

# ВИКОРИСТАННЯ ГРАДУЙОВАНОГО АНАЛІЗУ ГРОМАДСЬКИХ НАСТРОЇВ У СИСТЕМІ ПРОГНОЗУВАННЯ ДАНИХ

<sup>1</sup>Вінницький національний технічний університет

*Прогнозування даних на фінансових ринках — актуальне завдання у сучасному світі. Здатність передбачити напрямки руху ринку допомагає інвесторам уникнути очевидних ризиків і позбавити себе додаткових витрат. Розроблено багато різних торгових платформ, щоб швидко отримувати доступ до великих обсягів історичних даних, що дозволяє аналізувати фінансовий ринок з будь-якого куточка планети в режимі реального часу, використовуючи лише ноутбук або персональний комп'ютер. Такі платформи дозволяють розробляти унікальні стратегії та підходи на основі фундаментального або технічного аналізу, які враховують новини про певну компанію, її прибуток, капіталізацію та кількість дивідендів, які вона повинна виплачувати вчасно.*

*Новини про різні компанії допомагають потенційному інвестору виявляти певні ризики, зокрема, персоналу виробництва або, найчастіше в сучасних реаліях, репутаційні. Тому аналіз новинних текстів відіграє важливу роль у формуванні фундаментального аналізу, і саме тому його можна проводити найефективніше за допомогою нейромереж.*

*Аналіз настроїв за допомогою нейронних мереж є потужним інструментом для прогнозування ринків в сучасному бізнесі, оскільки він дозволяє аналізувати та розуміти глибокі відчуття та настрої споживачів та інвесторів на основі текстової інформації, такої як відеореєстри, соціальні медіа, новини та ін.*

*Аналіз настроїв використовує машинне навчання та природну мову, щоб автоматично визначити та категоризувати текстові дані як позитивні, негативні чи нейтральні. Ця інформація може бути використана для прогнозування реакції ринку на нові продукти, послуги, політичні події чи інші чинники. Такий аналіз допомагає підприємствам зрозуміти глибоке ставлення споживачів та інвесторів до їхнього бренду, товарів та послуг, що дозволяє розробляти ефективніші маркетингові стратегії та ухвалювати обґрунтовані рішення щодо розвитку бізнесу.*

*Використання градуйованого аналізу настроїв допоможе отримати ліпшу оцінку тому чи іншому тексту, який описує ринок або надає про нього певну інформацію, що допоможе інвестору ухвалити найменш ризиковане рішення. Імплементція цього методу сентиментального аналізу допоможе підвищити ефективність системи прогнозування даних, яка вже використовує технічний аналіз на фондових ринках.*

*Метою роботи є підвищення ефективності автоматизованого технічного аналізу даних на фондових ринках завдяки імплементції методів аналізу настроїв.*

**Ключові слова:** історичні дані, аналіз настроїв, фінансові ринки, градуйований аналіз настроїв, фундаментальний аналіз, система прогнозування даних.

## Вступ

Прогнозування на фондовому ринку є активною сферою досліджень вже тривалий час. Гіпотеза про ефективність ринку стверджує, що ціни на фондовому ринку значною мірою визначаються новою інформацією та випадково. Незважаючи на те, що ця гіпотеза загалом широко визнана дослідницькою спільнотою як центральна парадигма, що керує ринками, люди намагалися встановити закономірності в поведінці фондових ринків та реакції на зовнішні подразники.

Здійснення точних прогнозів щодо цін на акції є складною задачею, оскільки найкращий час для інвестування або утримання акцій залежить від таких факторів, як процентні ставки, інфляція, ризикованість інвесторів, зміни в пропозиції та попиту або щоквартальні фінансові звіти. Навіть якщо багато чинників визначають ціни на акції на ринку, психологічні чинники, такі як настрої користувачів щодо змін у політиці, нових інвестицій чи природних катастроф, також сильно впли-

вають на зміну цін на акції [1].

Зокрема, фундаментальний аналіз відіграє важливу роль у визначенні інвестиційної стратегії та ризиків. Цей метод аналізу фокусується на оцінці фундаментальних аспектів компаній, таких як фінансовий здоров'я, прибутковість та перспективи ринку. Враховуючи ці фактори, інвестори можуть приймати обґрунтованіші рішення та адаптувати свої стратегії до змін у фондовому середовищі.

