VISENET: Vietnamese Investment weighted-Scoring and Ensemble Network for Enhanced Trading

Trần Nam Hải, Nguyễn Vũ Trung Kiên, Phạm Tiến Dũng
Trường Công nghệ Thông tin và Truyền thông, Đại học Bách khoa Hà Nội, Việt Nam
Email: {trannamhai.5d, ngkienn89, tiendungcbh1801}@gmail.com

Đội thi HD4K - Vòng 02 cuộc thi Data Science Talent Competition 2025

Tóm tắt nội dung*—Chiến lược lựa chọn và giao dịch cổ phiếu giữ một vai trò quan trọng trong bài toán đầu tư nhưng những phương pháp chấm điểm có trọng số truyền thống còn nhiều hạn chế: chưa phản ánh đúng tác động của mỗi yếu tố đến hiệu quả danh mục, khó tìm ra tổ hợp tối ưu và thiếu linh hoạt với sở thích nhà đầu tư. Chúng tôi xây dựng một quy trình bắt đầu từ việc áp dụng thiết kế thực nghiệm hỗn hợp (mixture experimental design) để xác định trọng số yếu tố và xây dựng mô hình dự báo hiệu quả danh mục, từ đó chọn ra 30 cổ phiếu tiêu biểu trên thị trường chứng khoán Việt Nam (HNX, HOSE, UPCOM) giai đoạn 2018–2025. Trên tập hợp này, chúng tôi tiếp tục với chiến lược giao dịch dựa trên học tăng cường sâu theo hướng ensemble, kết hợp ba thuật toán actor-critic (PPO, A2C, TD3). Kết quả thực nghiệm cho thấy phương pháp hỗn hợp giúp dự báo chính xác và nhận diện tương tác giữa các trọng số, trong khi mô hình ensemble vượt trội so với từng thuật toán riêng lẻ và các chuẩn tham chiếu truyền thống về lợi nhuận đã điều chỉnh rủi ro. Phương pháp luận này cung cấp một khung toàn diện để tối ưu hóa danh mục và xây dựng chiến lược giao dịch phù hợp với sở thích nhà đầu tư. Mã nguồn khả dụng tại: https://github.com/CryAndRRich/visenet.

***Từ khoá*—Lựa chọn cổ phiếu, Thiết kế thực nghiệm hỗn hợp, Học tăng cường sâu, Chiến lược giao dịch ensemble, Thuật toán actor-critic, Thị trường chứng khoán Việt Nam**

I. GIỚI THIỆU

Thị trường chứng khoán Việt Nam đang mở rộng nhanh chóng, kéo theo nhu cầu về những chiến lược lựa chọn và giao dịch cổ phiếu hiệu quả. Tuy nhiên, phương pháp chấm điểm có trọng số (weighted-scoring) truyền thống vẫn còn nhiều hạn chế: khó phản ánh chính xác tác động của từng yếu tố đến hiệu quả danh mục, thiếu khả năng khám phá các tương tác phi tuyến và tổ hợp tối ưu giữa các yếu tố.

Song song với đó, sự phát triển của học tăng cường sâu (deep reinforcement learning - DRL) đã mở ra một hướng tiếp cận mới cho giao dịch định lượng. DRL cho phép các mô hình học trực tiếp từ dữ liệu thị trường, tối ưu hóa chiến lược, và thích ứng tốt hơn với biến động. Tuy nhiên, việc áp dụng một thuật toán DRL đơn lẻ thường dẫn đến khả năng khái quát hạn chế, từ đó đặt ra nhu cầu kết hợp nhiều mô hình (ensemble) để tận dụng ưu điểm bổ sung lẫn nhau.

Trong vòng 02 của **Data Science Talent Competition 2025**, thử thách đặt ra bao gồm:

- Xây dựng chiến lược lựa chọn cổ phiếu có tiềm năng sinh lời dựa trên dữ liệu lịch sử, dữ liệu thời gian thực và các tín hiệu tổng hợp từ phân tích kỹ thuật (technical analysis – TA) và phân tích cơ bản (fundamental analysis – FA).
- Thiết kế hệ thống có khả năng đưa ra quyết định đầu tư, được kiểm chứng bằng các thí nghiệm backtest và có thể thích ứng trước biến động phức tạp của thị trường.

Để giải quyết bài toán này, chúng tôi đề xuất **VISENET (Vietnamese Investment weighted-Scoring and Ensemble Network for Enhanced Trading)**, một khung phương pháp tích hợp giữa thiết kế thực nghiệm hỗn hợp (mixture experimental design) và học tăng cường sâu theo hướng ensemble. Bước đầu, phương pháp mixture experimental design được sử dụng để xác định và mô hình hóa trọng số các yếu tố tác động đến hiệu quả danh mục, từ đó lựa chọn 30 cổ phiếu tiêu biểu trên ba sàn giao dịch tại Việt Nam (HOSE, HNX, UPCOM) trong giai đoạn 2018–2025. Trên tập hợp cổ phiếu này, chúng tôi phát triển chiến lược giao dịch dựa trên sự kết hợp của ba thuật toán actor-critic hiện đại:

- **PPO**: Proximal Policy Optimization
- **A2C**: Advantage Actor-Critic
- **TD3**: Twin Delayed Deep Deterministic Policy Gradient

Kết quả cho thấy VISENET mang lại:

- Dự báo chính xác và nhận diện tương tác trọng số nhờ mixture design.
- Hiệu quả vượt trội về lợi nhuận điều chỉnh rủi ro nhờ ensemble DRL so với từng thuật toán riêng lẻ và phương pháp truyền thống.
- Khả năng thích ứng với đặc thù thị trường Việt Nam.

Bằng việc kết hợp thiết kế thực nghiệm và DRL, nghiên cứu này vừa đáp ứng yêu cầu cuộc thi (một hệ thống đầu tư tự động, kiểm chứng qua backtest), vừa cung cấp một khung phương pháp luận toàn diện cho việc tối ưu hóa danh mục và xây dựng chiến lược giao dịch cá nhân hóa trong thực tiễn.

II. DỮ LIỆU TỔNG QUAN

Chúng tôi thu thập dữ liệu cổ phiếu và chỉ số thị trường trực tiếp từ **FiinQuantX**, một thư viện Python cho phép truy xuất dữ liệu từ kho dữ liệu của **FiinGroup**. Quy trình gồm năm bước chính như sau:

Bước 1: Khởi tạo phiên FiinQuant

Trước hết, chúng tôi sử dụng lớp FiinSession để kết nối và đăng nhập vào API, thông qua tài khoản đã được cấp.

Bước 2: Lấy danh sách mã cổ phiếu

Tiếp theo, chúng tôi truy xuất danh sách toàn bộ mã cổ phiếu đang giao dịch trên ba sàn chứng khoán chính tại Việt Nam:

- Danh sách mã trên sàn **HOSE** (đại diện bởi VNINDEX)
- Danh sách mã trên sàn **HNX** (đại diện bởi HNXINDEX)
- Danh sách mã trên sàn **UPCOM** (đại diện bởi UPCOMINDEX)

Sau đó, 3 danh sách được gộp lại và loại bỏ trùng lặp để thu được tập hợp duy nhất, bao quát toàn bộ cổ phiếu niêm yết trên ba sàn.

Bước 3: Lấy dữ liệu giao dịch

Dữ liệu giao dịch được truy xuất thông qua phương thức `client.Fetch_Trading_Data`. Trong bước này, chúng tôi xác định rõ phạm vi dữ liệu và các trường dữ liệu cần thiết như sau:

- Phạm vi dữ liệu: Khoảng thời gian lấy dữ liệu kéo dài trong khoảng 7 năm (29/11/2018 – 29/08/2025).
- Trường dữ liệu:
  - Giá cổ phiếu: *Open, High, Low, Close* – lần lượt là giá mở cửa, cao nhất, thấp nhất, và đóng cửa trong ngày.
  - Thanh khoản: *Volume* – khối lượng giao dịch.
  - Các trường bổ sung: *Bu, Sd, Fb, Fs, Fn*.
  - Điều chỉnh sự kiện: *adjusted=True* để hiệu chỉnh dữ liệu theo các sự kiện doanh nghiệp (chia cổ tức, phát hành thêm), đảm bảo dữ liệu phản ánh chính xác biến động thị trường.

