

АНАЛІЗ КУРСІВ ФІНАНСОВИХ ІНСТРУМЕНТІВ З ВИКОРИСТАННЯМ МАШИННОГО НАВЧАННЯ ДЛЯ ФОРМУВАННЯ ІНВЕСТИЦІЙНОГО ПОРТФЕЛЮ

У статті розглянуто проблему формування та оптимізації інвестиційного портфеля в умовах високої волатильності фінансових ринків. Актуальність дослідження зумовлена необхідністю підвищення ефективності управління інвестиціями шляхом поєднання сучасних технологій збору даних та методів машинного навчання. Запропоновано систему, яка поєднує веб-скрапінг для автоматизованого збору фінансових даних із відкритих джерел та оптимізацію портфеля з використанням reinforcement learning. Методологія дослідження передбачає три етапи: формування вибірки фінансових інструментів, попередню обробку даних та оптимізацію портфеля із застосуванням класичних методів і алгоритму PPO (Proximal Policy Optimization). Результати показали, що нейромережева модель забезпечує вищий рівень диверсифікації та адаптивності порівняно з традиційним підходом Марковіца. Отримані висновки підтверджують доцільність інтеграції сучасних методів машинного навчання в процес управління інвестиційними портфелями для підвищення стійкості та ефективності фінансових стратегій.

Ключові слова: інвестиційний портфель, веб-скрапінг, оптимізація, машинне навчання, reinforcement learning, PPO, диверсифікація активів, ризик, дохідність.

The paper discusses the problem of forming and optimizing an investment portfolio in conditions of high volatility of financial markets. The study's relevance is determined by the need to improve investment management efficiency by combining modern data collection technologies and machine learning methods. A proposed system combines web scraping for automated financial data collection from open sources and portfolio optimization using reinforcement learning. The research methodology involves three stages: forming a sample of financial instruments, preliminary data processing, and portfolio optimization using classical methods and the PPO (Proximal Policy Optimization) algorithm. The results showed that the neural network model provides a higher level of diversification and adaptability than the traditional Markowitz approach. The conclusions confirm the feasibility of integrating modern machine learning methods into the investment portfolio management process to increase the stability and effectiveness of financial strategies.

Keywords: investment portfolio, web scraping, optimization, machine learning, reinforcement learning, PPO, asset diversification, risk, return.

Вступ

У сучасних умовах розвитку глобальної економіки фінансові ринки відіграють ключову роль у формуванні інвестиційних стратегій як приватних інвесторів, так і великих інституційних гравців. Швидкі зміни вартості активів, висока волатильність та зростання кількості доступних фінансових інструментів роблять задачу ефективного управління інвестиційним портфелем однією з найбільш актуальних у сфері фінансового менеджменту. Для прийняття обґрунтованих рішень необхідні дані, які відповідають вимогам оперативності, достовірності та повноти. Традиційні методи отримання інформації, що базуються на ручному зборі або використанні комерційних баз даних, дедалі частіше поступаються автоматизованим технологіям.

Одним із перспективних напрямів у цьому контексті є застосування веб-скрапінгу, який забезпечує можливість швидкого та систематизованого збору великих обсягів фінансових даних із відкритих джерел [7,8]. Використання подібних технологій дозволяє створювати власні набори даних для подальшого аналізу та оптимізації портфеля, що є особливо цінним у випадках, коли доступ до платних інформаційних ресурсів є обмеженим [9]. Автоматизація процесу збору інформації дає змогу не лише зменшити витрати часу, але й підвищити точність і актуальність результатів дослідження.

Разом із тим, проблема ефективного використання зібраних даних залишається відкритою. Класичні методи оптимізації інвестиційного портфеля, зокрема модель Марковіца та її модифікації, демонструють певні обмеження у випадках, коли ринок характеризується нестабільністю та наявністю складних нелінійних залежностей [1]. Ці підходи передбачають стаціонарність процесів і часто ігнорують мінливість ринкових умов. У результаті оптимізація, побудована виключно на традиційних статистичних припущеннях, не завжди відображає реальні ризики та перспективи інвестування [2-4].

