

BÁO CÁO VÒNG 3: DATA SCIENCE TALENT COMPETITION 2025

I. TỔNG QUAN

Link github: <https://github.com/m1htan/Round-3-CTE-FTU>

Hướng dẫn sử dụng: [README.md](#)

Danh sách thành viên nhóm **CTP**:

Họ và tên	Số báo danh
Nguyễn Đức Minh Tân	B0382
Phan Thị Minh Huyền	B0381
Nguyễn Như Huyền	B0528

Mục lục báo cáo:

I. TỔNG QUAN	1
II. PHẦN BÁO CÁO CHÍNH	4
Step 1: Thiết lập luồng dữ liệu real-time	4
1.1. Giới thiệu	4
1.2. Cấu hình hệ thống và môi trường	4
1.3. Chia lô và quản lý tập dữ liệu (Batching & Chunking)	4
1.4. Chuẩn hóa dữ liệu (Data Normalization)	4
1.5. Phát hiện và xác định mã hợp lệ (Valid Ticker Discovery)	5
1.6. Thu thập dữ liệu EOD lịch sử	5
1.7. Lưu trữ và quản lý đầu ra	5
1.8. Kết luận của Step 1	5
Step 2: Xây dựng logic cảnh báo	7
2.1. Giới thiệu	7
2.2. Dữ liệu đầu vào và tổ chức thư mục	7
2.3. Tiền xử lý và chuẩn hóa lược đồ	7
2.3.1. Khử trùng lặp tên cột	7

2.3.2. Chuẩn hóa mã giao dịch.....	7
2.3.3. Chuẩn hóa thời gian.....	7
2.3.4. Điều hòa tên biến (harmonization).....	8
2.4. Lớp quy tắc kỹ thuật thủ công (Handmade TI Rules).....	8
2.5. Lớp quy tắc kết hợp FI+TI (FI Rules).....	8
2.5.1. Bộ điều kiện TI cơ sở (điều chỉnh ngưỡng RSI):.....	8
2.5.2. Ràng buộc cơ bản (FI) nhằm sàng lọc chất lượng doanh nghiệp:.....	9
2.5.3. Mô-đun động lượng (momentum) tùy chọn:.....	9
2.5.4. Tổng hợp luật FI:.....	9
2.6. Hợp nhất tín hiệu và thứ tự ưu tiên.....	9
2.7. Kiểm soát chất lượng và tính bền vững.....	9
2.8. Độ phức tạp và khả năng mở rộng.....	10
2.9. Tái lập và tham số hóa thực nghiệm.....	10
2.10. Hạn chế và hướng mở rộng.....	10
2.11. Kết luận.....	10
Step 3: Triển khai hệ thống gửi cảnh báo.....	11
3.1. Giới thiệu.....	11
3.2. Cấu hình hệ thống và nguồn dữ liệu.....	11
3.3. Cơ chế nạp dữ liệu và làm mới.....	11
3.4. Bộ lọc tín hiệu và lựa chọn hiển thị.....	11
3.5. Giao diện tổng quan.....	12
3.6. Khám phá chi tiết theo mã.....	14
3.7. Cập nhật và tần suất làm mới.....	14
3.8. Kết luận.....	14
Step 4: Phân biện và kiểm thử.....	16
4.1. Cơ chế quản trị rủi ro và thoát lệnh.....	16
4.2. Mô phỏng danh mục và backtest.....	16
4.3. Thử nghiệm kịch bản cú sốc.....	16

4.4. Báo cáo và phản biện kết quả.....	17
4.5. Ứng dụng tư duy định lượng để cải tiến.....	17
4.6. Kết luận.....	17
PHỤ LỤC.....	18

II. PHẦN BÁO CÁO CHÍNH

Step 1: Thiết lập luồng dữ liệu real-time

1.1. Giới thiệu

Trong giai đoạn đầu tiên, hệ thống được thiết kế nhằm xây dựng một pipeline thu thập và chuẩn hóa dữ liệu giao dịch cuối ngày (End-of-Day - EOD). Dữ liệu EOD đóng vai trò nền tảng quan trọng cho các bước xử lý dữ liệu thời gian thực, bởi nó bảo đảm rằng các phân tích sau này được thực hiện trên cơ sở dữ liệu sạch, đồng nhất và có cấu trúc rõ ràng. Đặc biệt, việc thiết lập luồng dữ liệu EOD tạo tiền đề để mở rộng sang dữ liệu intraday và dữ liệu streaming theo thời gian thực.