Bước 4: Tính toán chỉ số TA và trích xuất chỉ báo FA

Sau khi thu thập dữ liệu giao dịch gốc, chúng tôi mở rộng tập dữ liệu bằng cách bổ sung các chỉ số phân tích kỹ thuật (TA) và một chỉ báo phân tích cơ bản (FA) trực tiếp vào dữ liệu:

- Chỉ số phân tích kỹ thuật (TA): Dựa trên các hàm có sẵn trong thư viện FiinQuant, chúng tôi tính toán các chỉ số quan trọng phản ánh động lượng, xu hướng và biến động thị trường, bao gồm:
  - *Volatility*: đo lường mức độ biến động giá.
  - *RSI*: chỉ số sức mạnh tương đối.
  - *Liquidity*: chỉ số phản ánh mức độ thanh khoản.
  - *MACD*: chỉ báo phân kỳ hội tụ đường trung bình động.
  - *CCI*: chỉ báo đo độ lệch giá so với trung bình.
  - *ADX*: chỉ số đo sức mạnh xu hướng.
- Chỉ báo phân tích cơ bản (FA): Chúng tôi lựa chọn duy nhất một chỉ báo FA mang tính chất quy mô và cấu trúc vốn của doanh nghiệp là *Shares Outstanding* (số lượng cổ phiếu đang lưu hành). Chỉ báo này được trích xuất từ thư viện **vnstock** cho tất cả các mã trên ba sàn HOSE, HNX, và UPCOM.

Bước 5: Lưu dữ liệu

Sau khi hoàn tất việc thu thập và gộp dữ liệu vào một Dataframe, chúng tôi lưu dữ liệu thành tệp định dạng `.csv` để phục vụ cho quá trình phân tích và mô hình hóa ở các bước tiếp theo, đồng thời hạn chế số lần gọi API không cần thiết.

III. MÔ HÌNH WEIGHTED-SCORING

A. Dữ liệu đầu vào

Bộ dữ liệu đầu vào của mô hình *weighted-scoring* được xây dựng từ **1.029 mã cổ phiếu** niêm yết trên ba sàn giao dịch chứng khoán tại Việt Nam, bao gồm: *Sở Giao dịch Chứng khoán Hà Nội (HNX)*, *Sở Giao dịch Chứng khoán Thành phố Hồ Chí Minh (HOSE)* và *UPCOM*.

Dữ liệu được thu thập trong giai đoạn từ **29/11/2018** đến **29/08/2025**, bao gồm các biến số thị trường cơ bản và chỉ báo kỹ thuật sau:

- Giá cổ phiếu: *Open, Close, High, Low*.
- Các chỉ báo kỹ thuật: *Volatility, RSI, Liquidity, MACD, CCI, ADX*.
- Thông tin tài chính bổ sung: *Shares Outstanding* (số lượng cổ phiếu lưu hành), lưu trữ trong tệp dữ liệu định dạng `.csv`.

Khoảng thời gian này được lựa chọn nhằm đảm bảo bao quát nhiều trạng thái thị trường khác nhau, từ tăng trưởng, suy giảm cho đến biến động mạnh.

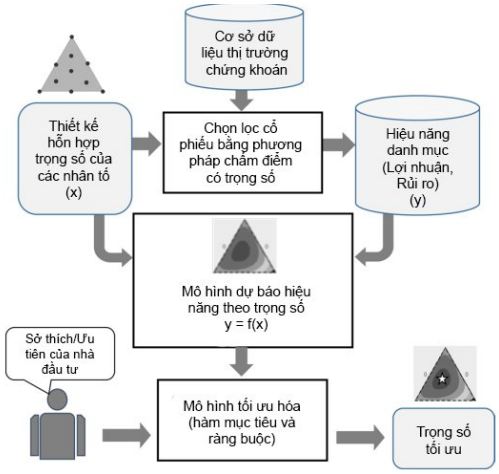
B. Mô hình

1) Khung tiếp cận mô hình:

Hướng tiếp cận sử dụng mô hình hỗn hợp được xây dựng như sau:

- Thiết kế các tổ hợp trọng số cho các *yếu tố cổ phiếu* theo phương pháp *mixture design* [1] [2], tạo ra tập hợp vector trọng số x.
- Với mỗi x, tiến hành backtest trên dữ liệu lịch sử để thu thập các **thước đo hiệu năng** y, từ đó hình thành bộ dữ liệu ghép cặp (x, y).
- Xây dựng mô hình dự báo hiệu năng $y = f(x)$ bằng hồi quy đa thức đa biến, nhằm phân tích mối quan hệ giữa hiệu năng và trọng số các yếu tố, đồng thời nhận diện tương tác giữa chúng;
- Sử dụng mô hình dự báo để tìm tổ hợp trọng số tối ưu, đáp ứng các ràng buộc phù hợp với sở thích nhà đầu tư (rủi ro, quy mô danh mục).
- Thực hiện kiểm định ngoài mẫu thông qua backtest độc lập để xác nhận hiệu quả của tổ hợp trọng số tối ưu.

Trong khung tiếp cận này, mô hình *weighted–scoring* là trục triển khai chính: mỗi yếu tố tố được chuẩn hoá điểm trong khoảng $[0, 100]$ theo thứ hạng (cao nhất = 100, thấp nhất = 0), sau đó tổng hợp thành điểm tổng $S_i = \sum_j w_j s_{ij}$ với $\sum_j w_j = 1, w_j \geq 0$. Danh mục được xây dựng từ nhóm cổ phiếu có S_i cao nhất.



Hình 1. Sơ đồ khung mô hình chọn lọc cổ phiếu bằng phương pháp chấm điểm có trọng số

2) Thiết kế thực nghiệm hỗn hợp (Mixture experimental design):

Để phân tích mối quan hệ giữa tổ hợp trọng số và hiệu năng danh mục, nghiên cứu sử dụng *simplex–centroid design*. Với q yếu tố, thiết kế này sinh ra $2^q - 1$ hỗn hợp thí nghiệm, tuân thủ ràng buộc $\sum_{i=1}^{q} w_i = 1, w_i \geq 0$.

Mỗi thước đo hiệu năng y được mô hình hoá bởi đa thức hỗn hợp (mixture polynomial) [1]:

$$E[y] = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j} \beta_{ij} x_i x_j + \sum_{i<j<k} \beta_{ijk} x_i x_j x_k + \cdots + \beta_{12\ldots q} \prod_{i=1}^{q} x_i, \quad (1)$$

trong đó x_i là trọng số của yếu tố thứ i, và các β_i là các hệ số hồi quy.

Hiệu năng danh mục được đặc trưng bởi ba chỉ tiêu: lợi nhuận vượt trội (α), rủi ro hệ thống (β), và quy mô/thanh khoản (MV). Trong đó α và β được ước lượng từ mô hình thị trường [3]:

$$(R_{p,t} - R_{f,t}) = \alpha + \beta(R_{m,t} - R_{f,t}) \quad (2)$$

với $R_{p,t}$ là lợi nhuận danh mục, $R_{m,t}$ lợi nhuận thị trường, $R_{f,t}$ lãi suất phi rủi ro. Chỉ tiêu MV được tính bằng log của trung bình theo thời gian của *trung vị vốn hoá* các cổ phiếu trong danh mục ở từng tháng.

3) Triển khai với bộ ba yếu tố Volatility–RSI–Liquidity*:*

Các yếu tố được sử dụng trong mô hình được quyết định bởi sở thích, ưu tiên của mỗi nhà đầu tư. Trong phần triển khai thực nghiệm này, chúng tôi quyết định sử dụng ba chỉ báo kỹ thuật được tính toán như sau:

- **Volatility**: độ biến động giá, ước lượng trong cửa sổ 20 ngày và annualized.
- **Liquidity**: độ thanh khoản, đo bằng trung bình giá trị dao động (TR) trong 14 ngày gần nhất.
- **RSI**: chỉ số sức mạnh tương đối với cửa sổ 14 ngày, phản ánh trạng thái quá mua/quá bán.

a) Chuẩn hoá điểm yếu tố:

Tại mỗi tháng t, các chỉ báo của cổ phiếu i được chuẩn hoá về thang điểm $[0, 100]$ theo thứ hạng (percentile):

- **Volatility**: xếp hạng tăng dần, điểm cao ứng với biến động thấp.
- **RSI**: xếp hạng giảm dần, điểm cao ứng với mức RSI lớn.
- **Liquidity**: xếp hạng giảm dần, điểm cao ứng với thanh khoản cao.

b) Điểm tổng và xây dựng danh mục:

Với vector trọng số $W = (w_{\text{vol}}, w_{\text{rsi}}, w_{\text{liq}})$, điểm tổng của cổ phiếu i tại tháng t được xác định bởi

$$S_{i,t} = w_{\text{vol}}\, s_{i,t}^{\text{vol}} + w_{\text{rsi}}\, s_{i,t}^{\text{rsi}} + w_{\text{liq}}\, s_{i,t}^{\text{liq}}. \quad (3)$$

Mỗi tháng, chọn nhóm cổ phiếu thuộc top 30 theo $S_{i,t}$ để xây dựng danh mục cân bằng đều. Chuỗi lợi nhuận danh mục hàng tháng $R_{p,t}$ được tính từ giá đóng cửa của danh mục.

