, що вже у 2014 р. студенти провідних технічних вузів зацікавилися даною тематикою та поділилися результатами власних досліджень. Так, у магістерській роботі [5] автор показує, яким чином можна комбінувати різноманітні економетричні моделі з аналізом даних соціальних мереж для прогнозування курсів акцій. Тоді ж дослідниками з Китаю [6] вперше опубліковано статтю щодо прогнозування курсів акцій на основі аналізу настроїв, застосованого до повідомлень у мережі Twitter. За допомогою емпіричної перевірки на історичних даних було продемонстровано ефективність шести різних фінансових моделей. У наступному, 2015 р. автор кваліфікаційної роботи [1], окрім уже зазначених підходів, застосовує алгоритми машинного навчання до даних із соціальних мереж. У роботі [8] показано, як різні методи аналізу настроїв впливають на точність прогнозу. Водночас автор [7] досліджує питання прогнозування курсів акцій за повідомленнями в соціальних мережах у рамках різних галузей економіки та описує характер руху спеціально розроблених «секторальних індексів настрою».

Останні дослідження 2016 р. вкотре підтверджують силу зв'язку соціальних мереж із реальними економічними явищами. Так, автор [2] доводить, що навіть повідомлення від користувачів, які не мають безпосереднього стосунку до інвестування та економіки, все одно впливають на формування вартості курсів акцій. У статті [4] зроблено огляд найчастіше використовуваних технік, алгоритмів та підходів, які використовуються на кожному з етапів.

Таким чином, можна побачити, що інтерес до моделювання на основі даних із соціальних мереж є актуальним, а кількість досліджень та спектр аналітичного інструментарію зростають, тим самим даючи можливість отримувати все точніші результати досліджень.

Методологія. Як було зазначено раніше, інструментарій, що використовується на кожному із кроків побудови математичних моделей, може варіювати. У даній роботі описано процес створення моделі нейронної мережі для короткострокового прогнозування курсів акцій на основі застосування так званого аналізу настроїв до текстових повідомлень із соціальної мережі Twitter. Найбільш трудомістким етапом моделювання на основі даних із соціальних мереж є збір, попередня обробка та трансформація даних із текстової форми в кількісну. Найперші кроки, а саме – збір текстових по-

відомлень із мережі та попередня обробка можуть бути реалізовані за допомогою технології Twitter API (доступ до потоку текстового контенту, що використовують для розробки програмних додатків на базі Twitter) та відповідних програмних кодів мовами Python та R. У нашому випадку аналіз настроїв було реалізовано за допомогою бібліотеки sentiment для мови R, а побудова нейронної мережі – бібліотеки neuralnet у тому ж середовищі. Далі розглянемо більш детально методи аналізу настроїв та застосування нейронних мереж.

Відповідно до визначення, наведеного у [18], аналіз настроїв – це автоматичне розпізнавання оцінок та емоційної лексики, представленої в текстах. При аналізі текстових масивів у мережі Інтернет усі повідомлення можна поділити на дві групи: факти та оцінки. Визначення та ідентифікація оцінок є ключовою задачею при аналізі настроїв. Існує два основні підходи до класифікації повідомлень: через оцінку емоційного забарвлення кожного окремого слова та на основі моделей машинного навчання, натренованих на великій кількості текстового матеріалу для розпізнавання емоцій. Другий метод є більш ефективним в силу того, що він оцінює семантичні конструкції загалом, а не окремі слова в реченнях.

Використання методу нейронних мереж для прогнозування курсів акцій обумовлено високою ефективністю даного типу моделей при аналізі фінансових ринків. Завданню прогнозування часових рядів найкраще відповідають моделі нейронних мереж прямого розповсюдження, що використовують навчання з учителем, оскільки, на відміну від задачі кластеризації, наприклад, завжди можна підготувати багато контрольних прикладів, посилаючись на історичні дані. З їх допомогою можна обчислити похибку, необхідну для роботи алгоритмів. До таких мереж належать одношаровий перцептрон, мережі Ворда, перцептрон Розенблатта, багатшаровий перцептрон, мережі на основі радіальних базисних функцій, машини опорних векторів та інші. У цих мереж багато спільного, перш за все – алгоритм зворотного розповсюдження похибки, що допомагає узагальнювати клас вирішуваних з їх допомогою задач, а результати та обмеження – схожими. Згідно з дослідженням [19] при моделюванні фінансових ринків найбільшу ефективність показує мережа – одношаровий перцептрон.