1.2. Cấu hình hệ thống và môi trường

Quy trình bắt đầu bằng việc thiết lập các biến môi trường và thông tin xác thực thông qua tệp cấu hình .env. Các thông số này bao gồm tên người dùng và mật khẩu để đăng nhập API FiinQuantX, qua đó bảo đảm kết nối an toàn tới nguồn dữ liệu. Đồng thời, hệ thống định nghĩa cấu trúc thư mục để lưu trữ dữ liệu, giúp cho việc quản lý và tái sử dụng kết quả trở nên minh bạch và có tổ chức.

1.3. Chia lô và quản lý tập dữ liệu (Batching & Chunking)

Do hạn chế về quota khi truy vấn dữ liệu qua API, tập universe các mã chứng khoán được chia thành nhiều nhóm nhỏ. Cơ chế chunking và batching được hiện thực hóa nhằm xử lý song song nhiều nhóm nhưng vẫn duy trì ổn định cho toàn bộ pipeline. Mỗi lô dữ liệu được thiết kế không vượt quá 92 mã, qua đó tối ưu hoá tốc độ truy vấn đồng thời đảm bảo khả năng kiểm soát lỗi.

1.4. Chuẩn hóa dữ liệu (Data Normalization)

Một thành phần cốt lõi của pipeline là hàm chuẩn hoá dữ liệu. Dữ liệu được đồng bộ theo các quy tắc sau:

- Tất cả tên cột được chuyển về dạng chữ thường.
- Mã chứng khoán (ticker) được chuẩn hoá về dạng chữ in hoa.

- Trường thời gian được chuyển đổi về dạng epoch milliseconds, bảo đảm tính nhất quán cho các phép xử lý thời gian.
- Các trường dữ liệu bắt buộc (open, high, low, close, volume) được kiểm tra tính đầy đủ, trong khi các trường bổ sung (bu, sd, fb, fs, fn) được tạo thêm khi API không trả về.

Nhờ đó, schema dữ liệu được duy trì nhất quán xuyên suốt toàn bộ quá trình, giảm thiểu rủi ro sai lệch giữa các nguồn và giai đoạn xử lý khác nhau.

1.5. Phát hiện và xác định mã hợp lệ (Valid Ticker Discovery)

Để đảm bảo tính toàn vẹn dữ liệu, hệ thống triển khai một cơ chế phát hiện các mã giao dịch hợp lệ từ universe ban đầu. Cách tiếp cận này bao gồm việc truy vấn thử dữ liệu với tập tickers, sau đó lọc ra những mã thực sự có dữ liệu. Quá trình này được tối ưu nhờ cơ chế fallback chia nhỏ dữ liệu khi có lỗi phát sinh. Kết quả cuối cùng được lưu lại trong một tệp cache, từ đó tăng tốc cho những lần chạy kế tiếp.

1.6. Thu thập dữ liệu EOD lịch sử

Sau khi xác định được tập mã hợp lệ, hệ thống tiến hành tải dữ liệu lịch sử EOD cho mỗi batch. Dữ liệu được lấy với chu kỳ một ngày (by=1d), kéo dài khoảng 800 phiên giao dịch (tương ứng gần 3 năm). Các batch dữ liệu sau đó được gom lại thành một DataFrame duy nhất, phục vụ cho các bước phân tích thống kê và huấn luyện mô hình.

1.7. Lưu trữ và quản lý đầu ra

Dữ liệu hợp nhất cuối cùng được xuất ra tệp CSV, trong đó tên tệp chứa timestamp (UTC) để phản ánh ngày chạy pipeline. Phương pháp đặt tên có hệ thống này giúp việc truy xuất và tái sử dụng dữ liệu trở nên trực quan, đồng thời tạo thuận lợi cho việc tích hợp với các bước phân tích tiếp theo.

