Giải thích các trường dữ liệu

Tóm tắt các Tệp:

**CMG\_foreign\_detail\_transactions\_processed.csv & FPT\_foreign\_detail\_transactions\_processed.csv**: Các tệp này chứa thông tin chi tiết về giao dịch của nhà đầu tư nước ngoài (khối ngoại) đối với riêng CMG và FPT, bao gồm cả giao dịch khớp lệnh trên sàn và giao dịch thỏa thuận (put-through).

**foreigner\_matched\_order\_trading\_statistics\_processed.csv**: Tệp này cung cấp cái nhìn tổng quan toàn thị trường về thống kê giao dịch của khối ngoại đối với tất cả các mã cổ phiếu trong một ngày cụ thể. Lưu ý quan trọng: tệp này chỉ hiển thị giao dịch khớp lệnh, không bao gồm giao dịch thỏa thuận.

**TARGET CMG\_detail\_transactions\_processed.csv & TARGET FPT\_detail\_transactions\_processed.csv**: Các tệp này chứa dữ liệu giao dịch chung cho CMG và FPT, bao gồm tổng khối lượng và giá trị giao dịch (trong nước + nước ngoài), vốn hóa thị trường và giá đóng cửa cùng với sự thay đổi giá. Đây là các tệp 'đích' cho thấy kết quả (giá đóng cửa).

Giải thích các Trường Dữ liệu theo Loại Tệp:

1. **CMG\_foreign\_detail\_transactions\_processed.csv & FPT\_foreign\_detail\_transactions\_processed.csv**

Hai tệp này có cấu trúc giống nhau và cung cấp chi tiết giao dịch khối ngoại cho một mã cổ phiếu cụ thể:

* StockID: Mã cổ phiếu (CMG hoặc FPT).
* Date: Ngày giao dịch (định dạng: DD/MM/YYYY).
* Foreign Investor Room: Tổng số lượng cổ phiếu mà nhà đầu tư nước ngoài được phép sở hữu tối đa theo quy định. Đây là giới hạn (room).
* Remaining Room Shares: Số lượng cổ phiếu còn khả dụng cho nhà đầu tư nước ngoài mua trong giới hạn (Foreign Investor Room - Sở hữu hiện tại của khối ngoại).
* Remaining Room %: Tỷ lệ phần trăm của tổng số cổ phiếu đang lưu hành của công ty mà nhà đầu tư nước ngoài còn có thể mua. (Cách tính: Remaining Room Shares / Tổng số cổ phiếu đang lưu hành). Lưu ý: Cơ sở tính toán KHÔNG phải là Foreign Investor Room. Tỷ lệ phần trăm thấp có nghĩa là giới hạn sở hữu nước ngoài sắp đạt mức tối đa.
* Matched Buy Volume: Số lượng cổ phiếu được khối ngoại mua thông qua hệ thống khớp lệnh thông thường của sàn giao dịch.
* Matched Buy Value: Tổng giá trị (tính bằng triệu VND, ví dụ: 4.611 = 4.611.000.000 VND) của cổ phiếu được khối ngoại mua qua khớp lệnh.
* Matched Sell Volume: Số lượng cổ phiếu được khối ngoại bán qua khớp lệnh.
* Matched Sell Value: Tổng giá trị (triệu VND) của cổ phiếu được khối ngoại bán qua khớp lệnh.
* Negotiated Buy Volume: Số lượng cổ phiếu được khối ngoại mua qua giao dịch thỏa thuận (put-through deals, ngoài sổ lệnh).
* Negotiated Buy Value: Tổng giá trị (triệu VND) của cổ phiếu được khối ngoại mua qua giao dịch thỏa thuận.
* Negotiated Sell Volume: Số lượng cổ phiếu được khối ngoại bán qua giao dịch thỏa thuận.
* Negotiated Sell Value: Tổng giá trị (triệu VND) của cổ phiếu được khối ngoại bán qua giao dịch thỏa thuận.
* Total Buy Volume: Tổng số lượng cổ phiếu khối ngoại đã mua (Matched Buy Volume + Negotiated Buy Volume).
* Total Buy Value: Tổng giá trị (triệu VND) cổ phiếu khối ngoại đã mua (Matched Buy Value + Negotiated Buy Value).
* Total Sell Volume: Tổng số lượng cổ phiếu khối ngoại đã bán (Matched Sell Volume + Negotiated Sell Volume).
* Total Sell Value: Tổng giá trị (triệu VND) cổ phiếu khối ngoại đã bán (Matched Sell Value + Negotiated Sell Value).

