

УДК 519.688

Канд. техн. наук доцент Сирота С. В., студент Байда О. Є.

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»

МАТЕМАТИЧНЕ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ОПТИМІЗАЦІЇ ІНВЕСТИЦІЙНОГО ПОРТФЕЛЯ

Abstract

Sergiy Syrota, assoc. prof., PhD; Oleksii Baida student
Mathematical and software framework for stock classification using transformer-based news sentiment analysis and automated portfolio construction.

This paper presents an integrated approach to analyzing financial markets by combining time series of stock prices and textual information from financial news. The developed system includes modules for data collection via the NewsAPI and Yahoo Finance interfaces, sentiment scoring of textual news using the FinBERT transformer model, and subsequent clustering of stocks by sentiment and price dynamics. The main objective is to improve portfolio diversification through sentiment-informed grouping of assets and to evaluate the correlation between news tone and short-term market movements.

Вступ

Останні досягнення в галузі обробки природної мови (NLP) відкрили можливість застосування трансформерних нейронних мереж, таких як FinBERT [1], спеціально натренованих на фінансових текстах, для автоматичного визначення позитивних, негативних або нейтральних настроїв новинних повідомлень. Це дозволяє перетворити неструктурований текст у кількісні показники, придатні для подальшого статистичного аналізу.

Постановка задачі

Мета роботи – розробити математичне та програмне забезпечення для автоматизованої класифікації акцій компаній за сукупністю фінансових та текстових характеристик, із подальшою побудовою оптимізованого інвестиційного портфеля.

Основні завдання:

1. Реалізувати збір новинних даних про компанії через NewsAPI
2. Завантажити часові ряди акцій через ufinance API
3. Виконати аналіз тональності текстів за допомогою моделі FinBERT

4. Об'єднати новинні та фінансові ознаки
5. Провести кореляційний аналіз між настроєм новин і ретурнами (доходністю) акцій.

Архітектура системи

Розроблена система має модульну архітектуру, що складається з семи основних компонентів, які взаємодіють між собою відповідно до UML-діаграми компонентів (рис. 1).

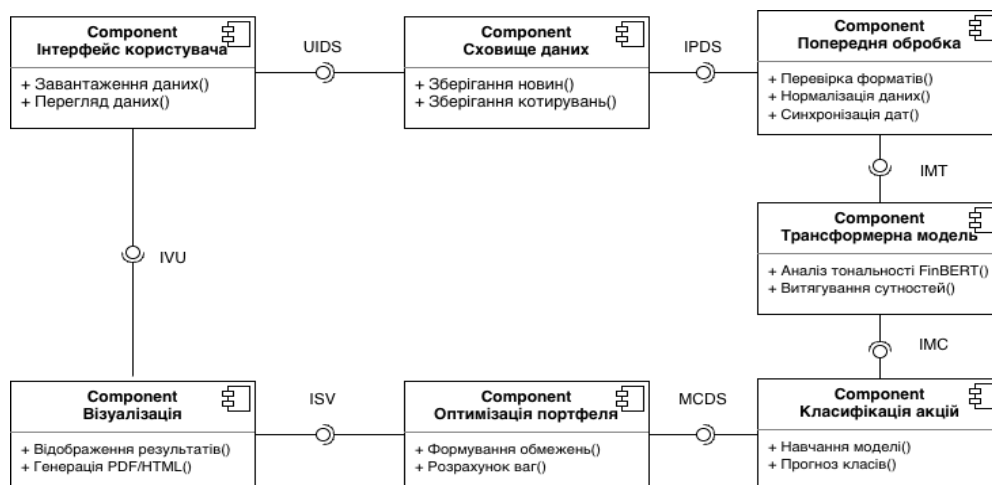


Рис. 1. UML-діаграма компонентів програмного забезпечення для класифікації акцій

Взаємодія компонентів організована за принципом послідовного потоку даних: від завантаження через API-інтерфейси до фінальної оптимізації та візуалізації.