1.8. Kết luận của Step 1

Bước 1 không chỉ dừng lại ở việc lấy dữ liệu thô từ API FiinQuantX, mà còn xây dựng một pipeline chuẩn hoá và kiểm định dữ liệu có khả năng mở rộng cho quy mô lớn. Quy trình này bảo đảm rằng dữ liệu EOD được xử lý sạch, đồng nhất và sẵn sàng phục vụ cho luồng dữ liệu thời gian thực. Đây là nền móng quan trọng để triển khai các bước tiếp

theo, bao gồm phân tích intraday, xây dựng tín hiệu giao dịch, và phát triển các mô hình dự báo dựa trên chuỗi thời gian.

Step 2: Xây dựng logic cảnh báo

2.1. Giới thiệu

Bước 2 hiện thực hóa tầng “ra quyết định” của hệ thống qua việc tạo tín hiệu cảnh báo mua/bán từ dữ liệu đã hợp nhất. Cấu phần này kết hợp các tín hiệu kỹ thuật (TI) với các ràng buộc cơ bản (FI), chuẩn hóa lược đồ dữ liệu đầu vào, kiểm soát chất lượng dữ liệu, và hợp nhất nhiều lớp quy tắc theo một cơ chế ưu tiên rõ ràng. Kết quả được xuất song song ở hai hạt mức: (i) toàn bộ chuỗi thời gian tín hiệu, và (ii) tập cảnh báo “mới nhất theo mã” phục vụ hiển thị/dashboard.

2.2. Dữ liệu đầu vào và tổ chức thư mục

- Nguồn đầu vào: tệp dữ liệu được lấy từ data sau khi thu thập ở step 1
- Đầu ra: thư mục data/step2_signals/ lưu hai file:
 - + signals_from_merged_full_<YYYYMMDD>.csv: toàn bộ tín hiệu theo thời gian.
 - + alerts_from_merged_<YYYYMMDD>.csv: cảnh báo tại nền mới nhất mỗi mã.

2.3. Tiền xử lý và chuẩn hóa lược đồ

2.3.1. Khử trùng lặp tên cột.

Nếu phát hiện cột trùng tên, giữ cột xuất hiện đầu tiên, loại bỏ phần còn lại, bảo đảm ma trận đặc trưng “một-tên-một-cột”.

2.3.2. Chuẩn hóa mã giao dịch.

- Trích cột nhận diện từ danh sách ưu tiên: ticker, Ticker, TICKER, symbol, Symbol, SYM.
- Tạo cột chuẩn ticker = UPPER(TRIM(found)); từ chối tập vào nếu toàn rỗng.

2.3.3. Chuẩn hóa thời gian.

- Ưu tiên timestamp, dự phòng date.

- Quy đổi về epoch milliseconds qua `_to_ms(.)`, đảm bảo UTC, kiểu Int64 (nullable) để xử lý an toàn giá trị thiếu.

2.3.4. Điều hòa tên biến (harmonization).

- *Kỹ thuật (TI)*: `rsi_14` → `rsi14`, `macd_12_26` → `macd`, `macd_signal_12_26_9` → `macd_signal`, `sma_20` → `sma20`, cùng các biến SMA/EMA thường dùng.
- *Cơ bản (FI)*: ưu tiên bản điền khuyết/TTM: `PE_TTM|PE_filled` → `pe`, `PB|PB_filled` → `pb`, `ROE` → `roe`, `EPS_TTM` → `eps_ttm`, cộng thêm cặp ưu tiên cho tỷ số đòn bẩy (`Debt_to_Equity|DebtToEquity` → `dte`).
- *Hàm `_pick_series(.)`* chọn biến theo thứ tự ưu tiên và ép kiểu số an toàn, bảo vệ pipeline trước lệch chuẩn nhỏ trong quá trình hợp nhất (merged).