1. **foreigner\_matched\_order\_trading\_statistics\_processed.csv**

Tệp này cung cấp cái nhìn tổng quan về giao dịch khớp lệnh của khối ngoại trên toàn thị trường:

* No: Số thứ tự dòng.
* Date: Ngày giao dịch.
* StockID: Mã cổ phiếu.
* Foreign Room: Số lượng cổ phiếu tối đa khối ngoại được sở hữu (giới hạn).
* Foreign Ownership %: Tỷ lệ phần trăm hiện tại của công ty đang được sở hữu bởi khối ngoại.
* Remaining Room Volume: Số lượng cổ phiếu còn lại cho khối ngoại (Foreign Room - Sở hữu hiện tại).
* Remaining Room %: Tỷ lệ phần trăm của tổng room khối ngoại còn lại. (Ví dụ: nếu giới hạn là 49% và sở hữu hiện tại là 40%, thì tỷ lệ này là ((49-40)/49)%). Giá trị 0% nghĩa là room đã hết.
* Matched Buy Volume: Số lượng cổ phiếu khối ngoại mua qua khớp lệnh (giống trong tệp chi tiết).
* Matched Buy %: Matched Buy Volume của khối ngoại chiếm bao nhiêu phần trăm tổng khối lượng khớp lệnh của mã đó trong ngày (bao gồm cả giao dịch trong nước). Cho biết mức độ chi phối của người mua nước ngoài trong giao dịch khớp lệnh.
* Matched Sell Volume: Số lượng cổ phiếu khối ngoại bán qua khớp lệnh (giống trong tệp chi tiết).
* Matched Sell %: Matched Sell Volume của khối ngoại chiếm bao nhiêu phần trăm tổng khối lượng khớp lệnh. Cho biết mức độ chi phối của người bán nước ngoài.
* Matched Buy Value: Giá trị khối ngoại mua qua khớp lệnh (triệu VND).
* Matched Buy Value %: Giá trị mua của khối ngoại chiếm bao nhiêu phần trăm tổng giá trị khớp lệnh.
* Matched Sell Value: Giá trị khối ngoại bán qua khớp lệnh (triệu VND).
* Matched Sell Value %: Giá trị bán của khối ngoại chiếm bao nhiêu phần trăm tổng giá trị khớp lệnh.
* Net Volume: Khối lượng mua/bán ròng của khối ngoại qua khớp lệnh (Matched Buy Volume - Matched Sell Volume). Dương = mua ròng, âm = bán ròng.
* Net Value: Giá trị mua/bán ròng của khối ngoại qua khớp lệnh (Matched Buy Value - Matched Sell Value, triệu VND).
* Exchange: Sàn giao dịch nơi cổ phiếu niêm yết (HOSE, HNX, UPCoM).

1. **TARGET CMG\_detail\_transactions\_processed.csv & TARGET FPT\_detail\_transactions\_processed.csv**

Các tệp này hiển thị dữ liệu thị trường chung và kết quả giá:

* StockID: Mã cổ phiếu (CMG hoặc FPT).
* Date: Ngày giao dịch.
* Total Volume: Tổng số lượng cổ phiếu được giao dịch trong ngày (trong nước + nước ngoài, khớp lệnh + thỏa thuận).
* Total Value: Tổng giá trị (triệu VND) của tất cả cổ phiếu được giao dịch.
* Market Cap: Vốn hóa thị trường (triệu VND) vào cuối ngày (Tổng số cổ phiếu đang lưu hành \* Closing Price).
* Closing Price: Giá đóng cửa của cổ phiếu vào ngày đó (tính bằng VND). Đây là mức giá cuối cùng mà cung và cầu gặp nhau khi kết thúc phiên (thường được xác định trong phiên ATC).
* Price Change: Mức thay đổi tuyệt đối của giá đóng cửa so với ngày giao dịch trước đó (tính bằng VND).
* Price Change %: Mức thay đổi phần trăm của giá đóng cửa so với ngày giao dịch trước đó.
* Matched Volume: Tổng số lượng cổ phiếu được giao dịch qua hệ thống khớp lệnh (trong nước + nước ngoài).
* Matched Value: Tổng giá trị (triệu VND) của các giao dịch khớp lệnh.