Результати та аналіз

На даному етапі дослідження було реалізовано повний цикл збору, обробки та інтеграції текстових і фінансових даних із двох незалежних джерел — NewsAPI [2] та Yahoo Finance [3]. Кожна новина, отримана через API, класифікувалася моделлю FinBERT за трьома категоріями емоційного забарвлення — позитивна, нейтральна та негативна (p_{pos} , p_{neg} , p_{neu}), які відображали відповідний клас настрою. На основі цих показників розраховувався інтегральний індекс настрою за формулою:

$$score = p_{pos} - p_{neg}$$

Додатні значення $score$ свідчать про позитивний настрій новини, від'ємні — про негативний. Для кожного дня t та компанії i визначався середній добовий показник настрою як середнє значення усіх оцінок за цей день

Паралельно з новинними даними з Yahoo Finance були отримані котирування акцій відповідних компаній. Для кожного дня були витягнуті майбутні дохідності (ретурни) за різних горизонтів прогнозу $h = 1, 2, \dots, 15$.

$$r_{i,t+h} = \frac{P_{i,t+h} - P_{i,t}}{P_{i,t}}$$

де $P_{i,t}$ — ціна закриття акції компанії i у день t . Таким чином, для кожного дня було сформовано узгоджені пари значень, що дозволило аналізувати взаємозв'язок між поточним настроєм новин та майбутньою зміною вартості акцій. Для кількісної оцінки сили цього зв'язку застосовувалися три метрики кореляції. Перша з них — коефіцієнт Пірсона, який вимірює лінійну залежність між двома змінними та розраховується за формулою

$$r_p = \frac{cov(x, y)}{\sigma_x \sigma_y}$$

де $x = \text{score}$, $y = r_{t+h}$, а σ_x і σ_y — стандартні відхилення відповідних величин. Друга метрика — коефіцієнт Спірмена, що базується на ранговому перетворенні даних і є менш чутливим до викидів та нелінійних залежностей. Його аналітичне вираження має вигляд

$$r_s = 1 - \frac{6 \sum d_i^2}{n(n^2 - 1)}$$

де d_i — різниця між рангами кожної пари спостережень, а n — кількість спостережень. Третій показник — детрендована кореляція, що вимірює силу зв'язку після усунення впливу загального ринкового тренду. Для цього обидва ряди апроксимуються лінійною функцією часу, і з них віднімаються отримані тренди:

$$r_d = \text{corr}(x - \hat{x}, y - \hat{y})$$

Кореляційний аналіз проведено на часовому інтервалі з 1 по 20 жовтня 2025 року, для вибірки, що містила близько 1000 новин для 7 компаній. Результати розрахунків подано в таблиці 1.

Таблиця 1

Коефіцієнти кореляції між середнім настроєм та майбутніми рентабельністю інвестиції (ROI) акцій:

Горизонт (днів)	Pearson r	p-value	Spearman r	p-value	Detrended r	p-value	n
1	0.189	0.117	0.257	0.032	0.189	0.116	70
2	0.200	0.097	0.239	0.046	0.203	0.091	70
3	0.202	0.093	0.217	0.072	0.208	0.085	70
4	0.127	0.296	0.164	0.175	0.132	0.275	70
5	0.059	0.625	0.080	0.512	0.066	0.587	70
6	0.206	0.100	0.208	0.097	0.213	0.088	65

7	0.198	0.129	0.212	0.104	0.206	0.115	60
8	0.286	0.034	0.278	0.040	0.293	0.030	55
9	0.349	0.013	0.340	0.016	0.353	0.012	50
10	0.358	0.016	0.263	0.080	0.365	0.014	45
11	0.288	0.072	0.184	0.255	0.289	0.070	40
12	0.033	0.852	0.091	0.602	0.029	0.867	35
13	0.087	0.648	0.092	0.629	0.088	0.642	30
14	0.217	0.298	0.148	0.481	0.223	0.285	25
15	0.287	0.219	0.274	0.243	0.284	0.225	20