2.4. Lớp quy tắc kỹ thuật thủ công (Handmade TI Rules)

Quy tắc áp dụng theo từng mã sau khi sắp xếp thời gian tăng dần:

- Mua (BUY): $RSI < 30$ và giao cắt lên của MACD:

$$MACD_t > Signal_t \wedge MACD_{t-1} \leq Signal_{t-1}.$$

- Bán (SELL): $RSI > 70$ hoặc giá đóng cửa cắt xuống SMA20:

$$Close_t < SMA20_t \wedge Close_{t-1} \geq SMA20_{t-1}.$$

Các cột tối thiểu được đảm bảo tồn tại qua `ensure_cols/`bổ sung NaN, cho phép tính toán “chịu lỗi” khi một số chỉ báo vắng mặt cục bộ.

2.5. Lớp quy tắc kết hợp FI+TI (FI Rules)

2.5.1. Bộ điều kiện TI cơ sở (điều chỉnh ngưỡng RSI):

- BUY (TI): $RSI < 35$ và MACD cắt lên Signal.
- SELL (TI): $RSI > 70$ hoặc Close cắt xuống SMA20.

2.5.2. Ràng buộc cơ bản (FI) nhằm sàng lọc chất lượng doanh nghiệp:

- *Định giá hợp lý*: $0 < PE \leq 30$, $0 < PB \leq 5$.
- *Sinh lời bền vững*: $ROE \geq 10\%$.
- *Đòn bẩy an toàn*: $D/E \leq 2.0$ (nói lỏng: NaN được coi là “không vi phạm” để giảm loại bỏ giả).

2.5.3. Mô-đun động lượng (momentum) tùy chọn:

- Nếu tồn tại rank_composite: chọn mã Top 50 mỗi ngày ($rank_composite \leq 50$).
- Nếu không có, bỏ ràng buộc động lượng (mặc định true).

2.5.4. Tổng hợp luật FI:

- BUY (FI): $BUY(TI) \wedge FI_OK \wedge MOM_OK$.
- SELL (FI): $SELL(TI) \vee (\neg FI_OK)$.

2.6. Hợp nhất tín hiệu và thứ tự ưu tiên

Để xử lý xung đột giữa các tầng, hệ thống dùng hàm gộp:

- *Ưu tiên loại lệnh*: $SELL > BUY$.
- *Ưu tiên nguồn*: $fi_rule > rule_signal > signal$.

Giá trị hợp nhất được lưu tại final_signal, đóng vai trò duy nhất trong việc phát sinh cảnh báo.

2.7. Kiểm soát chất lượng và tính bền vững

- *Ràng buộc bất biến sau xử lý*: bắt buộc tồn tại ticker, timestamp (có assert/raise rõ ràng).
- *Chống sai lệch lược đồ*: khử trùng lặp tên cột, hài hòa tên biến, ép kiểu số, và group-by theo ticker để bảo toàn ranh giới thực thể.
- *Chịu lỗi dữ liệu*: mọi phép toán giao cắt/so sánh đều qua `to_numeric(errors="coerce")`; điều kiện FI cho dte chấp nhận NaN theo chính sách “không phạt thiếu dữ liệu”.

2.8. Độ phức tạp và khả năng mở rộng

- Các phép biến đổi là *tuyến tính theo số bản ghi* $O(N)$, ngoại trừ group-by theo mã có chi phí $\tilde{O}(N)$ với hằng số nhỏ.
- Hệ thống không phụ thuộc trạng thái bên ngoài ngoài tệp hợp nhất đầu vào; có thể song song hóa theo mã hoặc theo lát thời gian khi quy mô dữ liệu tăng.

2.9. Tái lập và tham số hóa thực nghiệm

Giao diện dòng lệnh cho phép:

- --input-csv: chỉ định tệp hợp nhất; nếu bỏ trống sẽ tự chọn bản mới nhất.
- --emit-latest-only: in ra bảng cảnh báo “mới nhất theo mã” trên STDOUT (phục vụ kiểm thử/hồi quy nhanh).

Mọi tệp kết quả đóng dấu ngày theo UTC (<YYYYMMDD>), thuận tiện cho so sánh phiên bản và audit.