Для того, щоб оцінити можливість використання даних із соціальних мереж для прогнозування курсів акцій компаній необхідно довести наявність зв'язку між текстовим контентом повідомлень користувачів та рухом цін відповідних акцій на біржі. Тобто, метою даної роботи є доведення значущості зв'язку та побудова ефективної прогностичної моделі на основі інтелектуального аналізу даних онлайн-потоків із соціальних мереж. Для досягнення мети необхідним є вирішення низки завдань: підбір необхідного інструментарію для обробки, аналізу даних і реалізації відповідних моделей. Також у рамках даного дослідження слід оцінити можливість реалізації відповідних моделей на українському ринку.

Головним завданням даної роботи є встановлення взаємозв'язку між динамікою тематичних повідомлень стосовно роботи деякої компанії в соціальній мережі Twitter та коливаннями курсу акцій відповідної компанії. З цією метою за допомогою інструменту Twitter API були зібрані повідомлення користувачів із ключовими словами «Google» та «\$GOOG» у режимі онлайн за період з 3 по 6 травня 2016 р. На жаль, на даний момент Twitter не дає можливості вільного доступу до історичних даних, вони продаються самою компанією та посередниками, проте можна завантажити в режимі реального часу (як тільки повідомлення з'являється у мережі, його можна зберегти разом із багатьма супутніми параметрами: час та місце створення, автор повідомлення, характеристики автора тощо). Також обмеженою є пропускна можливість каналу API, що не дозволяє перевищувати деяку швидкість потоку інформації для зчитування.

Після збору усіх необхідних даних (котирування акції за відповідний період з джерела [20] та текстові повідомлення у Twitter) дані приводяться до вигляду часових рядів. Котирування подаються у вигляді ряду з часовим інтервалом у 15 хвилин. Вибір такого невеликого часового інтервалу пов'язаний, по-перше, із природою новин та швидкістю їх розповсюдження (завдяки соціальним мережам новина може розходитися між користувачами за лічені хвилини), а по-друге – із об'єктивною необхідністю виділення якомога більшої кількості точок спостереження (умова використання методу нейронних мереж). Текстові повідомлення за допомогою попередньої обробки переводяться у вигляд, прийнятний для аналізу за допомогою бібліотеки sentiment, а далі – у числову оцінку настрою користувачів за відповідний період часу (із таким самим інтервалом у 15 хвилин). Оцінювання відбувається за шкалою від мінімального значення –100 (усі повідомлення негативні) до максимального значення 100 (усі повідомлення позитивні).

Первинний вигляд зібраних повідомлень є складно структурованим і містить величезну кількість характеристик цього самого повідомлення: Twitter API визначає користувача, що створив повідомлення, локацію пристрою, з якого надіслане повідомлення, стилістику тексту, довжину тексту (кількість символів), кількість послідовників користувача, що створив повідомлення тощо. Усі ці дані записуються у вигляді одного рядка у 600–1000 символів, тому їх необхідно попередньо обробити з метою вичленення самого тексту. Деякі дослідники намагалися виділити ті характеристики повідомлень, що дозволили б виокремити найбільш інформативні з них (наприклад, людина з більшою кількістю послідовників має вищий авторитет і, власне, її думка з більшою вірогідністю вплине на загальний настрій користувачів та викличе відповідну реакцію). У табл. 1 наведено приклад попередньої обробки даних і визначення полярності повідомлень за пошуковими запитами «Google» та «GOOG».

