

УДК 004.8

Д. А. Ткачик, Р. Н. Кветний

АНАЛІЗ ЕМОЦІЙНОГО ЗАБАРВЛЕННЯ ТЕКСТУ ДЛЯ ПРОГНОЗУВАННЯ ДАНИХ НА ФІНАНСОВИХ РИНКАХ

Вінницький національний технічний університет, Вінниця

Анотація. Прогнозування даних на фінансових ринках є актуальним завданням в сучасному світі, оскільки змога передбачати напрямки руху ринку допомагає інвесторам уникати очевидних ризиків та позбутись додаткових фінансових витрат. Розроблено багато різних торгівельних платформ для швидкого отримання доступу до великих масивів історичних даних, що дозволяє аналізувати фінансовий ринок з будь якої точки планети та в реальному часі, використовуючи тільки ноутбук чи стаціонарний персональний комп'ютер. Такі платформи дозволяють розробляти спеціальні стратегії та підходи на основі фундаментального або технічного аналізу, які враховують новини про ту чи іншу компанію, її дохід, капіталізацію та суму дивідендів, які вона повинна вчасно виплачувати, а також різні технічні індикатори. Новини про ту чи іншу компанію допомагають визначити потенційному інвестору певні ризики, зокрема, кадрові, виробничі чи, найчастіше в сучасних реаліях, репутаційні. Тому при формуванні фундаментального аналізу велику роль грає аналіз текстів новин, і саме тому найбільш ефективно це можна зробити за допомогою використання нейронних мереж. За останнє десятиліття нейронні мережі завдяки технологічним інноваціям та розробкам займають важливе місце в аналізі великих масивів даних, зокрема текстових. Оскільки кожна новина про ту чи іншу компанію, яка є об'єктом для потенційного інвестора чи трейдера, має в собі певне емоційне забарвлення, наприклад, позитивне чи негативне, то його можна визначити за допомогою саме навченої нейронної мережі, що допоможе вносити коректні прогнози на фінансових ринках та розробляти ефективні стратегії. В комплексі з технічним аналізом розробка та дослідження такого підходу до прогнозування може давати точні результати. Саме тому наукове дослідження в цьому напрямку є актуальним. В даній науковій статті обґрунтовано актуальність розробки системи прогнозування даних на основі аналізу емоційного забарвлення текстів. Розроблено програмне забезпечення на основі мови програмування Python. Описано результати дослідження, а також представлено вихідні дані роботи розробленої програми з отриманою точністю аналізу. За результатами наукового дослідження зроблено висновки.

Ключові слова: прогнозування, дані, нейронні мережі, фінансові ринки, фондова біржа, Python, інформаційні технології, емоційне забарвлення тексту, заголовок, інтернет трейдинг, штучний інтелект.

Abstract. Data forecasting in financial markets is an urgent task in today's world, as the ability to predict the direction of market movement helps investors avoid obvious risks and get rid of additional financial costs. Many different trading platforms have been developed to quickly gain access to large amounts of historical data, which allows you to analyze the financial market from anywhere on the planet and in real time, using only a laptop or a stationary personal computer. Such platforms allow you to develop special strategies and approaches based on fundamental or technical analysis, which take into account news about a particular company, its income, capitalization and the amount of dividends it should pay on time, as well as various technical indicators. News about this or that company helps a potential investor identify certain risks, in particular, personnel, production or, most often in modern realities, reputational. Therefore, the analysis of news texts plays an important role in the formation of fundamental analysis, and that is why it can be done most effectively using neural networks. Over the last decade, thanks to technological innovations and developments, neural networks have taken an important place in the analysis of large data sets, in particular text. Since every piece of news about one or another company, which is an object for a potential investor or trader, has a certain emotional color, for example, positive or negative, it can be determined with the help of a specifically trained neural network, which will help make correct predictions on financial markets and develop effective strategies. In combination with technical analysis, the development and research of such a forecasting approach can produce accurate results. That is why scientific research in this direction is relevant. This scientific article substantiates the relevance of developing a data forecasting system based on the analysis of the emotional coloring of texts. The software was developed based on the Python programming language. The results of the research are described, as well as the output data of the developed program with the obtained accuracy of the analysis are presented. Based on the results of the scientific research, conclusions were drawn.