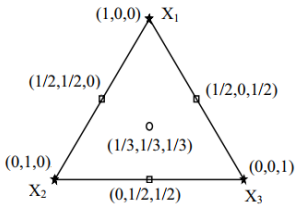
c) Thiết kế hỗn hợp cho ba yếu tố:

Với $q = 3$, thiết kế *simplex–centroid* tạo ra 7 tổ hợp trọng số:

$$(1,0,0),\ (0,1,0),\ (0,0,1),$$
$$\left(\tfrac{1}{2},\tfrac{1}{2},0\right),\ \left(\tfrac{1}{2},0,\tfrac{1}{2}\right),\ \left(0,\tfrac{1}{2},\tfrac{1}{2}\right),$$
$$\left(\tfrac{1}{3},\tfrac{1}{3},\tfrac{1}{3}\right).$$

Với mỗi tổ hợp W, thực hiện backtest toàn bộ giai đoạn để thu chuỗi $\{\alpha_t(W), \beta_t(W), \text{MV}_t(W)\}$, từ đó tính đặc trưng tổng hợp.

	X_1	X_2	X_3
1	1	0	0
2	0	1	0
3	0	0	1
4	1/2	1/2	0
5	1/2	0	1/2
6	0	1/2	1/2
7	1/3	1/3	1/3



Hình 2. Thiết kế simplex-centroid với 3 yếu tố

d) Mô hình dự báo hiệu năng:

Với mỗi thước đo $y \in \{\alpha, \beta, \text{MV}\}$, mô hình đa thức bậc 3 được ước lượng:

$$f_y[W] = \beta_1 w_{\text{vol}} + \beta_2 w_{\text{rsi}} + \beta_3 w_{\text{liq}} + \beta_{12} w_{\text{vol}} w_{\text{rsi}} + \beta_{13} w_{\text{vol}} w_{\text{liq}} + \beta_{23} w_{\text{rsi}} w_{\text{liq}} + \beta_{123} w_{\text{vol}} w_{\text{rsi}} w_{\text{liq}}. \quad (4)$$

trong đó các β_i là các hệ số hồi quy ứng với thước đo y.

e) Bài toán tối ưu hoá trọng số:

Từ các mô hình $f_\alpha(W), f_\beta(W), f_{\text{MV}}(W)$, tối ưu hoá tìm $W^\star$ phù hợp sở thích nhà đầu tư. Ví dụ, cực đại hoá α với ràng buộc trên β và MV [3]:

$$\begin{aligned} \max_{W} \quad & \alpha = f_\alpha(W) \\ \text{s.t.} \quad & \beta = f_\beta(W) \leq \beta^\star, \\ & \text{MV} = f_{\text{MV}}(W) \geq \text{MV}^\star, \\ & w_{\text{vol}} + w_{\text{rsi}} + w_{\text{liq}} = 1, \ \ w_j \geq 0. \end{aligned} \quad (5)$$

f) Diễn giải hệ số:

Khi phân tích các hệ số hồi quy:

- Hệ số tuyến tính β_i dương và lớn cho thấy việc tăng trọng số yếu tố i có xu hướng cải thiện thước đo hiệu năng y.
- Hệ số tương tác $\beta_{ij} > 0$ gợi ý hiệu ứng bổ sung giữa hai yếu tố, trong khi $\beta_{ij} < 0$ phản ánh sự triệt tiêu lẫn nhau.
- Hệ số bậc ba β_{123} biểu thị tương tác đồng thời của cả ba yếu tố.

C. Kết quả đầu ra

Sau khi thực hiện quá trình ước lượng hệ số hồi quy, tối ưu hoá tổ hợp trọng số và backtest trên toàn bộ giai đoạn dữ liệu, mô hình weighted-scoring cho ra danh mục cổ phiếu được chọn lọc cuối cùng (code minh hoạ tại tệp notebook `wscoring.ipynb`).

Với vector trọng số tối ưu $W^\star = (w^\star_{vol}, w^\star_{rsi}, w^\star_{liq})$, điểm tổng hợp của từng cổ phiếu i tại thời điểm t được tính theo công thức:

$$S^\star_{i,t} = w^\star_{\text{vol}}\, s^{\text{vol}}_{i,t} + w^\star_{\text{rsi}}\, s^{\text{rsi}}_{i,t} + w^\star_{\text{liq}}\, s^{\text{liq}}_{i,t}. \tag{6}$$

trong đó các chỉ báo được chuẩn hoá về thang điểm $[0, 100]$.

Mô hình tiến hành tính điểm thành phần cho toàn bộ cổ phiếu dựa trên ba chỉ báo RSI, Volatility và Liquidity, sau đó tổng hợp thành điểm số tổng $S^\star_{i,t}$. Toàn bộ các cổ phiếu được xếp hạng theo thứ tự giảm dần của $S^\star_{i,t}$, và **nhóm 30 cổ phiếu có điểm số cao nhất** được lựa chọn để xây dựng danh mục cuối cùng.

Tệp đầu ra:

- `top_30_score_after_train.csv`: chứa danh sách 30 cổ phiếu có điểm số cao nhất cùng với điểm RSI, Volatility, Liquidity và tổng điểm.
- `top_30_stocks_after_train.csv`: chứa dữ liệu giá của 30 cổ phiếu được chọn để phục vụ backtest và các phần phân tích tiếp theo.

Danh mục được xây dựng phản ánh trực tiếp mức độ ưu tiên giữa các yếu tố (Volatility-RSI-Liquidity) theo trọng số tối ưu. Kết quả đầu ra cho phép kiểm chứng hiệu quả của mô hình qua lợi nhuận vượt trội (α), rủi ro hệ thống (β) và quy mô thanh khoản (MV). Từ 30 loại cổ phiếu này, chúng tôi tiếp tục tiến hành lọc dữ liệu và sử dụng làm đầu vào cho **mô hình Ensemble**.

IV. MÔÔI TRƯỜNG GIAO DỊCH

Trước khi đi vào mô hình Ensemble, chúng tôi tiến hành mô hình hóa giao dịch chứng khoán như một Quy trình Quyết định Markov (Markov Decision Process - MDP), và xây dựng mục tiêu giao dịch dưới dạng tối đa hóa lợi nhuận kỳ vọng [4].

A. Mô hình MDP cho giao dịch chứng khoán

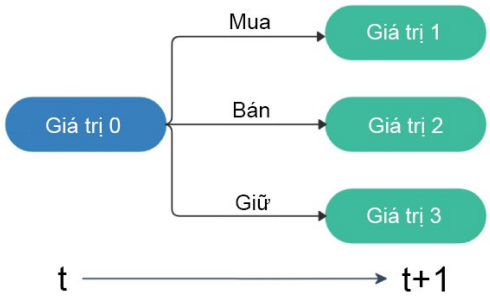
Để mô tả tính ngẫu nhiên của thị trường chứng khoán động, chúng tôi xây dựng mô hình MDP như sau:

- **Trạng thái (state)** $s = [p; h; b]$**:** một vector bao gồm giá cổ phiếu $p \in R^D_+$, số lượng cổ phiếu hiện nắm giữ $h \in Z^D_+$, và số dư tiền mặt $b \in R_+$, trong đó D là số lượng cổ phiếu và Z_+ là tập số nguyên không âm.
- **Hành động (action)** a**:** một vector hành động trên D cổ phiếu. Các hành động được phép gồm *bán*, *mua*, hoặc *giữ*, tương ứng làm giảm, tăng hoặc giữ nguyên số lượng cổ phiếu h.
- **Phần thưởng (rewward)** $r(s, a, s')$**:** phần thưởng trực tiếp khi thực hiện hành động a tại trạng thái s và chuyển sang trạng thái s'.
- **Chính sách (policy)** $\pi(s)$**:** chiến lược giao dịch tại trạng thái s, được định nghĩa như phân phối xác suất của các hành động tại trạng thái s.
- **Giá trị hành động (Q-value)** $Q^\pi(s, a)$**:** lợi nhuận kỳ vọng khi thực hiện hành động a tại trạng thái s theo chính sách π.