2.10. Hạn chế và hướng mở rộng

- *Ngưỡng cố định*: các ngưỡng RSI/PE/PB/ROE/D/E hiện tĩnh; có thể tinh chỉnh thích nghi theo chế độ thị trường (regime-aware).
- *Động lượng đơn biến*: rank_composite (nếu có) đang dùng ngưỡng tuyệt đối; có thể chuyển sang phân vị theo ngành/nhóm vốn hóa.
- *Đồng bộ FI-GIÁ*: đang “as-of” theo ngày; có thể cải tiến theo lịch công bố và độ trễ báo cáo (publication lag).
- *Quản trị rủi ro*: bước này chưa bao gồm sizing/stop-loss/take-profit; nên bổ sung ở Step 3+ để khép kín vòng đời lệnh.

2.11. Kết luận

Bước 2 xây dựng một *khung cảnh báo lai TI+FI*, có kiểm soát chất lượng và xử lý xung đột đa nguồn, tạo ra tín hiệu có khả năng khái quát cao và bền vững hơn so với các quy tắc đơn lẻ. Kiến trúc này đồng thời đảm bảo tính *tái lập - mở rộng - chống nhiễu*, là nền móng để tích hợp thời gian thực và triển khai vận hành trong môi trường sản xuất.

Step 3: Triển khai hệ thống gửi cảnh báo

“streamlit run ~/Documents/GitHub/Round-3-CTE-FTU/main/step_3/streamlit_app.py”

3.1. Giới thiệu

Bước 3 trong hệ thống tập trung vào việc *triển khai giao diện cảnh báo* cho người dùng cuối. Nếu các bước trước đã đảm bảo dữ liệu sạch, chuẩn hóa (Step 1) và xây dựng logic tín hiệu (Step 2), thì bước này hiện thực hóa việc phân phối thông tin một cách trực quan, kịp thời, và dễ sử dụng. Chúng tôi sử dụng nền tảng *Streamlit*, vốn nổi tiếng trong cộng đồng khoa học dữ liệu nhờ khả năng phát triển nhanh các ứng dụng phân tích tương tác. Mục tiêu là tạo ra một *dashboard tín hiệu giao dịch* có khả năng cập nhật tự động, cung cấp cái nhìn tổng quan và chi tiết theo từng mã chứng khoán.

3.2. Cấu hình hệ thống và nguồn dữ liệu

Ứng dụng được thiết lập với cấu hình giao diện rộng (layout = “wide”) nhằm tận dụng không gian trình bày. Dữ liệu đầu vào là kết quả tín hiệu từ Step 2, lưu trong thư mục `data/step2_signals/`. Các tệp này được đánh dấu bằng ngày (`signals_from_merged_full_<YYYYMMDD>.csv`), bảo đảm khả năng truy xuất theo dòng thời gian. Ngoài ra, thư mục `logs/` được duy trì để lưu trữ nhật ký hệ thống, phục vụ công tác kiểm tra và đánh giá sau vận hành.

3.3. Cơ chế nạp dữ liệu và làm mới

Ứng dụng định nghĩa hàm `load_latest()` để:

- Truy xuất tệp tín hiệu mới nhất.
- Xác định cột tín hiệu chính, ưu tiên theo thứ tự: `final_signal` → `rule_signal` → `signal`.
- Gán nhãn mặc định “BUY” trong trường hợp tín hiệu trống hoặc rỗng, nhằm tránh mất mát thông tin.
- Bộ nhớ đệm (`@st.cache_data`) với TTL 30 giây được sử dụng để tối ưu hiệu năng và giảm tải I/O, đồng thời vẫn đảm bảo tính cập nhật liên tục.

3.4. Bộ lọc tín hiệu và lựa chọn hiển thị

Người dùng có thể lựa chọn hiển thị tín hiệu *BUY* hoặc *SELL* thông qua công cụ lọc đa chọn (st.multiselect). Dữ liệu được xử lý để chỉ giữ lại nền giao dịch mới nhất cho từng mã (groupby tail(1)), qua đó cung cấp *snapshot hiện tại của toàn bộ universe*. Đây chính là tập cảnh báo cần thiết cho việc ra quyết định nhanh.