Легко помітити, що насправді більшість повідомлень, отриманих зі стриму Twitter, мають нейтральний зміст із нульовою полярністю. Проте кількість таких повідомлень навіть за досить короткі проміжки часу є великою, тим самим схилиючи дані в цілому до певної

Результати аналізу сентиментів для текстових повідомлень із мережі Twitter

Текст	Емоція	Полярність	Оцінка
The last time these insiders purchased this stock it skyrocketed dia qqq spy goog	Unknown	Positive	70
How to block location access in google chrome android guide read more goog	Unknown	Neutral	29
Google teaches self-driving cars to honk politely read more goog	Surprise	Neutral	32
Most misspelled words in each state google read more goog	Unknown	Neutral	-21
Review our daytrade and swingtrade history spy aapl amzn goog tsla uso	Unknown	Neutral	0
Google android n features we love in the new os read more	Joy	Positive	100
Google is recording your voice even when you're offline read more	Surprise	Neutral	-35

Джерело: авторська розробка.

полярності. Також слід зазначити наявність помилок при виведенні оцінки повідомлення: вона може значно відрізнятися від сприйняття тексту людиною.

Нагадаємо, що для дослідження були зібрані повідомлення за двома ключовими словами («Google» та «\$GOOG»), причому другий варіант пошуку, – фінансовий тег – зазвичай приводить до повідомлень на інвестиційну тематику (тег, що містить біржовий тикер компанії, використовують для повідомлень про привабливість фінансового інструмента). Як свідчить дослідження [2], для моделювання можуть бути використані обидва типи повідомлень. Проте для спрощення завдання та уникнення необхідності накладання додаткових обмежень на зібрані повідомлення було вирішено використати твіти за пошуковим запитом «\$GOOG». На рис. 1 зображено часові ряди котирувань курсу акції компанії Google та динаміки сентиментів повідомлень користувачів Twitter стосовно інвестиційної привабливості інструмента.

Легко помітити, що крива курсу акцій схожа на криву сентиментів із деяким часовим лагом. Це дозволяє висунути гіпотезу про те, що динаміка оцінок користувачів соціальної мережі на декілька годин випереджає рух ціни на біржі. На основі даних курсу акцій компанії Google з 15-хвилинним інтервалом та даних проведеного аналізу сентиментів будемо модель нейронної мережі з метою прогнозування курсу акцій у короткостро-

ковому періоді. Було візуально встановлено, що тренд сентиментів приблизно на 4 години випереджає тренд курсу акцій. Саме тому як вхідні змінні моделі були обрані такі параметри:

- ✦ значення оцінки сентиментів із лагом 16 (3,5 год);
- ✦ курс акції з лагом 2 (0,5 год);
- ✦ курс акції з лагом 4 (1 год);
- ✦ курс акції з лагом 6 (1,5 год).

Прогноз за даним фінансовим інструментом був побудований на 8 періодів вперед (рис. 2). Кількість спостережень для побудови моделей – 84, прогнозованих значень – 8. Напрямо динаміки тренду було спрогнозовано вірно, прогноз виявився близьким до реальних значень.

Як бачимо, загальний настрій користувачів соціальних мереж, що висловлюють свою думку, дійсно корелює з рухом ціни на біржі, що дає змогу прогнозувати динаміку в короткостроковому періоді. Подібне дослідження для інших акцій показало, що якість прогнозу сильно залежить від якості даних: кількості створюваних користувачами повідомлень, які дають змогу відобразити картину загальних настроїв у мережі; наявності історичної бази даних, що дозволяє ефективно натренувати нейронну мережу. Таким чином, прогнозування на основі даних із соціальних мереж є потужним інструментом та може бути використаним в інвестиційній діяльності.