Quá trình chuyển trạng thái trong giao dịch chứng khoán được mô tả như sau: tại mỗi thời điểm, một trong ba hành động được thực hiện trên cổ phiếu d ($d = 1, \ldots, D$):

- **Bán:** bán $k[d] \in [1, h[d]]$ cổ phiếu, dẫn đến $h_{t+1}[d] = h_t[d] - k[d]$, với $k[d] \in Z_+$.
- **Giữ:** $h_{t+1}[d] = h_t[d]$.
- **Mua:** mua $k[d]$ cổ phiếu, dẫn đến $h_{t+1}[d] = h_t[d] + k[d]$.

Tại thời điểm t, hành động được thực hiện và giá cổ phiếu cập nhật ở thời điểm $t + 1$, do đó giá trị danh mục đầu tư có thể thay đổi từ "giá trị 0" sang "giá trị 1, 2, hoặc 3". Giá trị danh mục đầu tư được tính là $p^T h + b$.



Hình 3. Một giá trị danh mục đầu tư khởi điểm với ba hành động sẽ dẫn đến ba giá trị danh mục có thể xảy ra. Lưu ý rằng hành động 'giữ' có thể dẫn đến các giá trị danh mục khác nhau do sự thay đổi của giá cổ phiếu

B. Tích hợp các ràng buộc trong giao dịch chứng khoán

Chúng tôi đưa vào các giả định và ràng buộc để phản ánh các yếu tố thực tế như chi phí giao dịch, thanh khoản thị trường, và mức độ chấp nhận rủi ro:

- **Thanh khoản thị trường:** giả định lệnh có thể được khớp nhanh chóng ở mức giá đóng cửa, và giao dịch của mô hình chúng tôi không gây ảnh hưởng đến thị trường.
- **Số dư không âm** $b \geq 0$**:** hành động không được làm số dư âm. Với tập cổ phiếu bán S, mua B, và giữ H (không giao nhau, $S \cup B \cup H = \{1, \ldots, D\}$), ràng buộc số dư được viết:

$$b_{t+1} = b_t + (p^S_t)^T k^S_t - (p^B_t)^T k^B_t \geq 0 \tag{7}$$

- **Chi phí giao dịch:** chi phí phát sinh cho mỗi lệnh mua/bán, giả định bằng 0.1% giá trị giao dịch:

$$c_t = p^T k_t \times 0.1\% \tag{8}$$

- **Rủi ro thị trường sụp đổ:** để phản ánh rủi ro từ các sự kiện khủng hoảng, chúng tôi sử dụng chỉ số *turbulence* [5]:

$$turbulence_t = (y_t - \mu)\Sigma^{-1}(y_t - \mu)^T \in R \quad (9)$$

trong đó $y_t \in R^D$ là lợi nhuận cổ phiếu tại thời điểm t, $\mu \in R^D$ là trung bình lợi nhuận lịch sử, và $\Sigma \in R^{D\times D}$ là ma trận hiệp phương sai. Khi $turbulence_t$ vượt ngưỡng, ngừng mua và bán toàn bộ cổ phiếu đang nắm giữ. Giao dịch được tiếp tục khi chỉ số này trở về dưới ngưỡng.

C. Mục tiêu tối đa hóa lợi nhuận

Chúng tôi định nghĩa *hàm phần thưởng* (reward function) là sự thay đổi giá trị danh mục đầu tư khi thực hiện hành động a tại trạng thái s và chuyển sang trạng thái mới s':

$$r(s_t, a_t, s_{t+1}) = (b_{t+1} + p_{t+1}^T h_{t+1}) - (b_t + p_t^T h_t) - c_t \quad (10)$$

trong đó hạng tử thứ nhất và thứ hai lần lượt biểu diễn giá trị danh mục tại thời điểm $t+1$ và t. Cụ thể hơn, chúng tôi định nghĩa sự chuyển dịch số lượng cổ phiếu như sau:

$$h_{t+1} = h_t - k_t^S + k_t^B \quad (11)$$

và sự thay đổi số dư b_t được định nghĩa trong (7). Khi đó (10) có thể viết lại thành:

$$r(s_t, a_t, s_{t+1}) = r^H - r^S + r^B - c_t \quad (12)$$

trong đó:

$$r^H = (p_{t+1}^H - p_t^H)^T h_t^H \quad (13)$$

$$r^S = (p_{t+1}^S - p_t^S)^T h_t^S \quad (14)$$

$$r^B = (p_{t+1}^B - p_t^B)^T h_t^B. \quad (15)$$

Ở đây, r^H, r^S, và r^B lần lượt đại diện cho sự thay đổi giá trị danh mục đến từ các cổ phiếu *giữ*, *bán*, và *mua* khi chuyển từ thời điểm t sang $t+1$. Phương trình (12) chỉ ra rằng chúng ta cần tối đa hóa thay đổi tích cực của giá trị danh mục bằng cách mua và giữ các cổ phiếu có giá sẽ tăng trong bước tiếp theo, đồng thời giảm thiểu thay đổi tiêu cực bằng cách bán những cổ phiếu có giá sẽ giảm.

Chỉ số *turbulence* $turbulence_t$ được tích hợp trong hàm phần thưởng để xử lý vấn đề chấp nhận rủi ro trong trường hợp thị trường sụp đổ. Khi chỉ số trong (9) vượt ngưỡng, phương trình (14) trở thành:

$$r^{sell} = (p_{t+1} - p_t)^T k_t, \quad (16)$$

hàm ý rằng ta cần giảm thiểu thay đổi tiêu cực bằng cách bán toàn bộ cổ phiếu đang nắm giữ, do giá tất cả cổ phiếu đều giảm.

Mô hình được khởi tạo như sau: p_0 được đặt bằng giá cổ phiếu tại thời điểm ban đầu, b_0 là số vốn khởi tạo, h và $Q^\pi(s,a)$ được khởi tạo bằng 0, và $\pi(s)$ là phân phối đồng đều trên các hành động ở mỗi trạng thái. Sau đó, $Q^\pi(s_t, a_t)$ được cập nhật thông qua quá trình tương tác với môi trường thị trường chứng khoán.

Chiến lược tối ưu được xác định bởi phương trình Bellman, trong đó lợi nhuận kỳ vọng khi thực hiện hành động a_t tại trạng thái s_t được tính bằng kỳ vọng của phần thưởng trực tiếp $r(s_t, a_t, s_{t+1})$ cộng với lợi nhuận kỳ vọng trong tương lai tại trạng thái s_{t+1}. Với hệ số chiết khấu $0 < \gamma < 1$ nhằm đảm bảo hội tụ, ta có:

$$Q^\pi(s_t, a_t) = E_{s_{t+1}}\Big[r(s_t, a_t, s_{t+1}) + \gamma\, E_{a_{t+1}\sim\pi(s_{t+1})}[Q^\pi(s_{t+1}, a_{t+1})]\Big]. \quad (17)$$

Mục tiêu cuối cùng là xây dựng một chiến lược giao dịch tối ưu hóa sự thay đổi tích cực tích lũy của giá trị danh mục trong môi trường động. Để giải quyết bài toán này, chúng tôi áp dụng phương pháp học tăng cường sâu (deep reinforcement learning - DRL).

D. Môi trường thị trường chứng khoán

Trước khi huấn luyện mô hình học sâu tăng cường, chúng tôi xây dựng một môi trường mô phỏng giao dịch thực tế, cho phép mô hình thực hiện tương tác và học hỏi. Trong giao dịch thực tế, nhiều yếu tố cần được xem xét, ví dụ như giá lịch sử, số cổ phiếu nắm giữ, và các chỉ báo kỹ thuật. Mô hình giao dịch sẽ thu thập các thông tin này thông qua môi trường, và thực hiện các hành động đã được định nghĩa. Chúng tôi sử dụng `OpenAI Gym` [6] [7] [8] để triển khai môi trường và huấn luyện.