Key words: forecasting, data, neural networks, financial markets, stock exchange, Python, information technology, sentiment analysis, headline, internet trading, artificial intelligence.

DOI: <https://doi.org/10.31649/1999-9941-2022-55-3-51-58>.

Вступ

Процеси глобалізації інформаційної сфери та цифрової трансформації, розвиток інформаційного суспільства, впровадження передових інформаційно-комунікаційних технологій відкривають для трейдерів нові можливості для торгівлі цінними паперами. Використовуючи засоби Інтернету, учасники біржової торгівлі мають надшвидкий доступ до різних даних щодо ринку, що забезпечує кращу дохідність та ефективність.

З розвитком інтернет-трейдингу з'являються нові методи та підходи до прогнозування даних, які з кожним разом дають більшу точність та ефективність. Аналіз емоційного забарвлення тексту має безліч застосувань: наприклад такі дані дозволяють передбачити поведінку біржових трейдерів щодо конкретної компанії з відгуків в соціальних мережах чи при аналізі відповідних новин.

Аналіз емоційного забарвлення тексту означає визначення погляду чи емоції, що стоїть за ситуацією. В основному, це аналітика за фрагментом тексту чи промови або іншого способу спілкування. Кожний сектор економіки, з яким стикається клієнт (роздрібна торгівля, телекомунікації, фінанси тощо), зацікавлений у визначенні настроїв своїх клієнтів, незалежно від того, позитивні ці настрої чи ні.

Актуальність

Розробка систем для прогнозування даних є актуальною, так як в сучасному комп'ютеризованому світі значна кількість людей має доступ до інтернету, а також володіє персональним комп'ютером, що

забезпечує швидкий доступ до фондових бірж та дозволяє переглядати котирування цін в реальному часі, не виходячи з дому [1].

Завдяки розвитку нейронних мереж та підходів до обробки даних, стає можливим швидке та ефективно створення різних торгових стратегій та алгоритмів до прогнозування даних.

Аналіз емоційного забарвлення тексту може застосовуватися в будь-якій галузі, від фінансів і роздрібною торгівлі до сфери послуг та технологій. Нижче ми перерахували деякі з найпопулярніших способів використання аналізу настроїв у бізнесі [1]:

- 1) моніторинг соціальних мереж;
- 2) моніторинг бренду;
- 3) відгуки;
- 4) обслуговування клієнтів;
- 5) маркетингове дослідження.

Мета

Метою даної статті є розробка підходу для прогнозування даних на основі аналізу емоційного забарвлення тексту та приклад реалізації його з використанням мови Python. Впровадження таких засобів дозволить інтернет-трейдерам підвищити ефективність торговельних операцій на фондових біржах.

Результати досліджень

Аналіз емоційного забарвлення зосереджується на полярності тексту (позитивний, негативний, нейтральний), але він також виходить за межі полярності, щоб виявити конкретні почуття та емоції (злий, радісний, сумний тощо), терміновість (терміново, не терміново) і навіть наміри (зацікавлені, не зацікавлені) [2].

Залежно від того, як потрібно інтерпретувати ті чи інші новини, відгуки, коментарі, можна визначити та налаштувати свої категорії відповідно до потреб для аналізу тих чи інших настроїв [3]. Найпопулярніші підходи до аналізу емоційного забарвлення в тексті:

- поступовий аналіз настроїв;
- виявлення емоцій;
- аналіз настроїв на основі аспектів;
- багатомовний аналіз настроїв.

Якщо точність полярності важлива, то можна розглянути можливість розширення категорій полярності, щоб включити різні рівні як позитивних так негативних:

- дуже позитивно;
- позитивно;
- нейтрально;
- негативно;
- дуже негативно.

Це зазвичай називають градованим або поступовим аналізом настроїв, і його можна використовувати для інтерпретації 5-зіркових оцінок у огляді, наприклад, дуже позитивний це 5 зірок або дуже негативний це 1 зірка.

Виявлення емоцій дозволяє вийти за межі полярності, щоб виявити емоції, такі як щастя, розчарування, гнів і смуток.

Багато систем виявлення емоцій використовують лексикони (тобто списки слів і емоцій, які вони передають) або складні алгоритми машинного навчання.

Одним із недоліків використання лексиконів є те, що люди по-різному виражають емоції. Деякі слова, які зазвичай виражають гнів, як-от «погано» або «вбивати», також можуть виражати щастя.