Зміни в цінах на акції зумовлюють не лише фінансові фактори, а й те, як загальна громадська думка стосується компаній на ринку, репутація бренду серед клієнтів тощо.

Аналіз настроїв на фондовому ринку — один з методів веб-скрапінгу для фінансів, що допомагає збирати дані для ухвалення обґрунтованих бізнес-рішень. Дослідження показують, що рухи цін на фондовому ринку корелюють із громадськими настроями щодо компаній. Коли настрої користувачів враховуються у прогнозах цін, точність моделей прогнозу зростає на 20 %, що демонструє додаткову цінність настроїв клієнтів у прогнозуванні цін [2].

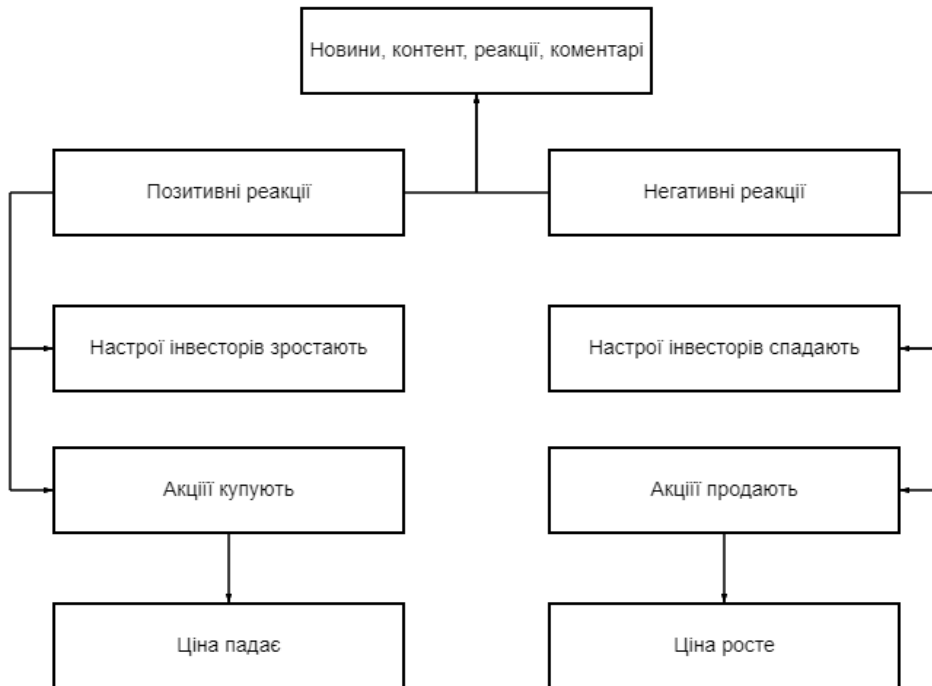


Рис. 1. Вплив громадських настроїв на ціни акцій на ринку

*Метою роботи* є підвищення ефективності автоматизованого технічного аналізу даних на фондових ринках завдяки імплементації методів аналізу настроїв.

### Результати досліджень

Сучасні підходи до аналізу настроїв поділяються на три категорії: бази знань, статистичні та гібридні. Аналіз настроїв виконується таким чином:

– на основі баз знань — цей підхід включає класифікацію тексту на основі слів, які виражають емоції.

– статистичний: цей підхід використовує алгоритми машинного навчання, такі як латентний семантичний аналіз і глибоке навчання для точного виявлення настрою.

– гібридний: цей підхід використовує як бази знань, так і статистичні техніки для точного аналізу настроїв.

Раніше для проведення аналізу настроїв потрібно було мати технічні знання в технологіях, таких як Python, R та машинне навчання. Але сьогодні існують декілька програмних засобів, які дозволяють проводити аналіз настроїв без або з мінімальним обсягом технічних знань.

Аналіз настроїв акцентує на полярності тексту (позитивний, негативний, нейтральний), але також виходить за межі полярності для виявлення конкретних почуттів і емоцій (сердитий, щасливий, сумний і т. д.), термінів (терміновий, не терміновий) і навіть намірів (зацікавлений чи не зацікавлений).