Саме тому актуальним є поєднання новітніх інструментів збору даних із сучасними методами аналізу та оптимізації. Веб-скрапінг може слугувати фундаментом для формування якісних вибірок, які надалі застосовуються у моделях оптимізації з використанням як класичних статистичних алгоритмів, так і методів машинного навчання [5, 6]. Такий підхід створює передумови для побудови більш стійких та адаптивних інвестиційних стратегій.

Метою цієї статті є розробка системи аналізу курсів фінансових інструментів на основі веб-скрапінгу з подальшою оптимізацією

інвестиційного портфеля. Дослідження передбачає інтеграцію технологій збору та обробки даних з методами оптимізації, що дозволяє оцінити ефективність різних підходів та обґрунтувати їх застосування в умовах сучасного фінансового ринку.

Огляд літератури

Дослідження проблематики формування інвестиційного портфеля бере свій початок у класичних роботах Гаррі Марковіца, який запропонував математичну модель оптимізації співвідношення між дохідністю та ризиком [1]. Запропонована ним теорія портфеля заклала основу сучасного фінансового аналізу та дозволила розглядати ризик як кількісну характеристику інвестиційного середовища. Подальший розвиток отримала методика оцінки ефективності інвестицій за допомогою коефіцієнта Шарпа, який дозволяє порівнювати портфелі з урахуванням їхньої дохідності відносно рівня ризику [2].

Згодом у наукових працях з'явилися додаткові інструменти оцінювання ризиків. Серед них варто виділити показник Value-at-Risk, який широко використовується банківськими та інвестиційними установами для оцінки можливих втрат у портфелі при заданому рівні довіри [3]. Значного поширення набули також методи аналізу максимального просідання (Maximum Drawdown), що дозволяють оцінити стабільність портфеля в умовах різких ринкових змін [4]. Поєднання класичних індикаторів і показників ризик-менеджменту дало змогу зробити оцінку інвестиційних стратегій більш повною.

Упродовж останніх десятиліть у фінансових дослідженнях активно застосовуються методи штучного інтелекту. Значну увагу приділено використанню рекурентних нейронних мереж, зокрема архітектури LSTM, для прогнозування цінкових рядів і динаміки фінансових інструментів. У працях Т. Фішера та К. Крауса доведено ефективність застосування LSTM у прогнозуванні котирувань фондових індексів, що підтверджує можливість використання глибинного навчання для побудови інвестиційних стратегій [5]. Паралельно розвивається підхід навчання з підкріпленням. Дослідження З. Цзяна та співавторів показали, що агент, навчений за допомогою deep reinforcement learning, може ефективно управляти структурою портфеля, адаптуючись до змін ринку в реальному часі [6].

Окремим напрямом є використання автоматизованих інструментів збору даних. Веб-скрапінг став важливим методом формування фінансових датасетів для подальшого аналізу. Праці Р. Мітчелла та Р. Лоусона описують практичні підходи до побудови систем веб-скрапінгу мовою Python та їх застосування в аналізі великих обсягів інформації [7,8]. У сфері фінансів ці технології дозволяють отримувати дані з відкритих джерел, таких як Yahoo Finance, Bloomberg або Investing, забезпечуючи

дослідників і практиків актуальною інформацією без необхідності використовувати комерційні платформи [9].

Таким чином, наявні дослідження охоплюють як класичні статистичні підходи, так і сучасні методи машинного навчання та технології автоматизації збору даних. Новизна даної роботи полягає у спробі інтегрувати веб-скрапінг як базовий інструмент отримання фінансової інформації з подальшим використанням методів оптимізації інвестиційного портфеля. Це дозволяє одночасно вирішити проблему доступності якісних даних і підвищити ефективність побудови інвестиційних стратегій.