Зазвичай, аналізуючи настрої текстів, потрібно знати, які конкретні аспекти чи особливості люди згадують у позитивному, нейтральному чи негативному ключі [4].

Ось де може допомогти аналіз настроїв на основі аспектів, наприклад, у цьому огляді продукту: «Час роботи батареї цього пристрою занадто малий», класифікатор на основі аспектів зможе визначити, що речення висловлює негативну думку щодо терміну служби акумулятора. відповідного продукту.

Багатомовний аналіз настроїв є більш складним. Це вимагає багато попередньої обробки даних та ресурсів. Більшість із цих ресурсів доступні в Інтернеті (наприклад, лексикони настроїв), тоді як інші потрібно створити, але потрібно знати, як кодувати, щоб їх використовувати.

Крім того, можна автоматично визначити мову в текстах за допомогою класифікатора мов, а потім навчити користувачку модель аналізу настроїв для класифікації текстів мовою за вибором.

Аналіз настроїв, інакше відомий як аналіз думок, працює завдяки обробці природної мови і алгоритмам машинного навчання, щоб автоматично визначити емоційний тон онлайн-розмов чи текстів.

Існують різні алгоритми, які можна застосувати в моделях аналізу настроїв, залежно від того, скільки даних потрібно проаналізувати, і наскільки точною очікується модель. Нижче ми розглянемо деякі з них більш детально.

Алгоритми аналізу настроїв поділяються на наступні три групи:

– на основі правил: ці системи автоматично виконують аналіз настроїв на основі набору правил, створених вручну;

– автоматичний: системи покладаються на методи машинного навчання;

– гібридні системи поєднують як засновані на правилах, так і автоматичні підходи.

Зазвичай система, на основі правил, використовує набір інструкцій, створених людиною, щоб допомогти визначити суб'єктивність, полярність або предмет певної точки зору чи позиції. Ці правила можуть включати різні методи, розроблені в комп'ютерній лінгвістиці, такі як:

- стемінг, лексема, тегування частини мови та синтаксичний аналіз;
- лексикони (тобто списки слів і виразів).

Наприклад як працює система на основі правил: визначає два списки поляризованих слів (наприклад, негативні слова, такі як поганий, найгірший, потворний тощо, та позитивні слова, такі як хороший, найкращий, красивий тощо), підраховує кількість позитивних і негативних слів, які зустрічаються в даному тексті. Якщо кількість позитивних слів більше, ніж кількість негативних слів, система повертає позитивний настрій, і навпаки. Якщо числа парні, система поверне нейтральні настрої.

Гібридні системи поєднують елементи методів, заснованих на правилах, і автоматичних методів в одну систему. Однією з переваг цих систем є те, що результати точніші [5].

На рисунку 1 зображено реалізацію класифікатора машинного навчання для аналізу емоційного забарвлення в тексті:

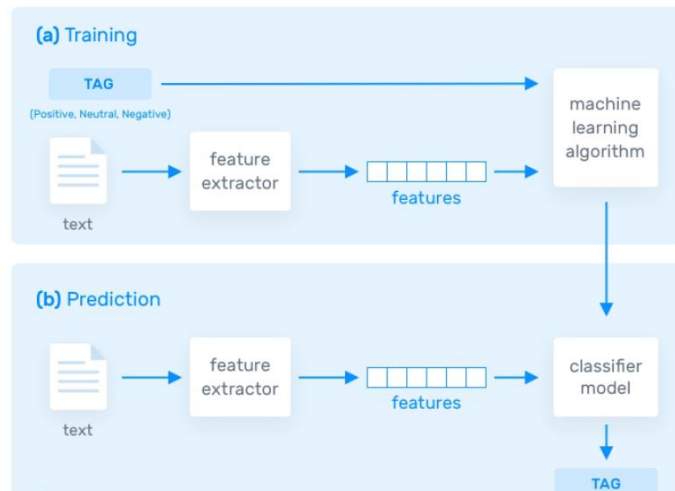


Рисунок 1 – Класифікатор машинного навчання

У процесі навчання (a) передається введений текст у вектор ознак. Пари векторів ознак і тегів (наприклад, позитивні, негативні або нейтральні) подаються в алгоритм машинного навчання для створення моделі.