Залежно від того, як потрібно інтерпретувати відгуки та запити клієнтів або деякі новини, ви



Бінарні мітки класів можуть бути достатніми для вивчення загальних тенденцій у вираженні позитивного/негативного настрою в текстових даних, таких як твіти, відгуки на продукти чи відгуки клієнтів, але вони мають свої обмеження [7]. Під час виконання скрапінгу інформації з порівняльними висловами, наприклад, «Ця річ набагато краща за іншу» — детальний аналіз може надавати точніші результати [7].

Перший крок у текстовому класифікаторі машинного навчання — це перетворення тексту або векторизація тексту, і класичним підходом є метод «мішок слів» або «мішок  $n$ -грам» з їхньою частотою.

Зазвичай крок класифікації включає в себе статистичну модель, таку як [8]:

- наївний баєсів класифікатор;
- машини опорних векторів;
- логістична регресія;
- нейронні мережі.

Розробка системи для аналізу емоційного забарвлення тексту базується на мові програмування Python, а також на бібліотеці vaderSentiment [9].

VADER (Valence Aware Dictionary and Sentiment Reasoner) — це лексична база даних і інструмент для аналізу настроїв на основі правил, який оптимізований для аналізу настроїв у соціальних мережах. Він використовує різноманітні техніки. Лексичний словник настрою — це колекція лексичних ознак (наприклад, слів), які класифікуються як позитивні або негативні в залежності від їхньої полярності настрою. Він відображає не лише оцінки позитивності та негативності, але й ступінь позитивного або негативного настрою.

Вибірка включає два набори даних: *df1* — це дані про заголовки новин про біржу Dow Jones, і *df2* містить дані про ціни на акції на біржі Dow Jones. Ми об'єднаємо ці два набори даних для комплексного аналізу [9].

На рис. 3 показано перший набір даних — новини про біржу Dow Jones.

	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...