Для візуалізації отриманих залежностей побудовано графік профілю зв'язку між середнім настроєм і майбутніми дохідностями (ROI), наведений на рисунку 2:

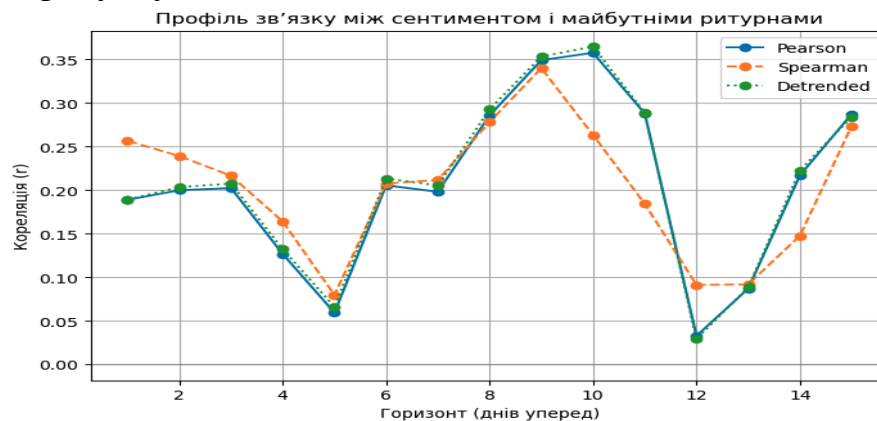


Рис. 2. Профіль зв'язку між настроєм новин і майбутніми (ROI)

З аналізу таблиці та графіка видно, що на коротких горизонтах прогнозу (1–3 дні) спостерігається слабка позитивна кореляція ($r \approx 0.2$), яка лише частково досягає статистичної значущості. Це вказує на наявність часткової миттєвої реакції ринку на інформаційний потік, проте без стійкої тенденції. У проміжку 4–5 днів кореляційний зв'язок практично зникає. Починаючи з 6–10 днів, сила зв'язку знову зростає і досягає максимуму на горизонтах 8–10 днів, де коефіцієнт Пірсона становить $r \approx 0.35$ при $p < 0.02$. Такий результат свідчить про наявність затриманої реакції інвесторів: позитивний інформаційний фон не призводить до миттєвого зростання цін, проте проявляє ефект протягом тижня після публікації новин. Після 10-го дня сила зв'язку поступово знижується, що може бути зумовлено згасанням впливу попередніх новин.

Збіг поведінки коефіцієнтів Пірсона, Спірмена та детрендованої кореляції свідчить, що виявлений ефект не є артефактом загального ринкового тренду, а має реальну поведінкову природу. Це означає, що зміни цін не просто слідують за зростанням ринку в цілому, а дійсно відображають

реакцію на емоційний контекст новин. Отримані результати узгоджуються з положеннями поведінкової фінансової теорії [4]

Висновки

Розроблено математичне та програмне забезпечення, яке:

- інтегрує фінансові часові ряди та текстові новини;
- реалізує автоматичний збір даних із NewsAPI та Yahoo Finance;
- оцінює емоційне забарвлення новин за допомогою FinBERT;

Проведений етап дослідження дозволяє зробити висновок про наявність прогностичного потенціалу сентимент-аналізу в короткострокових фінансових прогнозах. Система, створена у межах роботи, реалізує автоматизований збір новин, їх обробку за допомогою FinBERT, синхронізацію з фінансовими даними та кількісний аналіз взаємозв'язку між настроями інвесторів і ринковими результатами.

Література

1. Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In Proceedings of NAACL-HLT 2019, 4171–4186. Association for Computational Linguistics. <https://doi.org/10.18653/v1/N19-1423>
2. NewsAPI. (2025). NewsAPI – Worldwide News Aggregator API. Retrieved from <https://newsapi.org>
3. Yahoo Finance. (2025). Yahoo Finance API documentation. Retrieved from <https://finance.yahoo.com>
4. Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139–1168. <https://doi.org/10.1111/j.1540-6261.2007.01232.x>