3.5. Giao diện tổng quan

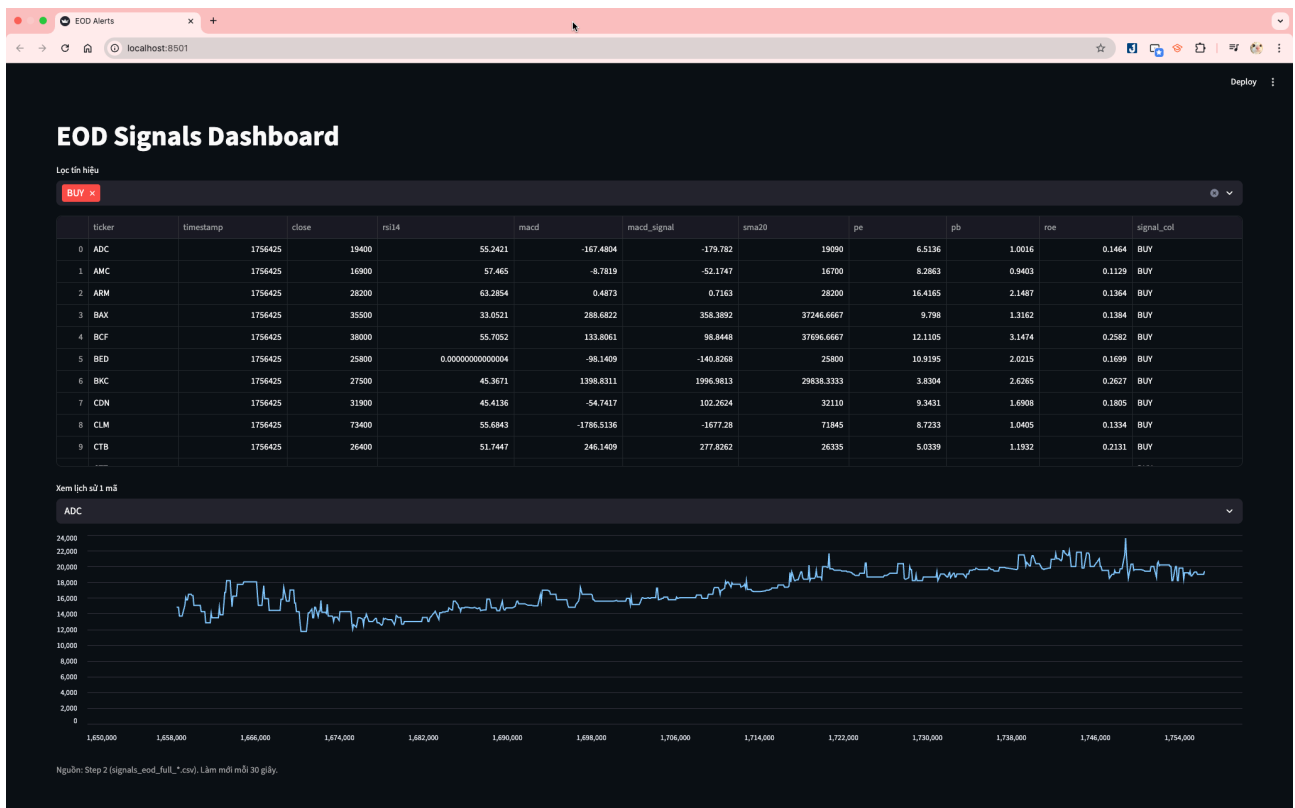
Ứng dụng hiển thị một bảng dữ liệu động bao gồm:

- Thông tin định danh: mã chứng khoán (ticker), thời gian (timestamp).
- Chỉ báo kỹ thuật: close, RSI14, MACD, MACD_signal, SMA20, EMA20, EMA50.
- Chỉ số cơ bản: OF_ratio, PE, PB, ROE, Debt-to-Equity.
- Tín hiệu hợp nhất (signal_col).

Người dùng có thể quan sát toàn cảnh universe, lọc theo tín hiệu, và sắp xếp theo mã. Điều này tái hiện chức năng bảng điều khiển (control panel) cho nhà phân tích hoặc quản lý danh mục.