Рис. 3. Новини про біржу Dow Jones

Рис. 4 показує другий набір даних, що містить інформацію про ціни на акції на ринку Dow Jones.

```
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

Рис. 4. Ціни на акції на ринку Dow Jones

Об'єднаємо ці два набори даних для створення комплексного зразка та отримуємо результат, показаний на рис. 5.

Наступним важливим кроком є отримання показників суб'єктивності та полярності.

Полярність означає силу думки. Вона може бути позитивною або негативною. Якщо до чогось додається сильне позитивне почуття або емоція, така як захоплення, довіра, любов, то вона, безперечно, буде мати вплив на всі інші аспекти існування об'єкта. Те саме стосується негативної полярності.

Date	Label	Top1	Open	High	Low	Close
2008-08-08	0	b'Georgia 'downs two Russian warplanes' as cou...	11432.089844	11759.959961	11388.040039	11734.32031

Рис. 5. Результат об'єднання двох наборів даних

Сила позитивної та негативної полярностей може варіюватися в залежності від ситуації, але їх все ж розглядають як позитивні чи негативні. Користувачі зазвичай виражають свої емоції, використовуючи слова, як «дуже» чи «трохи». Засоби аналізу настроїв розглядатимуть їх як позитивні або негативні. Тут найважливіше, як люди відчують щось через невербальну комунікацію. Використання цих слів може вказувати на те, наскільки суб'єктивно вони оцінюють об'єкт [2].

Суб'єктивність означає ступінь особистого включення особи в об'єкт. Тут найважливіше особисті зв'язки та індивідуальні досвіди з цим об'єктом, які можуть бути або не бути відмінними від погляду іншої людини. Як і полярність, сильна суб'єктивність може бути негативною або позитивною. Тут заява, вочевидь, є суб'єктивною, оскільки користувач фактично розповідає про свій досвід і те, як він реагує на об'єкт [1].

Засоби аналізу настроїв повинні вміти розрізнити полярність і суб'єктивність для коректного аналізу думок користувачів. Думка може мати великий ступінь суб'єктивності, якщо вона виражена як особистий досвід, в той час як низька ступінь може вказувати на точку зору іншої людини щодо чого-небудь іншого. Настрої можуть також мати різні рівні полярності в різних способах комунікації — електронні листи, чати або інші повідомлення можуть бути як позитивними, так і негативними; в інших випадках вони можуть навіть мати нейтральну полярність.

Після визначення полярності і суб'єктивності слід здійснити оцінку на основі чотирьох критеріїв для кожного заголовка: негативний; нейтральний; позитивний; комбінований, — обчислений шляхом нормалізації вищезазначених показників.

На наступному етапі можна отримати дані, показані на рис. 6.

Low	Volume	Суб'єктивність	Полярність	Комбінований	Негативний	Позитивний	Нейтральний
11388.040039	212830000	0.267549	-0.048568	-0.9982	0.235	0.041	0.724
11675.530273	183190000	0.374806	0.121956	-0.9858	0.191	0.089	0.721
11601.519531	173590000	0.536234	-0.044302	-0.9715	0.128	0.056	0.816
11453.339844	182550000	0.364021	0.011398	-0.9809	0.146	0.066	0.788
11450.889648	159790000	0.375099	0.040677	-0.9882	0.189	0.094	0.717

Рис. 6. Дані для створення та навчання моделі

На основі отриманих даних, показаних на рис. 6, створюємо модель для навчання та виконуємо прогнозування, як показано на рис. 7.

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

Рис. 7. Створення моделі для тренування

	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

Рис. 8. Результати програми для аналізу емоційного забарвлення тексту

Після аналізу даних за допомогою інструментів vaderSentiment, отримуємо зведену таблицю результатів навчання створеної моделі даних (рис. 8). З рисунку видно кількість відкритих позицій і точність виконання цілої системи, яка використовує градуований аналіз даних.

Алгоритм роботи програми зображений на рис. 9.

Точність є одним з показників для

оцінювання моделей класифікації. Неформально, точність — це частка прогнозів, які модель правильно визначила. Формально точність має таке визначення:



Рис. 9. Алгоритм роботи програми

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

Для бінарної класифікації точність також може бути обчислена за допомогою критерію  $k$  щодо позитивних і негативних рейтингів

$$k = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

де  $TP$  — правильно визначені позитивні результати,  $TN$  — правильно визначені негативні результати,  $FP$  — неправильно визначені позитивні результати і  $FN$  — неправильно визначені негативні результати.

Аналізуючи отримані результати, показані на рис. 8, можна підсумувати, що точність становить 0,84 або 84 % (84 правильних прогнозів зі 100 прикладів).

### Висновки

Розвиток систем прогнозування даних є актуальним, оскільки в сучасному комп'ютеризованому світі значна кількість людей має доступ до Інтернету і володіє особистим комп'ютером, що надає швидкий доступ до фондових бірж та дозволяє переглядати котирування в режимі реального часу, не залишаючи дім.

Градуїований аналіз настроїв дає можливість розширення набору категорій полярності для включення різних рівнів позитивного та негативного. Тексту присвоюється оцінка на певній шкалі, що дозволяє проводити набагато коректніший аналіз. Наприклад, на шкалі від 1 до 10, 1 може означати дуже негативний настрій, а 10 — дуже позитивний.