Для диверсифікації інвестиційного портфелю активи обираються з різних галузей, континентів і ринків [10, 11]. Для підвищення точності прогнозу застосовуються не лише дані часових рядів, показники фінансових звітів та інші числові предиктори, але і неструктурована інформація з відео, підкастів, графіків тощо [12, 13]. Традиційною методологією для розробки інвестиційних портфелів виступає теорія Марковиця, яка обирається як бенчмаркінг при розробці нових торговельних стратегій [14, 15].

Методологія дослідження

Дослідження виконано у три етапи, які охоплювали збір даних, їхню попередню обробку та оптимізацію інвестиційного портфеля. На першому етапі здійснювався збір історичних цін акцій п'ятдесяти компаній США за період з 2010 до 2024 року. Дані отримувалися з відкритих джерел за допомогою веб-скрапінгу мовою Python із застосуванням бібліотек BeautifulSoup та Selenium, а також через API уfinance [9]. Для формування вибірки були обрані компанії з різних секторів економіки, включно з технологіями, охороною здоров'я, фінансами, промисловістю, енергетикою, споживчими товарами, комунікаціями, нерухомістю та біотехнологіями. Після була проведена попередня обробка, залишалися тільки ті акції, які мали такі показники - сукупна (накопичена) дохідність повинна бути більше -0.05 та максимальна просадка менша за 0.3 [4].

Такий підхід дозволив забезпечити природну диверсифікацію портфеля, уникнути надмірної концентрації ризику в окремих галузях, та дозволив підтримувати стабільну ефективність портфеля.

На другому етапі проводилася попередня обробка даних. Для кожної акції розраховувалися щоденні доходності на основі змін цін закриття. Дані нормалізувалися для приведення до порівнянного масштабу шляхом центрування навколо нуля та масштабування стандартного відхилення до одиниці. Додатково були побудовані описові статистики, включно з середнім значенням, медіаною, стандартним відхиленням та мінімальними і максимальними показниками. Вони використовувалися для навчання агентів reinforcement learning [6].

Третій етап дослідження був спрямований на оптимізацію портфеля із застосуванням підходу навчання з підкріпленням на базі бібліотеки Stable-Baselines3 [6]. Для реалізації використано алгоритм PPO (Proximal Policy Optimization) з багатошаровою нейронною мережею. Вхідними даними для агента виступали послідовності щоденних доходностей активів, тоді як вихідні значення відображали розподіл ваг у портфелі, що нормалізувалися за допомогою моделі RL, яка використовувала кастомну функцію reward. Функція винагороди була сформована з урахуванням прибутковості портфеля, рівня ризику та штрафних коефіцієнтів за перевищення максимальної ваги окремого активу та надмірне зростання просідання капіталу. Навчання проводилося на історичних даних за період 2010–2024 років, тоді як тестування здійснювалося на відрізок 2024–2025 років. Така побудова експерименту дала можливість відокремити процес підготовки моделі від перевірки її ефективності та оцінити здатність агента функціонувати в умовах змін ринкової динаміки.

Результати

Розроблена система складалася з трьох основних модулів. Перший модуль виконував збір даних за допомогою веб-скрапінгу та API-запитів до платформи фінансової аналітики [7-9]. Другий модуль відповідав за попередню обробку даних, яка включала розрахунок доходностей, нормалізацію та побудову описових статистичних показників. Третій модуль реалізовував оптимізацію портфеля із використанням класичних методів та підходу на основі нейромережевої моделі з підкріплювальним навчанням [1-3,5].

У межах класичного підходу оптимізація дозволила виокремити обмежену кількість активів, які отримали ненульові ваги у структурі портфеля. Сформований портфель характеризувався певною концентрацією у вибраних компаніях, проте завдяки великій вибірці даних зберігав стабільність і демонстрував поєднання ризику та потенційної доходності (рис. 1).