У процесі прогнозування (b) ознаки використовуються для перетворення нерозпізнаних введених текстів у вектори ознак. Ці вектори ознак потім подаються в модель, яка генерує передбачені теги (негативні або нейтральні).

Першим кроком у класифікаторі тексту машинного навчання є перетворення вилученого тексту або його векторизації, і класичним підходом є bag-of-words або bag-of-ngrams [6].

Крок класифікації зазвичай включає статистичну модель, як

- наївний баєсів класифікатор;
- логістична регресія;
- опорні векторні машини;
- нейронні мережі.

Розробка системи для аналізу емоційного забарвлення тексту проводиться на основі мови програмування Python, а також бібліотеки vaderSentiment.

VADER (Valence Aware Dictionary and Sentiment Reasoner) — це інструмент аналізу настроїв на основі лексику та правил, який спеціально налаштований на емоції, виражені в соціальних мережах. Даний інструмент є повністю безкоштовним та відкритим, що значно спрощує його використання. Розроблений проект прогнозує, чи підвищиться або зменшиться ціна акцій компанії на основі заголовків головних новин.

Тестовий набір містить два датасети. На рисунку 3 зображено перший датасет – новини про фондовий ринок Доу-Джонса

	Date	Label	Top1	Top2	Top3	Top4	Top5
0	2016-07-01	0	b'Georgia 'downs two Russian warplanes' as cou...	b'BREAKING: Musharraf to be impeached.'	b'Russia Today: Columns of troops roll into So...	b'Russian tanks are moving towards the capital...	b'Afghan children raped with 'impunity,' U.N. ...
1	2016-06-30	1	b'Why wont America and Nato help us? If they w...	b'Bush puts foot down on Georgian conflict'	b'Jewish Georgian minister: Thanks to Israeli ...	b'Georgian army flees in disarray as Russians ...	b'Olympic opening ceremony fireworks 'faked''
2	2016-06-29	0	b'Remember that adorable 9-year-old who sang a...	b"Russia 'ends Georgia operation'"	b""If we had no sexual harassment we would hav...	b"Al-Qa'eda is losing support in Iraq because ...	b'Ceasefire in Georgia: Putin Outmaneuvers the...
3	2016-06-28	0	b' U.S. refuses Israel weapons to attack Iran:...	b"When the president ordered to attack Tskhinv...	b' Israel clears troops who killed Reuters cam...	b'Britain's policy of being tough on drugs is...	b'Body of 14 year old found in trunk; Latest (...
4	2016-06-27	1	b'All the experts admit that we should legalis...	b'War in South Osetia - 89 pictures made by a ...	b'Swedish wrestler Ara Abrahamian throws away ...	b'Russia exaggerated the death toll in South O...	b'Missile That Killed 9 Inside Pakistan May Ha...

Рисунок 3 – Дані нового заголовка про фондовий ринок Доу-Джонса

На рисунку 4 зображено другий датасет, що містить дані про ціну акцій на ринку Доу-Джонса.

```
df2= pd.read_csv('../input/stocknews/upload_DJIA_table.csv')
df2.head()
```

	Date	Open	High	Low	Close
0	2016-07-01	17924.240234	18002.380859	17916.910156	17949.369141
1	2016-06-30	17712.759766	17930.609375	17711.800781	17929.990234
2	2016-06-29	17456.019531	17704.509766	17456.019531	17694.679688
3	2016-06-28	17190.509766	17409.720703	17190.509766	17409.720703
4	2016-06-27	17355.210938	17355.210938	17063.080078	17140.240234

Рисунок 4 – Дані про ціну акцій на ринку Доу-Джонса

Потрібно об'єднати ці два датаести для бажаного набору даних. Результат об'єднання зображено на рисунку 5.