Розглянемо питання перспектив використання описаного підходу до акцій українських компаній. Українська біржа, у порівнянні з аналогічними в інших країнах,

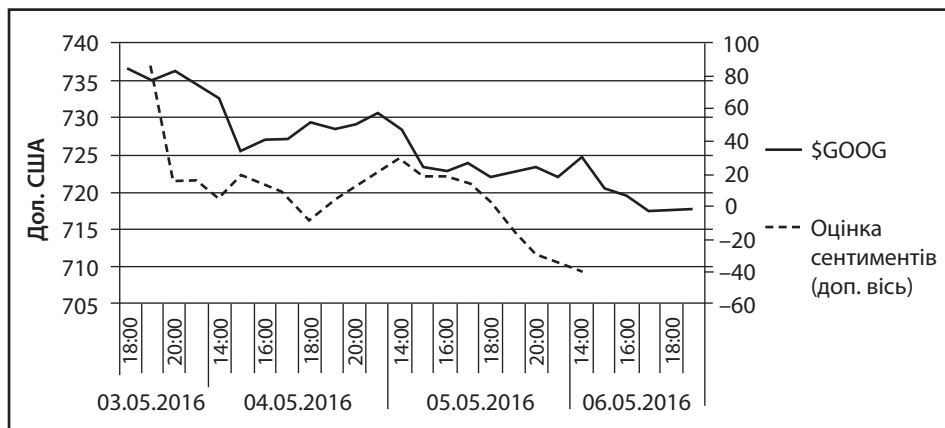


Рис. 1. Динаміка курсу акції \$GOOG та оцінки аналізу сентиментів повідомлень із відповідним фінансовим тегом

Джерело: складено за [20].

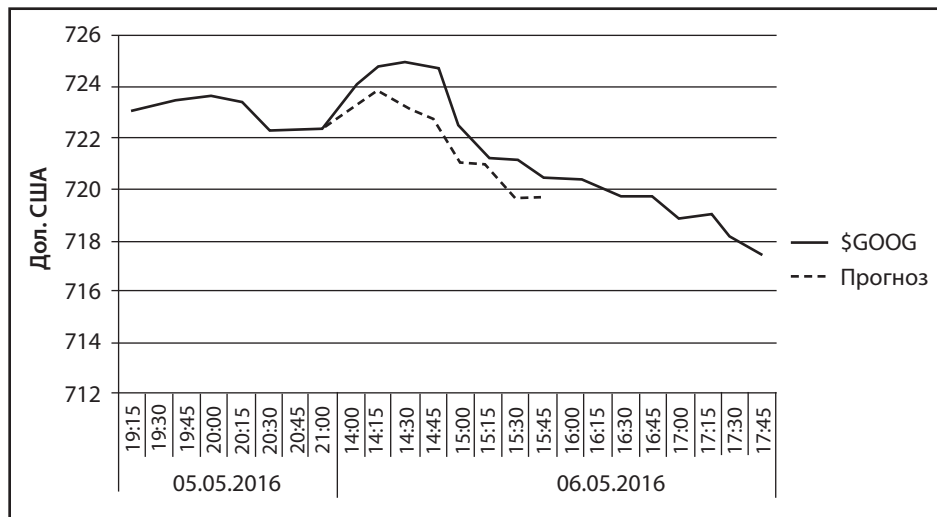


Рис. 2. Короткостроковий прогноз курсу акцій компанії Google

Джерело: авторська розробка.

особливо Європи, Північної Америки та країн Азії, слабо розвинута в контексті обсягів торгів на біржі та кількості учасників. Для порівняння, згідно з даними Української біржі, щоденні обсяги торгів коливаються в районі 10 млн грн [21], на біржі NYSE цей показник перевищує 3 млрд дол. [22]. На думку аналітиків та економістів, українська біржа в останні роки перетворилася на «кишенькову» та не має реального механізму для відображення реальної ринкової вартості активів підприємств.

Тим не менше, необхідно оцінити можливість проведення аналізу новин з метою виокремлення динаміки настроїв користувачів мережі Інтернет стосовно тих чи інших компаній. Оцінка технічних можливостей виконання цього завдання виглядає таким чином:

