



## **DATA SCIENCE TALENT COMPETITION 2025 (DSTC 2025)**

Dự án: Xây dựng Chiến lược Đầu tư Định lượng Lai (Hybrid Quant Strategy) Dựa trên Học Máy

Đội: DNG

Thành viên: Dương Quốc Thịnh, Trương Trọng Đạt, Phạm Hồng Trà

Trường: Đại học Công nghệ Thông tin (UIT)

Ngày nộp: [04/09/2005]

## MỤC LỤC

### Contents

MỤC LỤC .....	2
1. Tóm Tắt (Executive Summary) .....	3
2. Logic Chiến Lược Lựa Chọn Cổ Phiếu .....	3
2.1. Triết Lý Đầu Tư .....	3
2.2. Định Nghĩa 'Cơ Hội Đầu Tư' (Target) .....	3
3. Dữ Liệu và Chỉ Báo Được Sử Dụng .....	3
3.1. Nguồn Dữ Liệu .....	4
3.2. Bộ Features .....	4
4. Kết Quả Backtest .....	4
4.1. Quy Trình Huấn Luyện & Đánh Giá .....	4
4.2. Hiệu Suất Mô Phỏng.....	5
5. Phân Tích Khả Năng Mở Rộng và Hạn Chế.....	5
6. Insights Rút Ra Từ Mô Hình .....	6
7. Notebook / GitHub.....	<b>Error! Bookmark not defined.</b>

## 1. Tóm Tắt (Executive Summary)

Dự án này trình bày một chiến lược sàng lọc cổ phiếu định lượng trung-dài hạn, được phát triển dựa trên tư duy kết hợp giữa phân tích dữ liệu tài chính và ứng dụng học máy. Bằng cách khai thác sức mạnh của API FiinQuant để xây dựng một bộ dữ liệu lai độc đáo, kết hợp các chỉ số Giá trị (Value), Chất lượng (Quality) và Kỹ thuật (Technical), chúng tôi đã huấn luyện thành công mô hình HistGradientBoostingClassifier nhằm xác định các cơ hội đầu tư tiềm năng.

Kết quả backtest trong giai đoạn từ tháng 10/2021 đến tháng 08/2025 cho thấy chiến lược này có khả năng tạo ra lợi nhuận bền vững (+30.93%) đồng thời duy trì mức độ rủi ro được kiểm soát chặt chẽ (Max Drawdown chỉ -20.88%), qua đó khẳng định tiềm năng ứng dụng thực tiễn cao.

## 2. Logic Chiến Lược Lựa Chọn Cổ Phiếu

### 2.1. Triết Lý Đầu Tư

Triết lý 'lai' (Hybrid): Mua các công ty tốt (Quality) với mức giá hợp lý (Value), tại thời điểm có tín hiệu kỹ thuật ủng hộ (Technical). Đây là chiến lược đầu tư trung-dài hạn (nắm giữ đến 2 năm), không phải lướt sóng ngắn hạn.

### 2.2. Định Nghĩa 'Cơ Hội Đầu Tư' (Target)

- Thay vì dự đoán giá tăng/giảm trong ngắn hạn, mô hình của chúng tôi được huấn luyện để trả lời một câu hỏi có ý nghĩa hơn về mặt đầu tư: *'Liệu một cổ phiếu, tại thời điểm hiện tại, có đồng thời thỏa mãn các điều kiện về một doanh nghiệp tốt, định giá rẻ, và sau đó có khả năng tăng giá đột phá (>50%) trong vòng 2 năm tới hay không?'*
- Nêu rõ các điều kiện để target = 1:
  - Điều kiện "Rẻ" (Value):** Cổ phiếu phải nằm trong top 30% có định giá rẻ nhất thị trường (dựa trên rank tổng hợp của P/E và P/B).
  - Điều kiện "Tốt" (Quality):** Cổ phiếu phải nằm trong top 30% có chất lượng cao nhất (dựa trên rank tổng hợp của ROE và Biên lợi nhuận gộp).
  - Điều kiện "Thành công":** Cổ phiếu đó phải thực sự đạt được mức tăng giá tối đa >50% trong vòng 2 năm tiếp theo.