	Date	Label	Top1	Top2	Top3	Open	High	Low	Close
0	2016-07-01	0	b"Georgia 'downs two Russian warplanes' as cou...	b"BREAKING: Musharraf to be impeached."	b"Russia Today: Columns of troops roll into So...	11432.089844	11759.959961	11388.040039	11734.320312
1	2016-06-30	1	b'Why wont America and Nato help us? If they w...	b'Bush puts foot down on Georgian conflict'	b'Jewish Georgian minister: Thanks to Israeli ...	11729.669922	11867.110352	11675.530273	11782.349609
2	2016-06-29	0	b'Remember that adorable 9-year-old who sang a...	b"Russia 'ends Georgia operation'"	b""If we had no sexual harassment we would hav...	11781.700195	11782.349609	11601.519531	11642.469727

Рисунок 5 – Результат об'єднання двох вхідних датасетів

Далі з отриманого на рисунку 5 датасету потрібно видалити всі розриви та косі риси. Наступним важливим кроком є отримання суб'єктивності та полярності. Полярність означає силу думки [2]. Вона може бути позитивною або негативною. Якщо виникає з чимось пов'язане сильне позитивне почуття або емоція, наприклад, захоплення, довіра, любов; це справді матиме певну орієнтацію на всі інші аспекти існування цього об'єкта. Те ж саме стосується негативною полярності. Хорошим прикладом може бути такий: «Я не думаю, що куплю цей товар, тому що мій попередній досвід роботи з подібним товаром був не таким хорошим». Це матиме негативну полярність.

Сила позитивної і негативною полярностей може змінюватися в залежності від ситуації, але вони все ще вважаються позитивними або негативними. Користувачі зазвичай висловлюють свої емоції за допомогою таких слів, як «досить» або «трохи». Інструменти аналізу настроїв розглядатимуть їх як позитивні чи негативні. Найважливішим тут є те, що люди відчувають до чогось іншого через невербальне спілкування. Використання цих слів може вказувати на те, наскільки вони суб'єктивно оцінюють об'єкт [7].

Суб'єктивність означає ступінь особистої залученості людини в об'єкт. Тут найбільше значення мають особисті зв'язки та індивідуальний досвід з цим об'єктом, який може відрізнитися або не відрізнитися з точки зору когось іншого. Наприклад: «Я дуже задоволений своїм новим смартфоном, тому що він має найвищу продуктивність, доступну на ринку». Подібно до полярності, сильна суб'єктивність може бути негативною або позитивною. Твердження тут явно суб'єктивне, оскільки користувач насправді говорить про свій досвід і те, що він відчуває до об'єкта.

Інструменти аналізу настроїв повинні мати можливість розрізняти полярність і суб'єктивність, щоб правильно аналізувати думки користувачів. Думка може мати високий ступінь суб'єктивності, якщо вона висловлена як особистий досвід, тоді як низький ступінь може вказувати чужу точку зору на щось інше [8]. Настрої також можуть мати різний рівень полярності в різних способах спілкування – електронні листи, чати чи інші повідомлення можуть бути як позитивними, так і негативними; в деяких інших випадках вони можуть мати навіть нейтральну полярність.

Після визначення полярності та суб'єктивності слід провести оцінку на основі чотирьох критеріїв для кожного заголовку:

- негативний
- нейтральний
- позитивний
- комбінований (розраховано шляхом нормалізації наведених вище оцінок)

Після даних етапів отримуємо наступні дані, що зображені на рисунку 7:

	Label	Open	High	Low	Volume	Суб'єктивність	Полярність	Комбінований	Негативний	Позитивний	Нейтральний
0	0	11432.089844	11759.959961	11388.040039	212830000	0.267549	-0.048568	-0.9982	0.235	0.041	0.724
1	1	11729.669922	11867.110352	11675.530273	183190000	0.374806	0.121956	-0.9858	0.191	0.089	0.721
2	0	11781.700195	11782.349609	11601.519531	173590000	0.536234	-0.044302	-0.9715	0.128	0.056	0.816
3	0	11632.809570	11633.780273	11453.339844	182550000	0.364021	0.011398	-0.9809	0.146	0.066	0.788
4	1	11532.070312	11718.280273	11450.889648	159790000	0.375099	0.040677	-0.9882	0.189	0.094	0.717
...
1984	0	17355.210938	17355.210938	17063.080078	138740000	0.352311	-0.014015	-0.9644	0.148	0.094	0.758
1985	1	17190.509766	17409.720703	17190.509766	112190000	0.352649	0.046560	-0.9571	0.132	0.102	0.767
1986	1	17456.019531	17704.509766	17456.019531	106380000	0.389617	0.052622	-0.9975	0.225	0.091	0.684
1987	1	17712.759766	17930.609375	17711.800781	133030000	0.382566	0.011243	-0.9977	0.202	0.061	0.738
1988	1	17924.240234	18002.380859	17916.910156	82160000	0.320261	-0.035458	-0.9983	0.212	0.059	0.729