1. Джерелом новин для аналізу українських реалій може стати обмежене коло ресурсів у вигляді соціальних мереж та сайтів новин. Проте сайти новин часто не зовсім об'єктивно висвітлюють події, що трапляються, а найпопулярніші ресурси, такі як pravda.com.ua та ua.korespondent.net, не мають розділів, де публікувалися б новини стосовно роботи компаній або оцінок експертів у режимі коротких текстових повідомлень (зазвичай це готові статті, що містять вичерпну інформацію, проте підготовлені вони заздалегідь та на момент публікації вже втрачають актуальність для ринку акцій). Мережа Twitter є не такою популярною в Україні, перевірка кількості повідомлень, знайдених за результатами пошуку відомих українських компаній, що представлені на біржі, видає результат у декілька повідомлень на день. В Україні досить популярною є соціальна мережа Facebook, проте сайт розроблений таким чином, що робот зібрати дані в автоматичному режимі з нього не може, лише для чіткого переліку сторінок за числовим ідентифікатором. Єдиним джерелом даних може бути інформація GoogleTrends, проте вона містить лише динаміку кількості запитів за ключовим словом та не дає інформації про настрої користувачів стосовно діяльності компанії.

2. Задачі аналізу настроїв для текстів українською мовою поки що не можуть бути вирішені в автоматичному режимі через низку об'єктивних причин.

По-перше, ще не розроблений український словник для визначення тональності слів, а бази для обробки природної мови (*Natural Language Processing*) перебувають у зародковому стані. Проте, за ініціативи американсько-української компанії Grammarly, нарешті почалася розробка спеціальних корпусів української мови для автоматичної обробки текстів та проведення аналізу настроїв, зокрема [23].

Таким чином, наявні великі перспективи розробки аналітичних додатків для визначення полярності текстів українською мовою, проте застосувати їх до реальних даних з метою прогнозування курсів акцій українських компаній практично неможливо за поточних умов розвитку біржової торгівлі в Україні в цілому та ринку акцій зокрема. Натомість, аналіз настроїв українських текстів міг би добре спрацювати у сферах політики, соціальних досліджень тощо.

ВИСНОВКИ

У даній роботі було досліджено можливість використання даних із соціальних мереж з метою створення прогностичних моделей для курсів акцій компанії. Відповідно до теорії ефективного ринку вся наявна на ринку інформація про фінансовий інструмент та діяльність компанії формує ціну акцій. Беручи до уваги, що соціальні та медійні мережі є найбільшим джерелом корисних даних і, за умови симетрії інформації на ринку, можуть повністю та всебічно описати стан компанії, ці дані, переведені в кількісні показники, можна використати для створення математичних моделей, які описують рух цін на ринку акцій.

Було проведено повний цикл обробки отриманих із соціальної мережі Twitter даних для того, щоб на їх базі побудувати математичну модель прогнозування курсів акцій компанії Google. Процес дослідження цього зв'язку складався з декількох етапів. Перший етап передбачав налагодження онлайн-потоків повідомлень із соціальної мережі за ключовим словом. Цього вдалося досягти завдяки створенню програмного коду мовою

Python та підключення до ресурсу даних Twitter API, що дає змогу розробникам програмних застосунків на базі даних із мережі отримувати вичерпну інформацію щодо опублікованих повідомлень у режимі онлайн. Далі, за допомогою спеціальних пакетів обробки текстових даних для статистичної мови програмування R, було здійснено підготовку та очищення даних з метою подальшого проведення аналізу настроїв. Оцінка настроїв у кількісному вимірі дала змогу представити дані у вигляді часових рядів, що відображали наявну в мережі інформацію стосовно інвестиційної привабливості компаній. Залежність цін акцій від лагових значень та оцінок настроїв було подано у вигляді моделі штучної нейронної мережі, проведено порівняння прогнозованих значень із реальними.

У процесі дослідження зв'язку між динамікою повідомлень у соціальних мережах та курсом акцій були одержані такі висновки:

1. Існує тісний зв'язок між оцінкою діяльності компанії в соціальних мережах і курсом акцій тієї самої компанії. Цей зв'язок має часовий лаг, тобто новини в мережах дещо передують реальній зміні цін на ринку.

2. За допомогою динаміки оцінки настроїв повідомлень можна прогнозувати курс акцій у короткостроковий перспективі (декілька годин). Такий невеликий часовий проміжок пояснюється природою даних у соціальній мережі: вони поширюються дуже динамічно та всеохоплююче, тим самим спонукаючи інвесторів та дослідників до швидкої реакції на реальну ситуацію.