1) Môi trường cho nhiều cổ phiếu:

Chúng tôi sử dụng không gian hành động liên tục [9] để mô hình hóa giao dịch nhiều cổ phiếu, giả định rằng danh mục có tổng cộng 30 cổ phiếu.

a) Không gian trạng thái:

Trạng thái được biểu diễn bởi vector 181 chiều, bao gồm 7 thành phần thông tin $[b_t; p_t; h_t; M_t; R_t; C_t; X_t]$, mỗi thành phần được định nghĩa như sau:

- $b_t \in R_+$: số dư tại thời điểm t.
- $p_t \in R_+^{30}$: giá đóng cửa điều chỉnh của từng cổ phiếu.
- $h_t \in Z_+^{30}$: số cổ phiếu đang nắm giữ.
- $M_t \in R^{30}$: MACD (Moving Average Convergence Divergence), tính từ giá đóng cửa, là chỉ báo động lượng phổ biến [10].
- $R_t \in R_+^{30}$: RSI (Relative Strength Index), tính từ giá đóng cửa, đo lường mức độ thay đổi gần đây. Nếu giá quanh đường hỗ trợ, cổ phiếu bị bán quá mức và có thể mua; nếu quanh đường kháng cự, cổ phiếu bị mua quá mức và có thể bán [10].
- $C_t \in R_+^{30}$: CCI (Commodity Channel Index), tính từ giá cao, thấp và đóng cửa, so sánh giá hiện tại với giá trung bình để gợi ý mua hoặc bán [11].
- $X_t \in R^{30}$: ADX (Average Directional Index), tính từ giá cao, thấp và đóng cửa, đo lường độ mạnh của xu hướng [12].

b) Không gian hành động:

Với một cổ phiếu, không gian hành động được định nghĩa là $\{-k, \ldots, -1, 0, 1, \ldots, k\}$, trong đó k và $-k$ lần lượt biểu diễn số lượng cổ phiếu tối đa có thể mua hoặc bán ($k \le h_{\max}$, tham số giới hạn số lượng cổ phiếu trong mỗi hành động mua). Do đó, kích thước toàn bộ không gian hành động là $(2k+1)^{30}$. Không gian này được chuẩn hóa về $[-1, 1]$ để phù hợp với các thuật toán học tăng cường như A2C và PPO, vốn định nghĩa chính sách trực tiếp trên phân phối Gaussian cần đối xứng và chuẩn hóa.

V. MÔ HÌNH ENSEMBLE

A. Dữ liệu đầu vào

Đầu vào của mô hình được xây dựng từ tập dữ liệu lịch sử của 30 cổ phiếu tiêu biểu trên thị trường chứng khoán Việt Nam, được chọn ra sau quá trình huấn luyện giai đoạn đầu. Chúng tôi phát triển một hàm tiền xử lý (`preprocess_top30`) để đảm bảo rằng mỗi ngày trong tập dữ liệu đều có đủ thông tin của đúng 30 cổ phiếu được lựa chọn. Cách thực hiện như sau:

- Tại một mốc thời gian (timestamp), nếu thiếu mất dữ liệu cho một số cổ phiếu, chúng tôi bổ sung từ những ngày lân cận (trước hoặc sau) để đảm bảo tính liên tục, giả định trong khoảng thời gian này cổ phiếu không thay đổi nhiều.
- Kết quả cuối cùng được chuẩn hóa dưới dạng bảng gồm đúng 30 cổ phiếu cho mỗi ngày, được sắp xếp nhất quán theo timestamp.

Các đặc trưng đầu vào bao gồm cả dữ liệu gốc lẫn các chỉ báo đã tính toán:

- Giá, thanh khoản: *Open, High, Low, Close, Vol*
- Chỉ báo kỹ thuật (TA): *Liq, Rsi, Macd, Cci, Adx*.
- Chỉ báo rủi ro/thị trường: *Turbulence*.

Bộ dữ liệu sau tiền xử lý này được lưu thành file với định dạng `.csv` và sử dụng làm đầu vào trực tiếp cho mô hình ensemble actor-critic trong giai đoạn xây dựng chiến lược giao dịch.

B. Mô hình

Chúng tôi sử dụng ba thuật toán actor-critic để triển khai mô hình giao dịch. Ba thuật toán lần lượt là A2C, TD3, và PPO. Một chiến lược ensemble được đề xuất để kết hợp ba thuật toán này nhằm xây dựng một chiến lược giao dịch bền vững.

1) Advantage Actor Critic (A2C):

A2C [13] là một thuật toán actor-critic tiêu biểu và chúng tôi sử dụng nó như một thành phần trong chiến lược ensemble. A2C được đề xuất nhằm cải thiện policy gradient updates. A2C sử dụng một hàm advantage để giảm phương sai. Thay vì chỉ ước lượng hàm giá trị, mạng critic ước lượng hàm advantage. Do đó, việc đánh giá một hành động không chỉ phụ thuộc vào mức độ tốt của hành động mà còn xét hành động đó tốt hơn bao nhiêu so với bình thường. Điều này giúp giảm phương sai tốt và làm mô hình trở nên bền vững hơn.

A2C sử dụng nhiều bản sao của cùng một mô hình để cập nhật policy gradient với các mẫu dữ liệu khác nhau. Mỗi mô hình hoạt động độc lập để tương tác với cùng một môi trường. Trong mỗi vòng lặp, sau khi tất cả các mô hình hoàn thành việc tính gradient, A2C sử dụng một bộ điều phối để truyền trung bình các gradient từ tất cả các mô hình về một mạng toàn cục. Nhờ đó mạng toàn cục có thể cập nhật cả actor và critic. Sự tồn tại của mạng toàn cục làm tăng tính đa dạng của dữ liệu huấn luyện. Việc cập nhật gradient được đồng bộ hoá có hiệu quả chi phí hơn, nhanh hơn và hoạt động tốt với kích thước dữ liệu lớn. A2C là một mô hình phù hợp cho giao dịch chứng khoán do tính ổn định của nó.

Hàm mục tiêu cho A2C là:

$$\nabla_\theta J(\theta) \;=\; E\left[\sum_{t=1}^{T} \nabla_\theta \log \pi_\theta(a_t \mid s_t)\, A(s_t, a_t)\right] \tag{18}$$

trong đó $\pi_\theta(a_t \mid s_t)$ là mạng chính sách, và $A(s_t, a_t)$ là hàm advantage, có thể viết dưới dạng:

$$A(s_t, a_t) \;=\; Q(s_t, a_t) - V(s_t) \tag{19}$$

hoặc

$$A(s_t, a_t) \;=\; r(s_t, a_t, s_{t+1}) + \gamma V(s_{t+1}) - V(s_t) \tag{20}$$

2) Twin Delayed DDPG (TD3):

TD3 [14] được sử dụng nhằm tối đa hoá lợi nhuận đầu tư trong môi trường có không gian hành động liên tục. TD3 kế thừa ý tưởng của DDPG [15] nhưng khắc phục các nhược điểm về việc đánh giá quá lạc quan (*overestimation bias*) trong huấn luyện. Cốt lõi của TD3 là kết hợp ba kỹ thuật chính:

- *Clipped Double Q-learning*
- *Delayed Policy Update*
- *Target Policy Smoothing*

Tại mỗi bước thời gian, mô hình TD3 thực hiện một hành động a_t tại trạng thái s_t, nhận phần thưởng r_t và chuyển sang trạng thái s_{t+1}. Các chuyển tiếp (s_t, a_t, r_t, s_{t+1}) được lưu vào bộ đệm kinh nghiệm $\mathcal{R}$. Một lô mẫu gồm N chuyển tiếp được lấy ngẫu nhiên từ $\mathcal{R}$ để huấn luyện.

Giá trị mục tiêu cho critic được tính như sau:

$$y_i \;=\; r_i + \gamma \min_{j=1,2} Q'_{\theta'_{Q_j}}\Big(s_{i+1},\, \mu'_{\theta'_\mu}(s_{i+1}) + \epsilon\Big) \tag{21}$$

trong đó $\epsilon \sim \text{clip}(\mathcal{N}(0, \sigma), -c, c)$ là nhiễu Gauss bị chặn, nhằm thực hiện *target policy smoothing* và giảm dao động của hành động mục tiêu. Hai critic $Q'_{\theta'_{Q_1}}$ và $Q'_{\theta'_{Q_2}}$ được huấn luyện song song, và lấy giá trị nhỏ hơn để giảm thiên lệch đánh giá quá lạc quan.