Hình 1: Lọc các cổ phiếu có tín hiệu Bán



Hình 2: Lọc các cổ phiếu có tín hiệu Mua



Hình 3: Lọc các cổ phiếu theo cả 2 tín hiệu Mua và Bán



Hình 4: Báo cáo tất cả cổ phiếu

3.6. Khám phá chi tiết theo mã

Bên cạnh tổng quan, hệ thống cho phép người dùng chọn một mã cụ thể từ danh sách và xem lịch sử giá đóng cửa dưới dạng biểu đồ đường (st.line_chart). Tính năng này hỗ trợ phân tích trực quan mối quan hệ giữa tín hiệu phát ra và diễn biến giá, từ đó đánh giá chất lượng cảnh báo.

3.7. Cập nhật và tần suất làm mới

Dashboard được gắn nhãn nguồn (“Step 2 – signals_eod_full_*.csv”) và chú thích rõ *tần suất làm mới 30 giây/lần*. Điều này vừa cung cấp tính minh bạch cho người dùng, vừa bảo đảm hệ thống luôn đồng bộ với dữ liệu mới nhất mà không yêu cầu thao tác thủ công.

3.8. Kết luận

Bước 3 đã hiện thực hóa một *hệ thống gửi cảnh báo dựa trên dashboard*, đóng vai trò cầu nối giữa tầng xử lý dữ liệu (Step 1–2) và người sử dụng cuối. Việc kết hợp các yếu tố sau giúp hệ thống đáp ứng yêu cầu thực tiễn:

- *Tự động*: cập nhật tín hiệu theo thời gian gần thực (near real-time).

- *Tương tác*: bộ lọc linh hoạt và khả năng drill-down chi tiết.
- *Trực quan*: bảng dữ liệu tổng quan và biểu đồ diễn biến lịch sử.

Nhờ vậy, giải pháp này không chỉ hỗ trợ ra quyết định giao dịch nhanh chóng mà còn tăng cường khả năng giám sát, phân tích và đánh giá chiến lược đầu tư trong môi trường thực tế.

Step 4: Phản biện và kiểm thử

Trong giai đoạn này, hệ thống cảnh báo và giao dịch tự động được kiểm định thông qua một bộ khung thử nghiệm định lượng nhằm đánh giá khả năng thích ứng, mức độ ổn định và hiệu quả trong các bối cảnh thị trường khác nhau, bao gồm cả các tình huống biến động mạnh hoặc xuất hiện cú sốc thông tin. Các quy trình chính bao gồm: (i) xây dựng cơ chế quản trị rủi ro và thoát lệnh, (ii) mô phỏng hiệu quả danh mục qua backtest, (iii) thực hiện thử nghiệm kịch bản cú sốc (shock test), và (iv) tổng hợp báo cáo định lượng để phản biện các giả định chiến lược.

4.1. Cơ chế quản trị rủi ro và thoát lệnh

Một mô-đun *risk guard* được thiết kế để kiểm soát và bảo vệ danh mục trước các rủi ro bất định. Cơ chế vào/ra lệnh được chuẩn hoá theo cấu trúc tín hiệu định sẵn (BUY/SELL) và có thể cấu hình thời gian nắm giữ cố định hoặc thoát theo điều kiện dừng lỗ (stop-loss) và chốt lời (take-profit). Các mức dừng lỗ/chốt lời được định nghĩa theo tỷ lệ phần trăm, từ đó tạo ra khả năng ngăn chặn thua lỗ kéo dài và bảo toàn lợi nhuận trong bối cảnh biến động mạnh. Cách tiếp cận này giúp đảm bảo tính linh hoạt khi chuyển đổi giữa các giai đoạn thị trường tăng trưởng, điều chỉnh, hoặc khủng hoảng.

4.2. Mô phỏng danh mục và backtest

Mô-đun *backtest* mô phỏng quá trình phân bổ vốn và hiệu suất danh mục đầu tư. Vốn được phân bổ đều vào các vị thế thoả mãn tiêu chí lọc, đồng thời giới hạn số lượng mã tối đa trong cùng một thời điểm để tránh tập trung rủi ro. Các chỉ số quan trọng được tính toán bao gồm lợi nhuận trung bình, tỷ lệ giao dịch có lãi (win-rate), mức suy giảm cực đại (max drawdown), và đường cong vốn (equity curve) qua thời gian. Kết quả backtest cho phép đánh giá độ bền vững của chiến lược, cũng như xác định những giai đoạn hệ thống hoạt động yếu kém.