3. На основі динаміки новин, які набули поширення за період, поки біржа не працює, можна прогнозувати ціну акцій на момент відкриття робочого дня. Це є дуже цінним фактом, адже значні стрибки курсів акцій, як правило, припадають саме на початок торгового дня.

4. Якість обробки текстових даних і результатів аналізу настроїв часто має значну похибку, а визначення полярності повідомлень може не співпадати з реальною. Такий результат є закономірним в силу об'єктивної складності обробки неструктурованих даних за допомогою методів машинного навчання. Проте узагальнена динаміка на основі великої кількості повідомлень дає доволі реалістичну оцінку настроїв користувачів мережі.

Також слід зазначити, що була оцінена можливість використання подібної методики аналізу та моделювання для українського ринку акцій. На нашу думку, вона є низькою в силу того, що національні біржі не є тим інструментом врівноваження ринку, яким вони виступають у розвинених країнах, окрім того, обсяг торгів на вітчизняних платформах є надто низьким. Більше того, на даний час не існує інструментів для проведення аналізу настроїв текстів українською мовою.

Значним обмеженням у реалізації дослідження була складність видобутку необхідних текстових даних через неможливість отримання доступу до історичних повідомлень. Це не дало можливості всебічно оцінити вплив наявної в соціальних мережах інформації на курс акцій компанії та дослідити динаміку за тривалий період часу. Полем для подальшої роботи є відкрите питання щодо

можливості фільтрації найбільш інформативних даних серед усього арсеналу наявних у соціальній мережі, що стосуються діяльності окремих компаній на основі характеристик повідомлень, створених користувачами, та впливовості цих користувачів на медійному просторі. ■

ЛІТЕРАТУРА

1. **Alsing O., Bahceci O.** Stock Market Prediction Using Social Media Analysis: Degree Project in Computer Science at KTH Royal Institute of Technology. Stockholm, 2015. 50 p.

2. **Azar P. D., Lo A. W.** The Wisdom of Twitter Crowds: Predicting Stock Market Reactions to FOMC Meetings via Twitter Feeds. *SSRN*. March 2016. 21 p. URL: <http://dx.doi.org/10.2139/ssrn.2756815>

3. **Beer F., Hervé F., Zouaoui M.** Is Big Brother Watching Us? Google, Investor Sentiment and the Stock Market. *Economics Bulletin*. 2013. Vol. 33 (1). P. 454–466.

4. **Buche A., Chandak M. B.** Stock Market Prediction Using Text Opinion Mining: A Survey. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2016. Vol. 6 (6). P. 566–569.

5. **Burki T. R.** Social Media Aided Stock Market Predictions by Sparsity Induced Regression: Master of Science Thesis for the Degree of Master of Science in Systems and Control at Delft University of Technology. Delft, 2014. 93 p.

6. News Impact on Stock Price Return via Sentiment Analysis/ Li X., Xie H., Chen L., Wang J. and others. *Knowledge-Based Systems*. October 2014. Vol. 69. P. 14–23. URL: <http://dx.doi.org/10.1016/j.knsys.2014.04.022>.

7. Predicting Sector Index Movement with Microblogging Public Mood Time Series on Social Issues/Y. Lu, J. Guo, K. Sakamoto, H. Shibuki, T. Mori // Proceedings in Pacific Asia Conference on Language, Information and Computation. Shanghai, 2015. P. 563–571.

8. **Nguyen T. H., Shirai K., Velcin J.** Sentiment Analysis on Social Media for Stock Movement Prediction. *Expert Systems with Applications*. 2015. Vol. 42. URL: <http://dx.doi.org/10.1016/j.eswa.2015.07.052>.