Mạng critic được cập nhật bằng cách tối thiểu hoá hàm mất mát:

$$L(\theta_{Q_j}) \;=\; E_{(s,a,r,s')\sim\mathcal{R}}\Big[\big(y_i - Q(s, a \mid \theta_{Q_j})\big)^2\Big] \tag{22}$$

Trong khi đó, mạng actor chỉ được cập nhật với tần suất thấp hơn critic (theo cơ chế *delayed policy update*), cụ thể sau mỗi d lần cập nhật critic, actor mới được tối ưu thông qua gradient:

$$\nabla_{\theta_\mu} J(\theta_\mu) \;\approx\; E_{s\sim\mathcal{R}}\big[\nabla_a Q(s, a \mid \theta_{Q_1})\big|_{a=\mu(s)}\, \nabla_{\theta_\mu}\mu(s \mid \theta_\mu)\big] \tag{23}$$

Cuối cùng, các tham số mục tiêu được cập nhật theo phương pháp Polyak averaging:

$$\theta' \;\leftarrow\; \tau\theta + (1-\tau)\theta' \tag{24}$$

với $\tau \ll 1$.

Nhờ các cải tiến trên, TD3 tỏ ra hiệu quả hơn DDPG trong việc xử lý không gian hành động liên tục, giảm phương sai và nâng cao độ ổn định của quá trình huấn luyện, do đó phù hợp cho bài toán giao dịch cổ phiếu.

3) Proximal Policy Optimization (PPO):

PPO [16] được giới thiệu để kiểm soát cập nhật policy gradient và đảm bảo rằng chính sách mới không quá khác biệt so với chính sách trước đó. PPO cố gắng đơn giản hoá mục tiêu của Trust Region Policy Optimization (TRPO) bằng cách đưa vào một thành phần cắt (clipping) trong hàm mục tiêu.

Giả sử tỉ lệ xác suất giữa chính sách cũ và chính sách mới được biểu diễn là:

$$r_t(\theta) \;=\; \frac{\pi_\theta(a_t \mid s_t)}{\pi_{\theta_{\text{old}}}(a_t \mid s_t)} \tag{25}$$

Hàm mục tiêu surrogate cắt của PPO là:

$$J^{\text{CLIP}}(\theta) \;= E_t\Big[\min\big(r_t(\theta)\,\hat{A}_t, \\ \text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)\,\hat{A}_t\big)\Big] \tag{26}$$

trong đó $r_t(\theta)\hat{A}_t$ là mục tiêu gradient chính sách thông thường, và $\hat{A}_t$ là giá trị advantage ước lượng. Hàm $\text{clip}(r_t(\theta), 1-\epsilon, 1+\epsilon)$ giới hạn tỉ lệ $r_t(\theta)$ trong khoảng $[1-\epsilon,\ 1+\epsilon]$. Hàm mục tiêu của PPO lấy giá trị nhỏ hơn giữa mục tiêu cắt và mục tiêu thông thường. PPO ngăn chặn việc thay đổi chính sách quá lớn vượt ra ngoài khoảng cắt, từ đó cải thiện tính ổn định của quá trình huấn luyện mạng chính sách. Chúng tôi chọn PPO cho giao dịch chứng khoán vì nó ổn định, nhanh và đơn giản trong triển khai và hiệu chỉnh.

4) Chiến lược Ensemble:

Mục tiêu của chúng tôi là tạo ra một chiến lược giao dịch có độ bền cao. Vì vậy chúng tôi sử dụng một chiến lược ensemble để tự động chọn tác nhân có hiệu năng tốt nhất trong các mô hình PPO, A2C và TD3 để giao dịch dựa trên chỉ số Sharpe. Quy trình ensemble được mô tả như sau:

- **Bước 1.** Sử dụng một cửa sổ huấn luyện tăng dần có độ dài n tháng để huấn luyện lại (retrain) đồng thời ba mô hình. Trong bài báo cáo này chúng tôi huấn luyện lại ba tác nhân mỗi ba tháng.
- **Bước 2.** Xác thực cả ba mô hình bằng cách sử dụng một cửa sổ rolling 3 tháng sau cửa sổ huấn luyện để chọn tác nhân có hiệu năng cao nhất với Sharpe ratio lớn nhất [17]. Sharpe ratio được tính như:

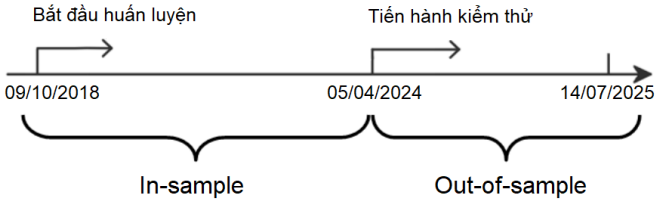
$$\text{Sharpe ratio} \;=\; \frac{\bar{r}_p - r_f}{\sigma_p} \tag{27}$$

  trong đó $\bar{r}_p$ là lợi suất kỳ vọng của danh mục, r_f là lãi suất phi rủi ro, và σ_p là độ lệch chuẩn của danh mục. Chúng tôi cũng điều chỉnh mức độ aversion rủi ro bằng cách sử dụng chỉ số turbulence trong giai đoạn xác thực.
- **Bước 3.** Sau khi chọn được tác nhân tốt nhất, chúng tôi sử dụng mô hình đó để dự báo và giao dịch cho quý tiếp theo.

Lý do đằng sau lựa chọn này là mỗi mô hình giao dịch nhạy cảm với các loại xu hướng khác nhau. Một mô hình hoạt động tốt trong xu hướng tăng nhưng hoạt động kém trong xu hướng giảm; mô hình khác có thể thích nghi tốt hơn với thị trường biến động. Chỉ số Sharpe càng cao nghĩa là tác nhân đó đã đạt được lợi nhuận tốt hơn so với mức rủi ro đã chịu. Do đó, chúng tôi chọn mô hình giao dịch có khả năng tối đa hoá lợi nhuận đã điều chỉnh theo rủi ro ngày càng gia tăng.

C. Kết quả đầu ra

Với 30 mã cổ phiếu có điểm số cao nhất từ **mô hình weight-scoring**, chúng tôi chia dữ liệu lịch sử trong giai đoạn 2018–2025 thành hai giai đoạn chính: giai đoạn *in-sample* dùng cho huấn luyện (09/10/2018–04/04/2024) và giai đoạn *out-of-sample* dùng cho kiểm thử (05/04/2024–14/07/2025), mỗi quý các mô hình sẽ được huấn luyện để lựa chọn cho quý tiếp theo, bắt đầu từ tháng 01/2024. Trong giai đoạn huấn luyện, chúng tôi tiến hành xây dựng và tối ưu hóa lần lượt 3 mô hình là A2C, PPO và TD3. Sau đó, trong giai đoạn kiểm thử, hiệu quả sinh lợi của các mô hình được đánh giá trên tập dữ liệu chưa từng thấy trước đó, nhằm phản ánh khả năng thích ứng của mô hình với động thái thị trường thực tế. Cụ thể, mô hình ensemble bắt đầu với giá trị tài khoản 1,000,000 vào ngày 05/04/2024. Đường giá trị tài khoản thể hiện những biến động đáng kể, đặc biệt trong nửa cuối năm 2024 khi có các pha suy giảm mạnh. Tuy nhiên, sau giai đoạn điều chỉnh, tài khoản dần hồi phục và bứt phá rõ rệt từ đầu năm 2025. Đến ngày 14/07/2025, giá trị tài khoản đạt khoảng 1,453,218, tương ứng với mức tăng trưởng 45,32% so với vốn ban đầu (code minh hoạ tại tệp notebook `visenet_run_model.ipynb`).