Рисунок 7 – Дані для створення та навчання моделі

За отриманими даними, що зображено на рисунку 7, створюємо модель для навчання, як показано на рисунку 8.

```
model= LinearDiscriminantAnalysis().fit(x_train, y_train)

predict= model.predict(x_test)
```

Рисунок 8 – Створення моделі для навчання

Після аналізу даних з допомогою інструментів vaderSentiment отримуємо підсумкову таблицю результатів навчання згенерованої моделі даних (рисунк 9).

	precision	recall	f1-score	support
0	0.86	0.79	0.83	193
1	0.82	0.88	0.85	205
accuracy			0.84	398
macro avg	0.84	0.84	0.84	398
weighted avg	0.84	0.84	0.84	398

Рисунок 9 – Результати роботи програми аналізу емоційного забарвлення тексту

Точність є одним з показників для оцінки моделей класифікації. Неформально, точність – це частка прогнозів, які модель отримала правильно. Формально точність має таке визначення:

$$\text{Точність} = \frac{\text{Кількість правильних прогнозів}}{\text{Кількість усіх прогнозів}} \quad (1)$$

Для двійкової класифікації точність також можна розрахувати в термінах позитивних і негативних показників таким чином:

$$\text{Точність} = \frac{TP+TN}{TP+TN+FP+FN}, \quad (2)$$

де TP = правильно позитивні результати, TN = правильно негативні результати, FP = помилкові позитивні результати та FN = помилкові негативні результати.

Проаналізувавши отримані результати, які зображені на рисунку 9, можна зробити висновок, що точність становить 0,84, або 84% (84 правильних прогнозів із 100 прикладів).

Висновки

Розробка систем для прогнозування даних є актуальною, так як в сучасному комп'ютеризованому світі значна кількість людей має доступ до інтернету, а також володіє персональним комп'ютером, що забезпечує швидкий доступ до фондових бірж та дозволяє переглядати котирування цін в реальному часі, не виходячи з дому [6].

Такі системи, як представлена у статті, призначені для полегшення аналізу ринку інвестором, вони допомагають фільтрувати ті чи інші ринки на фондових ринках для концентрації уваги на такі, що можуть принести більший і швидший прибуток.

Підхід, використаний в системі прогнозування даних, а саме аналіз емоційного забарвлення тексту, дозволяє підвищити ефективність прийняття рішень на фінансових біржах, зокрема, його можна використовувати в комплексі з іншими системами. Наприклад, система для прийняття рішень, яка базується на прогнозуванні історичних даних тільки за допомогою технічного аналізу, дає результат прогнозування приблизно 70%, як це зображено на рисунку 10 [9].

```
Total BUY positions: 505
Profit rate: 370 ( 73.26 % )
Stoploss rate: 134 ( 26.53 % )
```

Рисунок 10 – Результат роботи системи для підтримки прийняття рішень, що базується на технічному аналізі

Враховуючи точність описаного в статті підходу, результат роботи системи можна покращити щонайменше до 80%, за рахунок уникнення помилкових вхідних даних, навчання нейронної мережі та відбору більш стабільних фінансових об'єктів (компаній, біржових індексів і т.д.).

Даний підхід також може бути вдосконалений: зокрема, потрібно аналізувати не лише заголовки новин, а й самі новини. Також потрібно звернути увагу на критерії для поділу тексту: використовувати тільки позитивний та негативний, щоб уникати сумнівних прогнозів та покращити продуктивність.