9. **Oliveira N., Cortez P., Areal N.** Some Experiments on Modeling Stock Market Behavior Using Investor Sentiment Analysis and Posting Volume from Twitter // Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics. New York, 2013. P. 1–8. URL: <http://doi.acm.org/10.1145/2479787.2479811>

10. **Preis T., Moat S., Stanley H. E.** Quantifying Trading Behavior in Financial Markets Using Google Trends. *Scientific Reports*. April 2013. Vol. 3. P. 16–24. URL: <http://dx.doi.org/10.1038/srep01684>

11. **Rao T., Srivastava S.** Modeling Movements in Oil, Gold, Forex and Market Indices Using Search Volume Index and Twitter Sentiments // Proceedings of the 5th Annual ACM Web Science Conference. New York, 2013. P. 336–345. URL: <http://doi.acm.org/10.1145/2464464.2464521>.

12. **Піскун О. В., Міщенко Т. Г.** Застосування нейронних мереж для моніторингу стану фондового ринку // Інноваційна стратегія і тактика фінансово-економічного розвитку суб'єктів національного господарства: матеріали Міжнарод. наук.-практ. конф., Чернівці, 19–20 грудня 2014 р. У 3-х частинах. Чернівці: Видавничий дім «Гельветика», 2014. Ч. 3. С. 211–214.

13. **Слабченко О. О., Сидоренко В. М.** Аналіз і синтез моделей на основі машинного навчання для імпутації даних із персональних акаунтів соціальних мереж. *Вісник КрНУ імені Михайла Остроградського*. Травень 2014. № 88. С. 105–111.

14. **Ландэ Д. В.** Поиски знаний в Internet. Професиональная работа. М.: Диалектика, 2005. 272 с.

15. **Паклин Н. Б., Орешков В. С.** Бизнес-анализ: от данных к знаниям. СПб.: Питер, 2013. 704 с.

16. Черняк О. І., Захарченко П. В. Інтелектуальний аналіз даних: підручник. Київ: Знання, 2014. 599 с.
17. Чубукова О. Ю., Івченко Н. О., Кучер С. Л. Алгоритмізація та програмування процедур обробки економічної інформації. Київ: КНУТД, 2009. 138 с.
18. Pang B., Lee L. Opinion Mining and Sentiment Analysis. *Foundations and Trends in Information Retrieval*. 2008. Vol. 2, No. 1-2. P. 1–135. URL: <http://dx.doi.org/10.1561/15000000011>
19. Kryuchin O. V., Kozadaev A. S., Dudakov V. P. Forecasting Time Sets Using Artificial Neural Networks and Regression Models as the Example of Forecasting Currency Pairs Quotation. URL: <http://dx.doi.org/10.15827/0236-235X.110.070-074>
20. Yahoo Finance Official Web-site. AphabetInc. (GOOG). URL: <http://finance.yahoo.com/quote/GOOG?p=GOOG>
21. Офіційний сайт ПАО «Українська біржа». Результати торгів. URL: <http://www.ux.ua/ru/marketdata/marketresults.aspx> (01.10.2016).
22. New York Stock Exchange Official Web-site. NYSE: Transactions, Statistics and Data Library. URL: <https://www.nyse.com/data/transactions-statistics-data-library> (01.10.2016).
23. Grammarly Official Web-site. URL: <https://www.grammarly.com/> (01.10.2016).