Hình 4. Phân chia dữ liệu lịch sử các mã cổ phiếu

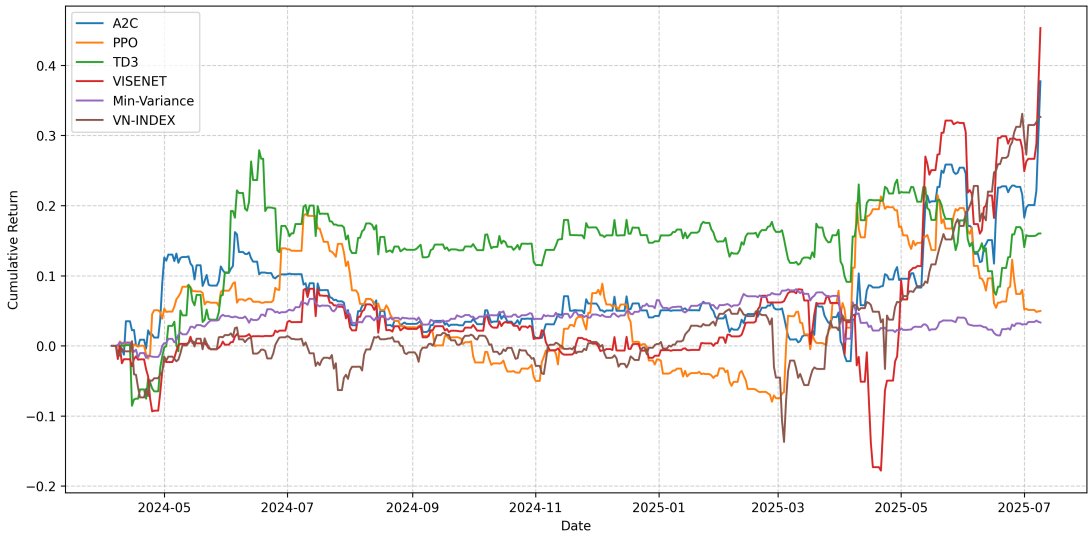
VI. ĐÁNH GIÁ VÀ PHÂN TÍCH CHIẾN LƯỢC

A. Lựa chọn mô hình cho VISENET

Từ Bảng I, có thể thấy rằng A2C đạt giá trị Sharpe ratio cao nhất là 0.314 trong giai đoạn 04/2024–07/2024, do đó mô hình này được lựa chọn để giao dịch trong quý tiếp theo (07/2024–10/2024). Tiếp theo, TD3 thể hiện ưu thế với Sharpe ratio tốt nhất (0.109) trong giai đoạn 07/2024–10/2024 và được sử dụng cho quý 10/2024–01/2025. Trong các giai đoạn tiếp theo, A2C liên tục được lựa chọn nhờ có Sharpe ratio tương đối vượt trội, mặc dù tại quý 10/2024–01/2025 giá trị này âm. Kết quả cho thấy hiệu quả của các mô hình thay đổi theo từng giai đoạn thị trường, và việc lựa chọn mô hình tối ưu cho từng quý mang lại lợi thế trong giao dịch.

Bảng I
SHARPE RATIO THEO QUÝ CỦA CÁC MÔ HÌNH

Quý giao dịch	A2C	PPO	TD3	Mô hình lựa chọn
04/2024-07/2024	**0.314**	0.104	0.302	A2C
07/2024-10/2024	0.001	-0.725	**0.109**	TD3
10/2024-01/2025	**-0.181**	-0.368	-0.184	A2C
01/2025-04/2025	**0.065**	-0.195	-0.032	A2C
04/2025-07/2025	**0.579**	-0.235	-0.012	A2C



Hình 5. Cumulative Return của VISENET với ba mô hình A2C, PPO, TD3, chiến lược phân bổ danh mục Min-Varience và chỉ số VNINDEX (Giá trị danh mục ban đầu là 1.000.000, trong giai đoạn từ 05/04/2024 đến 14/07/2025)

Bảng II
SO SÁNH HIỆU SUẤT ĐẦU TƯ

(05/04/2024-14/07/2025)	VISENET	A2C	PPO	TD3	Min-Variance	VNINDEX
Cumulative Return	**45.32%**	37.73%	4.97%	16.02%	3.32%	16.15%
Annual Return	**34.98%**	29.30%	3.97%	12.67%	2.66%	14.37%
Annual Volatility	28.17%	23.25%	23.39%	24.86%	**8.10%**	21.10%
Sharpe Ratio	1.20	**1.22**	0.28	0.60	0.36	0.74
Max Drawdown	-24.00%	-15.84%	-22.53%	-16.12%	**-6.96%**	-18.11%
Total Trade	4707	4549	4016	4644		
Win Rate	**52.86%**	48.09%	42.68%	50%	50.32%	52.67%

Để đánh giá toàn diện hiệu quả giao dịch, chúng tôi sử dụng 7 chỉ số chính:

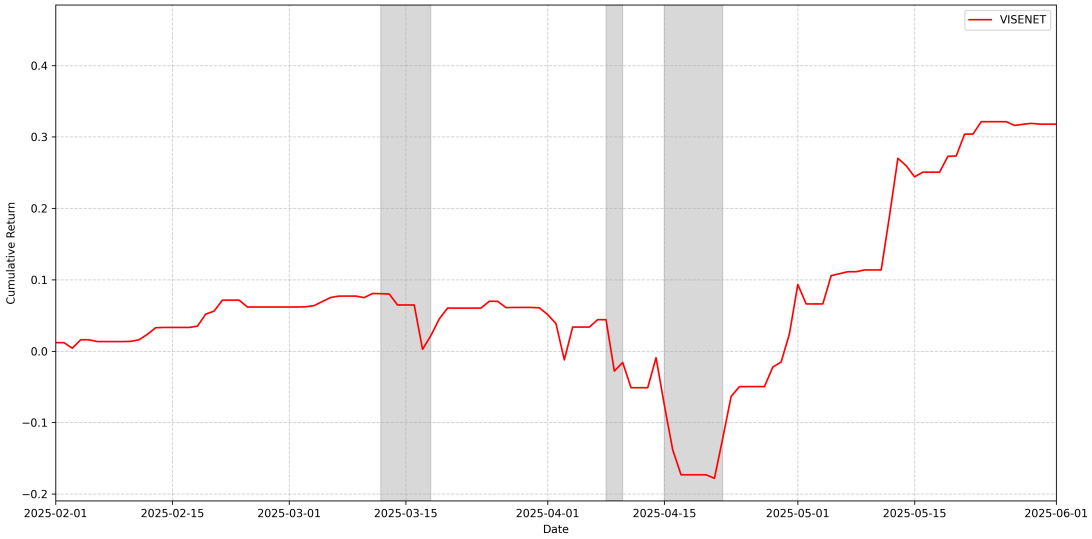
- **Cumulative Return (Tỷ suất sinh lợi tích lũy)**: đo lường tổng lợi nhuận thu được so với vốn ban đầu.
- **Annualized Return (Tỷ suất sinh lợi hàng năm)**: lợi nhuận trung bình hình học mà mô hình đạt được hàng năm.
- **Annualized Volatility (Độ biến động hàng năm)**: độ lệch chuẩn hàng năm của lợi nhuận danh mục.
- **Sharpe Ratio**: đo lường lợi nhuận điều chỉnh theo rủi ro, được tính bằng cách lấy lợi nhuận hàng năm trừ đi lãi suất phi rủi ro rồi chia cho độ biến động hàng năm.
- **Max Drawdown (Mức sụt giảm tối đa)**: phần trăm thua lỗ lớn nhất trong giai đoạn giao dịch.
- **Total Trade (Tổng số lần giao dịch)**: số lượng giao dịch được thực hiện trong toàn bộ giai đoạn.
- **Win Rate (Tỷ lệ giao dịch có lãi)**: phần trăm ngày giao dịch có lợi nhuận dương so với tổng số ngày giao dịch.

B. So sánh VISENET với Benchmark

Dựa trên Bảng II và Hình 5, có thể thấy về điểm mạnh, VISENET đạt **lợi nhuận tích lũy 45.32%** và **lợi nhuận trung bình hàng năm 34.98%**, vượt trội so với các mô hình đơn lẻ. Tỷ lệ giao dịch có lãi của VISENET đạt **52.86%**, cao hơn rõ rệt so với hầu hết mô hình khác, cho thấy khả năng ra quyết định có lợi hơn theo thời gian. Về rủi ro điều chỉnh theo lợi nhuận, VISENET đạt **Sharpe ratio 1.20**, cao hơn nhiều so với PPO, TD3 và Min-Variance [18], chỉ thua nhẹ so với A2C (1.22), điều này cho thấy ensemble vẫn duy trì hiệu quả lợi nhuận trên mỗi đơn vị rủi ro ở mức cạnh tranh.

Tuy nhiên, VISENET lại có **độ biến động hàng năm cao (28.17%)** và **max drawdown sâu (-24.00%)**, lớn hơn so với một số mô hình đơn lẻ (ví dụ A2C, TD3) và đặc biệt so với chiến lược Min-Variance (8.10% volatility; max drawdown 6.96%). Điều này cho thấy VISENET đánh đổi mức biến động cao để thu được lợi nhuận lớn hơn - hệ quả là chịu rủi ro sụt giảm sâu hơn trong các giai đoạn thị trường bất lợi. Thêm nữa, VISENET có tổng số giao dịch tương đối lớn (Total Trade = 4707), điều này có thể dẫn tới chi phí giao dịch và trượt giá cao hơn trong triển khai thực tế nếu không mô hình hóa chi phí đầy đủ.