Такий метод аналізу настроїв разом з технічним аналізом історичних даних допоможе підвищити точність прогнозування на фондових ринках, фільтруючи ризикованіші та неволатильні дані, що дозволить ефективніше ухвалювати рішення.

Використання градуїованого аналізу в автоматизованих системах прогнозування даних збільшить кількість прибуткових угод та дозволить автоматично заздалегідь визначати тренди тих чи інших учасників фондового ринку.

### СПИСОК ВИКОРИСТАНОЇ ЛІТЕРАТУРИ

- [1] Д. А. Ткачик, і Р. Н. Кветний, «Розробка ефективних комбінацій моделей технічного аналізу для прогнозування ринку», *Матеріали XLIX Науково-технічної конференції ВНТУ*, Вінниця, 27-28 квітня 2020 р., [Електронний ресурс]. Режим доступу: <https://conferences.vntu.edu.ua/index.php/all-fksa/all-fksa-2020/paper/view/9600>.
- [2] L. Zhang, S. Wang, and B. Liu, "Deep learning for sentiment analysis," *A survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253, 2018.
- [3] *A review on sentiment analysis and emotion detection from text*, 2021. [Electronic resource]. Available: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8402961/>.
- [4] R. Prabowo, and M. Thelwall, "Sentiment analysis: A combined approach," *Journal of Informetrics*, no. 3(2), pp. 143-157, 2009.
- [5] F. Hamborg, and 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]. <https://doi.org/10.18653/v1/2021.eacl-main.142>.
- [6] *Stock Market Sentiment Analysis in 2023*, 2022. [Electronic resource]. Available: <https://research.aimultiple.com/sentiment-analysis-stock-market/>.
- [7] *Types of Sentiment Analysis and How Brands Perform Them*, 2020. [Electronic resource]. Available: <https://www.analyticsinsight.net/types-of-sentiment-analysis-and-how-brands-perform-them/>.
- [8] B. Liang et al., "Aspect-based sentiment analysis via affective knowledge enhanced graph convolutional networks," *Knowledge-Based Systems*, vol. 235, pp. 107643, 2022. [Electronic resource]. <https://doi.org/10.1016/j.knsys.2021.107643>.
- [9] M. Birjali, M. Kasri, and A. Beni-Hssane, "A comprehensive survey on sentiment analysis: Approaches, challenges and trends," *Knowledge-Based Systems*, 226, 107134, 2021.
- [10] M. Wongkar, and A. Angdresey, "Sentiment analysis using Naive Bayes Algorithm of the data crawler," *IEEE Twitter. In 2019 Fourth International Conference on Informatics and Computing (ICIC)*, pp. 1-5, October, 2019.
- [11] *VADER (Valence Aware Dictionary and sentiment Reasoner) Sentiment Analysis*, 2022. [Electronic resource]. Available: <https://medium.com/mllearning-ai/vader-valence-aware-dictionary-and-sentiment-reasoner-sentiment-analysis-28251536698>.

**Ткачик Денис Анатолійович** — аспірант кафедри автоматизації та інтелектуальних інформаційних технологій, e-mail: true.tkachyk@gmail.com .

Вінницький національний технічний університет, Вінниця

**D. A. Tkachyk<sup>1</sup>**

## **Using of Public Sentiment Gradient Analysis in Data Forecasting System**

<sup>1</sup>Vinnitsia National Technical University

*Data forecasting data in financial markets is a critical task in today's world. The ability to predict the market movements helps the investors avoid obvious risks and reduce potential losses. Various trading platforms have been developed to provide quick access to vast volumes of historical data, allowing real-time market analysis from anywhere in the world using just a laptop or personal computer. These platforms enable the creation of unique strategies based on fundamental or technical analysis, which take into account news about the specific companies, their profits, capitalization, and dividend payments.*

*News about different companies helps potential investors identify various risks, including personnel, production, and most commonly in modern times, reputation risks. Therefore, textual news analysis plays a crucial role in forming fundamental analysis, which is most effectively conducted using neural networks.*

*Sentimental analysis by means of neural networks is a powerful tool for the markets forecasting as it enables to analyze and understand deep sentiments and emotions of the consumers and investors, based on textual information, such as feedback, social media, news, etc.*

*Over the past decade, due to technological innovations and advances in neural networks, these networks have become instrumental in analyzing large datasets, including textual data. As each news piece about a company targeted by potential investors or traders carries emotional sentiment, such as positive or negative, this sentiment can be determined using specially trained neural networks. This enables making accurate predictions in financial markets and developing effective strategies. When combined with technical analysis, the development and investigation of such an approach to forecasting can yield precise results. Hence, scientific research in this field remains relevant.*

*This article substantiates the sentiment analysis approach for forecasting historical data in financial markets, describes similar approaches, and outlines their advantages and drawbacks. Solutions are provided for a data forecasting system using selected sentiment analysis methods.*

**Keywords:** historical data, sentiment analysis, financial markets, sentiment gradient analysis, fundamental analysis.

**Tkachyk Denys A.** — Post-Graduate Student of the Chair of Automation and Intelligent Information Technologies, e-mail: true.tkachyk@gmail.com