Список літератури

- [1] F. Hamborg, K. Donnay, "NewsMTSC: A Dataset for (Multi-)Target-dependent Sentiment Classification in Political News Articles," Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Stroudsburg, PA, USA, 2021. [Online]. Available: <https://doi.org/10.18653/v1/2021.eacl-main.142>.
- [2] What is polarity and subjectivity in sentiment analysis? [Online]. Available: <https://www.quora.com/What-is-polarity-and-subjectivity-in-sentiment-analysis>.
- [3] Sentimental analysis using vader. [Online]. Available: <https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664>.
- [4] Santur Y. Sentiment Analysis Based on Gated Recurrent Unit. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 21–22 September 2019. 2019. [Online]. Available: <https://doi.org/10.1109/idap.2019.8875985>.
- [5] K. V. Nguyen et al., "UIT-VSFC: Vietnamese Students' Feedback Corpus for Sentiment Analysis," 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, 1–3 November 2018. 2018. [Online]. Available: <https://doi.org/10.1109/kse.2018.8573337>.
- [6] Д. Ткачик, О. Захарчук, "Розпізнавання емоційного забарвлення тексту для фундаментального аналізу на фінансових ринках," *Збірник наукових праць SCIENTIA, Валенсія, Королівство Іспанія: Європейська наукова платформа*. 2021.
- [7] Ł. Augustyniak et al., "Comprehensive Study on Lexicon-based Ensemble Classification Sentiment Analysis," *Entropy*, vol. 18, no. 1, p. 4. 2015. [Online]. Available: <https://doi.org/10.3390/e18010004>.
- [8] B. Liang et al., "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, p. 107643. 2022. [Online]. Available: <https://doi.org/10.1016/j.knosys.2021.107643>.
- [9] Д. А. Ткачик, Р. Н. Кветний, "Розробка ефективних комбінацій моделей технічного аналізу для прогнозування ринку," *Матеріали XLIX науково-технічної конференції підрозділів ВНТУ, Вінниця, 27-28 квітня 2020 р.* 2020. [Електронний ресурс]. Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2020/paper/view/9600>.

Стаття надійшла: 18.09.2022.

References

- [1] F. Hamborg, K. Donnay, "NewsMTSC: A Dataset for (Multi-)Target-dependent Sentiment Classification in Political News Articles," Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Stroudsburg, PA, USA, 2021. [Online]. Available: <https://doi.org/10.18653/v1/2021.eacl-main.142>.
- [2] What is polarity and subjectivity in sentiment analysis? [Online]. Available: <https://www.quora.com/What-is-polarity-and-subjectivity-in-sentiment-analysis>.
- [3] Sentimental analysis using vader. [Online]. Available: <https://towardsdatascience.com/sentimental-analysis-using-vader-a3415fef7664>.
- [4] Santur Y. Sentiment Analysis Based on Gated Recurrent Unit. 2019 International Artificial Intelligence and Data Processing Symposium (IDAP), Malatya, Turkey, 21–22 September 2019. 2019. [Online]. Available: <https://doi.org/10.1109/idap.2019.8875985>.
- [5] K. V. Nguyen et al., "UIT-VSFC: Vietnamese Students' Feedback Corpus for Sentiment Analysis," 10th International Conference on Knowledge and Systems Engineering (KSE), Ho Chi Minh City, 1–3 November 2018. 2018. [Online]. Available: <https://doi.org/10.1109/kse.2018.8573337>.
- [6] Tkachyk, O. Zakharchuk, "Rozpiznavannia emotsiinoho zabarvlennia tekstu dlia fundamentalnogo analizu na finansovykh rynkakh," *Zbirnyk naukovykh prats SCIENTIA, Valensiia, Korolivstvo Ispaniia: Yevropeiska naukova platforma*. 2021 [in Ukrainian].
- [7] Ł. Augustyniak et al., "Comprehensive Study on Lexicon-based Ensemble Classification Sentiment Analysis," *Entropy*, vol. 18, no. 1, p. 4. 2015. [Online]. Available: <https://doi.org/10.3390/e18010004>.
- [8] B. Liang et al., "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, p. 107643. 2022. [Online]. Available: <https://doi.org/10.1016/j.knosys.2021.107643>.
- [1] D. A. Tkachyk, R. N. Kvietyi, "Rozrobka efektyvnykh kombinatsii modelei tekhnichnogo analizu dlia prohnouzuvannia rynku," *Materialy XLIX nauково-tekhnichnoi konferentsii pidrozdiliv VNTU*,

*Vinnytsia, 27-28 kvitnia 2020 r. 2020. [Online]. Available:
<https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2020/paper/view/9600> [in Ukrainian].*

Відомості про авторів

Ткачик Денис Анатолійович – аспірант кафедри АІТ, факультет інтелектуальних інформаційних технологій та автоматизації.

Квєтний Роман Наумович – доктор технічних наук, професор.

D. A. Tkachyk, R. N. Kvyetnyy

**SENTIMENT ANALYSIS OF THE TEXT FOR FORECASTING
DATA IN FINANCIAL MARKETS**

Vinnytsia National Technical University, Vinnytsia