Tóm lại, VISENET thể hiện **ưu thế tổng thể** về lợi nhuận và tỷ lệ thắng, đồng thời duy trì Sharpe ratio cao, cho thấy hiệu quả của việc kết hợp (ensemble) các tác tử. Tuy nhiên, nhà đầu tư/triển khai thực tế cần cân nhắc và bù đắp nhược điểm về biến động, drawdown và turnover - ví dụ bằng các biện pháp quản trị rủi ro (kỹ thuật *position sizing* [19], giới hạn đòn bẩy [20], cơ chế cắt lỗ [21]), tối ưu trọng số ensemble theo điều kiện thị trường để đảm bảo hiệu suất duy trì ưu thế.



Hình 6. Top 3 giai đoạn drawdown lớn nhất của VISENET trong giai đoạn out-of-sample

C. Đánh giá hiệu suất VISENET trong giai đoạn thị trường biến động mạnh

Quan sát Hình 6 cho thấy VISENET phải chịu một giai đoạn drawdown đáng kể kéo dài từ tháng 3 đến tháng 5 năm 2025. Đây là thời điểm thị trường tài chính toàn cầu chịu tác động mạnh bởi những thay đổi chính sách thương mại từ Mỹ, đặc biệt sau khi Tổng thống Donald Trump tái đắc cử và tuyên bố áp dụng các mức thuế mới đối với nhiều mặt hàng nhập khẩu, trong đó có Việt Nam. Động thái này làm gia tăng lo ngại về nguy cơ tái bùng phát căng thẳng thương mại, tác động trực tiếp đến tâm lý giới đầu tư và triển vọng xuất khẩu của các nền kinh tế mới nổi.

Tại Việt Nam, diễn biến này nhanh chóng phản ánh vào thị trường chứng khoán: **áp lực bán từ cả khối ngoại lẫn nhà đầu tư trong nước gia tăng**, đặc biệt ở các nhóm cổ phiếu xuất khẩu chủ lực như dệt may, thủy sản, gỗ và linh kiện điện tử. Chỉ số VN-Index trong giai đoạn này liên tục mất mốc hỗ trợ quan trọng, trong khi thanh khoản sụt giảm cho thấy dòng tiền có xu hướng rút lui để phòng thủ. Các quỹ ETF ngoại cũng ghi nhận rút vốn ròng, làm khuếch đại thêm áp lực giảm điểm.

Những cú sốc kết hợp từ yếu tố quốc tế (thương mại Mỹ - Việt Nam) và yếu tố nội tại (dòng vốn rút ra, tâm lý phòng thủ) đã khiến VISENET khó tránh khỏi mức drawdown trong quý II/2025. Tuy nhiên, điểm sáng là sau khi thị trường dần tiêu hóa thông tin chính sách và thích ứng với mặt bằng giá mới, VISENET thể hiện khả năng **phục hồi nhanh, đưa lợi nhuận tích lũy quay lại xu hướng tăng trưởng ổn định** từ khoảng giữa năm 2025.

Như vậy, mô hình ensemble thể hiện ưu thế vượt trội trong việc phân bổ và điều chỉnh chiến lược giao dịch, qua đó không chỉ giúp giảm thiểu những tác động tiêu cực kéo dài từ các cú sốc bên ngoài mà còn duy trì được hiệu quả đầu tư ổn định trong một môi trường thị trường đầy biến động và khó dự đoán.

VII. KẾT LUẬN

Trong bài báo cáo này, chúng tôi đã kết hợp phương pháp thiết kế thực nghiệm hỗn hợp để xác định trọng số các yếu tố ảnh hưởng và xây dựng mô hình dự báo hiệu quả danh mục, từ đó chọn ra 30 cổ phiếu tiêu biểu trên thị trường chứng khoán Việt Nam giai đoạn 2018–2025. Trên tập dữ liệu này, chúng tôi triển khai chiến lược giao dịch dựa trên học tăng cường sâu theo hướng ensemble, kết hợp ba thuật toán actor-critic (PPO, A2C, TD3). Kết quả thực nghiệm cho thấy chiến lược ensemble không chỉ khai thác được ưu điểm riêng của từng mô hình mà còn vượt trội hơn các thuật toán đơn lẻ và các chuẩn tham chiếu truyền thống về lợi nhuận đã điều chỉnh rủi ro.

Trong tương lai, để thích nghi tốt hơn với biến động thị trường, hướng nghiên cứu có thể tập trung vào việc tích hợp cơ chế quản lý rủi ro linh hoạt và bổ sung thêm đặc trưng từ dữ liệu vĩ mô hoặc tin tức tài chính.

TÀI LIỆU THAM KHẢO

[1] D.C. Montgomery, *Design and analysis of experiments*. Wiley, New York, 2012, pp. 611–622.

[2] R.H. Myers and D.C. Montgomery, *Response surface methodology*. Wiley, New York, 2008.

[3] I-Cheng Yeh and Yu-Chen Liu, "Discovering optimal weights in weighted-scoring stock-picking models: a mixture design approach," *Financial Innovation*, vol. 6, no. 41, 2020.

[4] A. Ilmanen, "Expected returns: An investor's guide to harvesting market rewards," 05 2012.

[5] Mark Kritzman and Yuanzhen Li, "Skulls, financial turbulence, and risk management," *Financial Analysts Journal*, vol. 66, 10 2010.

[6] Ashley Hill, Antonin Raffin, Maximilian Ernestus, Adam Gleave, Anssi Kanervisto, Rene Traore, Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, and Yuhuai Wu, "Stable baselines", 2018.

[7] Prafulla Dhariwal, Christopher Hesse, Oleg Klimov, Alex Nichol, Matthias Plappert, Alec Radford, John Schulman, Szymon Sidor, Yuhuai Wu, and Peter Zhokhov, "OpenAI baselines", 2017.

[8] Greg Brockman, Vicki Cheung, Ludwig Pettersson, Jonas Schneider, John Schulman, Jie Tang, and Wojciech Zaremba, "OpenAI gym," 2016.

[9] Hongyang Yang, Xiao-Yang Liu, Shan Zhong, and Anwar Walid, "Deep reinforcement learning for automated stock trading: an ensemble strategy," 2020.

[10] Terence Chong, Wing-Kam Ng, and Venus Liew, “Revisiting the performance of MACD and RSI oscillators,” *Journal of Risk and Financial Management*, vol. 7, pp. 1–12, 03 2014.

[11] Mansoor Maitah, Petr Prochazka, Michal Cermak, and Karel Sredl, “Commodity channel index: evaluation of trading rule of agricultural commodities,” *International Journal of Economics and Financial Issues*, vol. 6, pp. 176–178, 03 2016.

[12] Ikhlaas Gurrib, “Performance of the average directional index as a market timing tool for the most actively traded USD based currency pairs,” *Banks and Bank Systems*, vol. 13, pp. 58–70, 08 2018.

[13] Volodymyr Mnih, Adrià Puigdomènech Badia, Mehdi Mirza, Alex Graves, Timothy P. Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu, “Asynchronous methods for deep reinforcement learning,” *CoRR*, abs/1602.01783, 2016.

[14] Scott Fujimoto, Herke van Hoof, and David Meger, “Addressing Function Approximation Error in Actor-Critic Methods,” in *Proceedings of the 35th International Conference on Machine Learning*, PMLR 80, pp. 1587–1596, 2018.

[15] Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra, “Continuous control with deep reinforcement learning,” *arXiv preprint arXiv:1509.02971*, 2015.

[16] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov, “Proximal Policy Optimization Algorithms,” *arXiv preprint arXiv:1707.06347*, 2017.

[17] W.F. Sharpe, “The sharpe ratio,” *Journal of Portfolio Management*, 01 1994.

[18] H. Markowitz, *Portfolio Selection: Efficient Diversification of Investments*. Wiley, New York, 1959.

[19] H. Scholz, "The influence of position sizing on the performance of technical trading rules,"*Discussion Papers of the Chair of Statistics, Econometrics and Quantitative Methods*, No. 31, 2012.

[20] N.-H. Chan and H. Ma, "A framework for stop-loss analysis on trading strategies,"*The Journal of Trading*, vol. 10, no. 3, pp. 24–34, 2015.

[21] R. Baviera and S. Baldi, "Optimal statistical arbitrage with stop-loss and leverage constraints,"*arXiv preprint*, arXiv:1706.07021, 2017.