REFERENCES

- Alsing, O., and Bahceci, O. *Stock Market Prediction Using Social Media Analysis: Degree Project in Computer Science at KTH Royal Institute of Technology*. Stockholm, 2015.
- Azar, P. D., and Lo, A. W. "The Wisdom of Twitter Crowds: Predicting Stock Market Reactions to FOMC Meetings via Twitter Feeds. SSRN. March 2016" <http://dx.doi.org/10.2139/ssrn.2756815>
- Beer, F., Herve, F., and Zouaoui, M. "Is Big Brother Watching Us? Google, Investor Sentiment and the Stock Market". *Economics Bulletin*. Vol. 33 (1) (2013): 454-466.
- Buche, A., and Chandak, M. B. "Stock Market Prediction Using Text Opinion Mining: A Survey". *International Journal of Advanced Research in Computer Science and Software Engineering*. Vol. 6 (6) (2016): 566-569.
- Burki, T. R. *Social Media Aided Stock Market Predictions by Sparsity Induced Regression*: Master of Science Thesis for the Degree of Master of Science in Systems and Control at Delft University of Technology: Delft, 2014.
- Cherniak, O. I., and Zakharchenko, P. V. *Intelektualnyi analiz danykh* [Data mining]. Kyiv: Znannia, 2014.
- Chubukova, O. Yu., Ivchenko, N. O., and Kucher, S. L. *Alhorytmizatsiia ta prohramuvannia protsedur obrobky ekonomichnoyi informatsii* [Algorithmization and programming procedures for the processing of economic information]. Kyiv: KNUITD, 2009.
- Grammarly Official Web-site. <https://www.grammarly.com/>
- Kryuchin, O. V., Kozadaev, A. S., and Dudakov, V. P. "Forecasting Time Sets Using Artificial Neural Networks and Regression Models as the Example of Forecasting Currency Pairs Quotation". <http://dx.doi.org/10.15827/0236-235X.110.070-074>
- Lande, D. V. *Poiski znaniy v Internet. Professionalnaya rabota* [The search for knowledge on the Internet. Professional work]. Moscow: Dialektika, 2005.
- Li, X. et al. "News Impact on Stock Price Return via Sentiment Analysis" *Knowledge-Based Systems*. <http://dx.doi.org/10.1016/j.knosys.2014.04.022>
- Lu, Y. et al. "Predicting Sector Index Movement with Microblogging Public Mood Time Series on Social Issues". In *Proceedings in Pacific Asia Conference on Language, Information and Computation*, 563-571. Shanghai, 2015.
- "New York Stock Exchange Official Web-site. NYSE: Transactions, Statistics and Data Library". <https://www.nyse.com/data/transactions-statistics-data-library>
- Nguyen, T. H., Shirai, K., and Velcin, J. "Sentiment Analysis on Social Media for Stock Movement Prediction". *Expert Systems with Applications*. <http://dx.doi.org/10.1016/j.eswa.2015.07.052>
- Oliveira, N., Cortez, P., and Areal, N. "Some Experiments on Modeling Stock Market Behavior Using Investor Sentiment Analysis and Posting Volume from Twitter". <http://doi.acm.org/10.1145/2479787.2479811>
- Piskun, O. V., and Mishchenko, T. H. "Zastosuvannia neuronnykh merezh dlia monitorynhu stanu fondovoho rynku" [Application of neural networks for monitoring the state of the stock market]. *Innovatsiina stratehiia i tatyka finansovo-ekonomichnoho rozvytku sub'ektiv natsionalnoho hospodarstva*. Vol. 3. Chernivtsi: Helvetyka, 2014. 211-214.
- Pang, B., and Lee, L. "Opinion Mining and Sentiment Analysis" *Foundations and Trends in Information Retrieval*. <http://dx.doi.org/10.1561/15000000011>
- Paklin, N. B., and Oreshkov, V. S. *Biznes-analiz: ot danykh k znaniyam* [Business analysis: from data to knowledge]. St. Petersburg: Piter, 2013.
- Preis, T., Moat, S., and Stanley, H. E. "Quantifying Trading Behavior in Financial Markets Using Google Trends". <http://dx.doi.org/10.1038/srep01684>
- "Rezultaty torhiv" [The results of the auction]. Ofitsiinyi sait PАО «Ukrainska birzha». <http://www.ux.ua/ru/marketdata/marketresults.aspx>
- Rao, T., and Srivastava, S. "Modeling Movements in Oil, Gold, Forex and Market Indices Using Search Volume Index and Twitter Sentiments". <http://doi.acm.org/10.1145/2464464.2464521>
- Slabchenko, O. O., and Sydorenko, V. M. "Analiz i syntez modelei na osnovi mashynnoho navchannia dlia imputatsii danykh iz personalnykh akauntiv sotsialnykh merezh" [Analysis and synthesis of models based on machine learning for mutaz data from personal social networking accounts]. *Visnyk KrNU imeni Mykhaila Ostrohradskoho*, no. 88 (2014): 105-111.
- "Yahoo Finance Official Web-site. AphabetInc. (GOOG)". <http://finance.yahoo.com/quote/GOOG?p=GOOG>