Рис. 1. Динаміка вартості інвестиційного портфелю



Нейромережева модель, навчена за допомогою reinforcement learning, показала більш рівномірний розподіл ваг між активами порівняно з класичною оптимізацією. Це свідчило про вищий рівень диверсифікації та здатність моделі адаптуватися до нових ринкових умов [6]. У результаті портфель, сформований на основі такого підходу, характеризувався більш збалансованим співвідношенням ризику та доходності, а також меншою залежністю від обмеженої кількості активів (рис. 2).



Рис. 2. Оцінка інвестиційного портфеля для машинного навчання з підкріпленням

Порівняння двох методів показало, що класична оптимізація здатна забезпечувати вищий рівень очікуваної доходності, проте з ризиком концентрації портфеля. Натомість використання нейромережевої моделі

створює більш гнучку та диверсифіковану структуру активів. Це дає підстави стверджувати, що поєднання цих методів може бути доцільним для побудови стратегій управління інвестиційним портфелем, де враховуються як ефективність, так і диверсифікація.

ЛІТЕРАТУРА

1. Markowitz, H. *Portfolio Selection: Efficient Diversification of Investments*. New York: John Wiley & Sons, 1959.
2. Sharpe, W. F. The Sharpe Ratio. *Journal of Portfolio Management*, 21(1), 49–58, 1995.
3. Jorion, P. *Value at Risk: The New Benchmark for Managing Financial Risk*. 3rd edition. McGraw-Hill, 2006.
4. Magdon-Ismael, M., Atiya, A. F., Prata, A., & Abu-Mostafa, Y. On the Maximum Drawdown of a Portfolio. *Journal of Risk*, 4(2), 27–42, 2001.
5. Fischer, T., & Krauss, C. Deep learning with long short-term memory networks for financial market predictions. *European Journal of Operational Research*, 270(2), 654–669, 2018.
6. Jiang, Z., Xu, D., & Liang, J. A Deep Reinforcement Learning Framework for the Financial Portfolio Management Problem. *arXiv:1706.10059*, 2017.
7. Mitchell, R. *Web Scraping with Python: Collecting Data from the Modern Web*. O'Reilly Media, 2018.
8. Lawson, R. *Practical Web Scraping for Data Science*. Packt Publishing, 2020.
9. Yahoo Finance. *Financial Data API Documentation*. <https://finance.yahoo.com>
10. Savchenko, S., & Kobets, V. (2025). Allocation of investment portfolio assets classes using machine learning. In *Communications in computer and information science* (pp. 285–296). https://doi.org/10.1007/978-3-031-81372-6_21
11. Ivanov, O., & Kobets, V. (2025). Integrating semantic analysis and financial indicators of business reports for predicting stock prices. In *Communications in computer and information science* (pp. 297–308). https://doi.org/10.1007/978-3-031-81372-6_22
12. Ivanov, O., & Kobets, V. (2025c). Future Financial Impact Analysis from Sentiment and Indicators Analysis. *Computational Economics*. <https://doi.org/10.1007/s10614-025-10891-7>
13. Ivanov, O., & Kobets, V. (2023). Data analysis for predicting stock prices using financial indicators based on business reports. In *Communications in computer and information science* (pp. 227–239). https://doi.org/10.1007/978-3-031-48325-7_17
14. Savchenko, S., & Kobets, V. (2023). Increasing Investment Portfolio Profitability with Computer Analysis Trading Strategies. In *Communications in computer and information science* (pp. 252–264). https://doi.org/10.1007/978-3-031-48325-7_19
15. Kobets, V., Savchenko, S.: Building an Optimal Investment Portfolio with Python Machine Learning Tools. In: Anisimov, A. et al. (eds.) *Proceedings of the 4th International Scientific Conference "Information Technology and Implementation", IT and I 2022*, pp. 307–315. CEUR-WS 3347, Aachen University (2022). https://ceur-ws.org/Vol-3347/Short_1.pdf