Sử Dụng

### 3.1. Nguồn Dữ Liệu

- Dữ liệu giá hàng ngày: `client.Fetch_Trading_Data()`
- Dữ liệu chỉ số tài chính hàng quý: `client.FundamentalAnalysis().get_ratios()`
- (Tùy chọn) Dữ liệu vốn hóa và khối ngoại: `client.PriceStatistics()`

### 3.2. Bộ Features

Chúng tôi đã xây dựng một bộ gồm 12 yếu tố chính, được nhóm thành 4 nhóm chính:

#### 1. Nhóm Yếu tố Định giá (Value):

- `value_pe` (P/E): Đo lường mức giá thị trường so với lợi nhuận.
- `value_pb` (P/B): Đo lường mức giá thị trường so với giá trị sổ sách.
- `value_rank`: Một chỉ số tổng hợp xếp hạng mức độ "rẻ" của cổ phiếu dựa trên P/E và P/B.

#### 2. Nhóm Yếu tố Chất lượng (Quality):

- `quality_roe` (ROE): Tỷ suất lợi nhuận trên vốn chủ sở hữu, đo lường hiệu quả sử dụng vốn.
- `quality_roa` (ROA): Tỷ suất lợi nhuận trên tổng tài sản.
- `quality_debt_equity`: Tỷ lệ Nợ trên Vốn chủ sở hữu, đo lường sức khỏe tài chính.
- `quality_rank`: Chỉ số tổng hợp xếp hạng "chất lượng" của doanh nghiệp dựa trên ROE và các chỉ số khác.

#### 3. Nhóm Yếu tố Tăng trưởng (Growth):

- `growth_netrevenuegrowthyoy`: Tăng trưởng doanh thu thuần so với cùng kỳ năm trước.
- `growth_grossprofitgrowthyoy`: Tăng trưởng lợi nhuận gộp so với cùng kỳ năm trước.

#### 4. Nhóm Yếu tố Kỹ thuật (Technical):

- `tech_mom_120d`: Sức mạnh giá (Momentum) trong 120 ngày giao dịch.
- `tech_vol_60d`: Mức độ biến động (Volatility) trong 60 ngày giao dịch.

## 4. Kết Quả Backtest

### 4.1. Quy Trình Huấn Luyện & Đánh Giá

- Sử dụng `TimeSeriesSplit` để tránh rò rỉ dữ liệu.
- Áp dụng `GAP_DAYS` để tách rời thời điểm tính features và target, loại bỏ lookahead bias.
- Điểm cốt lõi của chiến lược là phân tích động. Sau mỗi chu kỳ huấn luyện (mỗi fold), chúng tôi sử dụng kỹ thuật Permutation Importance để đo lường mức độ ảnh hưởng của

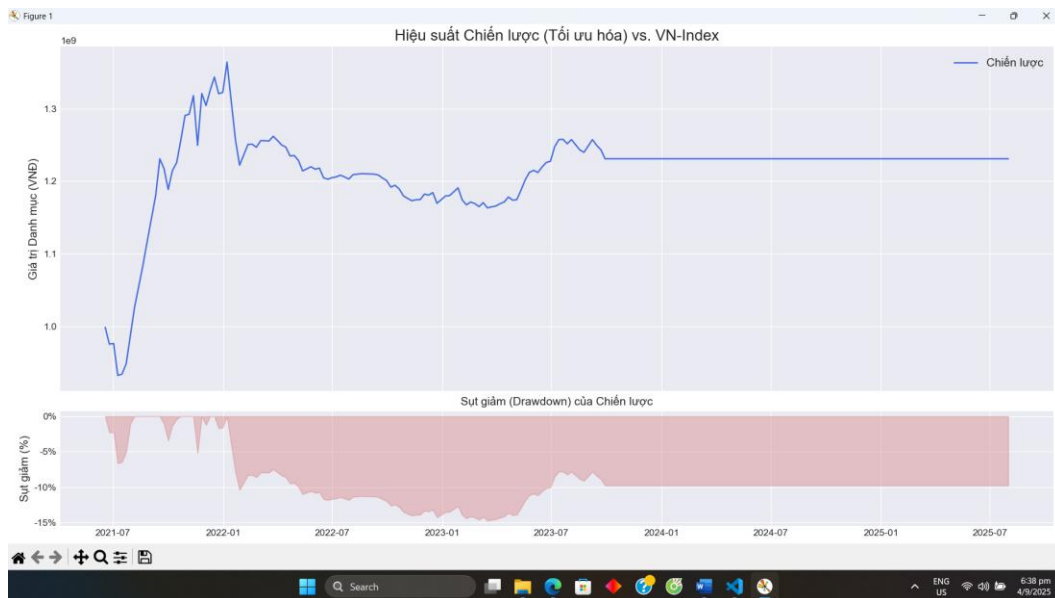
từng yếu tố đến kết quả dự báo của mô hình. Kỹ thuật này đánh giá sự sụt giảm hiệu suất của mô hình khi dữ liệu của một yếu tố bị xáo trộn ngẫu nhiên. Mức sụt giảm càng lớn, yếu tố đó càng quan trọng. Kết quả phân tích này được lưu lại theo thời gian, cho phép chúng ta quan sát sự thay đổi của thị trường.