4.3. Thử nghiệm kịch bản cú sốc

Để phản biện khả năng ứng phó với biến động phi tuyến, mô-đun *shock test* được triển khai. Phương pháp này giả lập các cú sốc giá (ví dụ: giảm 10% toàn thị trường hoặc trong một ngày giao dịch cụ thể) và đo lường tác động đến tín hiệu và hiệu quả giao dịch.

Cách tiếp cận này phản ánh các tình huống thực tiễn như khủng hoảng ngành, tin tức bất lợi, hoặc sự kiện nâng hạng/thay đổi chính sách vĩ mô. Việc thử nghiệm với cú sốc giá giúp phát hiện nguy cơ tín hiệu sai hoặc tín hiệu đến quá muộn, từ đó cung cấp cơ sở định lượng để thiết lập các cơ chế phòng vệ bổ sung.

4.4. Báo cáo và phản biện kết quả

Mô-đun *report* tự động tổng hợp kết quả của backtest và shock test thành một báo cáo định lượng chi tiết. Báo cáo này vừa đóng vai trò phản biện các giả định của mô hình, vừa cung cấp minh chứng khách quan cho khả năng vận hành của hệ thống. Từ báo cáo, nhóm nghiên cứu có thể trả lời các câu hỏi quan trọng:

- Hệ thống có duy trì tỷ lệ thắng ổn định qua nhiều giai đoạn không?
- Có xuất hiện tình trạng tín hiệu giả (false signal) hoặc trễ quá mức trong bối cảnh thị trường bất thường không?
- Các mức dừng lỗ và chốt lời đã đủ bảo vệ vốn trước kịch bản khủng hoảng chưa?

4.5. Ứng dụng tư duy định lượng để cải tiến

Kết quả kiểm thử cho thấy tiềm năng ứng dụng tư duy định lượng vào việc tinh chỉnh chiến lược. Ví dụ, việc điều chỉnh trọng số các chỉ số định giá và động lượng trong hàm tổng hợp có thể cải thiện tỷ lệ thắng, trong khi các mô hình quản lý rủi ro phi tuyến (như dynamic stop-loss theo độ biến động) giúp giảm thiểu drawdown. Hơn nữa, kết quả shock test gợi ý rằng việc tích hợp thêm bộ lọc theo chỉ báo vĩ mô hoặc tâm lý thị trường sẽ nâng cao khả năng dự báo và bảo vệ rủi ro.

4.6. Kết luận

Bộ khung kiểm thử tại Step 4 không chỉ đánh giá hiệu quả định lượng của hệ thống mà còn đóng vai trò phản biện, nhằm phát hiện hạn chế và hướng đến các cải tiến. Việc kết hợp cơ chế quản trị rủi ro, mô phỏng danh mục, thử nghiệm cú sốc và báo cáo định lượng đã hình thành một quy trình toàn diện, đảm bảo hệ thống có thể vận hành liên tục, thích ứng linh hoạt và duy trì độ tin cậy trong môi trường thị trường biến động cao.

PHỤ LỤC

Phụ lục A - Tóm tắt quy tắc (dạng công thức)

- **BUY_handmade:** $RSI < 30 \wedge \text{MACD crosses up Signal}$
- **SELL_handmade:** $RSI > 70 \vee \text{Close crosses down SMA20}$
- **BUY_FI:** $BUY_TI \wedge (0 < PE \leq 30) \wedge (0 < PB \leq 5) \wedge (ROE \geq 0.10) \wedge (D/E \leq 2.0 \text{ or NaN}) \wedge MOM_OK$
- **SELL_FI:** $SELL_TI \vee \neg FI_OK$
- Ưu tiên hợp nhất: $SELL > BUY, fi_rule > rule_signal > signal.$