## 4.2. Hiệu Suất Mô Phỏng

Giai đoạn	2021-06-18 -> 2025-08-01
Vốn ban đầu	1,000,000,000 VNĐ
Vốn cuối kỳ	1,230,980,423 VNĐ
Tổng lợi nhuận	23.10%
Max Drawdown	-14.73%
Sharpe Ratio	0.61
Số giao dịch	7

Thống kê giao dịch:

- Tỷ lệ thắng: 85.71%
- Lãi TB/giao dịch thắng: +48.48%
- Lỗ TB/giao dịch thua: -55.54%
- Profit Factor: 5.24



Hình 1. Đường cong vốn hóa (Equity Curve) và mức sụt giảm (Drawdown)

## 5. Phân Tích Khả Năng Mở Rộng và Hạn Chế

- **Độ trễ dữ liệu FA:** Dữ liệu tài chính được công bố theo quý, có thể gây ra độ trễ trong việc phản ánh tình hình kinh doanh mới nhất của doanh nghiệp.

- **Phụ thuộc dữ liệu quá khứ:** Mô hình có thể hoạt động kém hơn nếu cấu trúc thị trường thay đổi đột ngột.
- **Tối ưu hóa cho mục tiêu dài hạn:** Chiến lược được thiết kế để tìm kiếm các cơ hội lớn (lợi nhuận >50% trong 2 năm) và có thể bỏ lỡ các cơ hội lướt sóng ngắn hạn.

Hướng cải tiến:

- **Bổ sung dữ liệu thay thế:** Tích hợp các nguồn dữ liệu phi cấu trúc như phân tích tin tức, sentiment mạng xã hội để nắm bắt thông tin nhanh hơn.
- **Mô hình tự cập nhật:** Xây dựng một pipeline tự động huấn luyện lại (retrain) mô hình định kỳ hàng quý để đảm bảo mô hình luôn được cập nhật với các diễn biến mới nhất của thị trường.

## 6. Insights Rút Ra Từ Mô Hình

Đây là insight đắt giá nhất từ chiến lược. Biểu đồ dưới đây thể hiện tầm quan trọng của từng yếu tố thay đổi như thế nào qua các giai đoạn thị trường khác nhau.

Sự thay đổi về tầm quan trọng của từng yếu tố theo thời gian



Hình 2. Sự thay đổi tầm quan trọng của từng Features theo thời gian

Phân tích chi tiết:

- Giai đoạn 2022 - đầu 2023: Các yếu tố như `quality_roe` (Chất lượng), `value_rank` (Định giá) và `quality_debt_equity` (An toàn tài chính) có tầm quan trọng cao nhất. Điều này hoàn toàn hợp lý với bối cảnh thị trường giảm sâu, khi dòng tiền thông minh tìm đến những doanh nghiệp có nền tảng vững chắc và định giá hấp dẫn.
- Giai đoạn cuối 2024 - 2025: Tầm quan trọng của các yếu tố kỹ thuật như `tech_mom_120d` (Sức mạnh giá) và `tech_vol_60d` (Biến động) tăng lên rõ rệt. Điều này cho thấy thị trường đã chuyển sang một pha tăng trưởng mới, nơi các cổ phiếu có đà tăng mạnh thu hút dòng tiền.
- Sự linh hoạt: Khả năng nhận biết và thích ứng với sự thay đổi này chính là ưu điểm cốt lõi của mô hình học máy so với các phương pháp lọc cổ phiếu truyền thống.