Tuyệt vời! Dựa trên tất cả các nội dung đã thảo luận, đây là hướng đi chi tiết và có cấu trúc để bạn xây dựng hệ thống mô hình dự đoán phức hợp cho cổ phiếu (ví dụ: trên sàn HNX), kết hợp dự đoán hướng đi, giá trị và rủi ro thua lỗ khi dự đoán hướng đi sai:

**Mục Tiêu Tổng Thể:** Xây dựng một hệ thống hỗ trợ ra quyết định đầu tư cổ phiếu, bao gồm:

1. Dự đoán hướng đi (Tăng/Giảm).
2. Dự đoán mức độ thay đổi giá (Giá cụ thể).
3. Ước tính/Dự đoán rủi ro thua lỗ tiềm năng *nếu dự đoán hướng đi ban đầu sai*.

**Các Bước Xây Dựng Hệ Thống:**

**Bước 1: Nền Tảng - Dữ Liệu và Môi Trường**

1. **Xác định Cổ Phiếu/Thị Trường:** Chọn các cổ phiếu cụ thể trên HNX hoặc toàn bộ chỉ số HNX-Index mà bạn muốn phân tích.
2. **Thu Thập Dữ Liệu Toàn Diện:**
   * **Giá & Khối lượng:** Lịch sử OHLCV (Mở, Cao, Thấp, Đóng, Khối lượng) với tần suất phù hợp (ngày, giờ...).
   * **Chỉ Báo Kỹ Thuật:** Tính toán các chỉ báo phổ biến (SMA, EMA, MACD, RSI, Bollinger Bands, ATR, Stochastic Oscillator...).
   * **Dữ Liệu Thị Trường:** Chỉ số HNX-Index, VN-Index, các chỉ số ngành liên quan, dữ liệu giao dịch khối ngoại...
   * **(Tùy chọn) Dữ Liệu Cơ Bản:** EPS, P/E, P/B, ROE... (thường cập nhật theo quý/năm).
   * **(Nâng cao) Dữ Liệu Phi Cấu Trúc:** Tin tức, phân tích thị trường (yêu cầu xử lý ngôn ngữ tự nhiên).
3. **Làm Sạch & Tiền Xử Lý:**
   * Xử lý dữ liệu thiếu, dữ liệu ngoại lai.
   * Đồng bộ hóa các chuỗi thời gian.
   * Chuẩn hóa hoặc Scaling dữ liệu (ví dụ: MinMaxScaler, StandardScaler) nếu thuật toán yêu cầu.
4. **Feature Engineering:**
   * Tạo các đặc trưng mới từ dữ liệu gốc (ví dụ: độ dốc của đường MA, khoảng cách giá so với MA, độ rộng Bollinger Bands...).
   * Tạo các đặc trưng trễ (lag features) của giá, khối lượng, chỉ báo.
5. **Thiết Lập Môi Trường:** Sử dụng Python với các thư viện như Pandas, NumPy, Scikit-learn, Statsmodels, và các thư viện cho mô hình cụ thể (TensorFlow/Keras, PyTorch, XGBoost, LightGBM).

**Bước 2: Xây Dựng Các Mô Hình Thành Phần**

1. **Model 1: Dự Đoán Hướng Đi (Classification)**
   * **Mục tiêu:** Dự đoán giá đóng cửa ngày T+1 (hoặc T+n) sẽ Tăng hay Giảm so với giá đóng cửa ngày T. (Có thể thêm lớp Đi ngang nếu cần).
   * **Target Variable:** Tạo cột target: 1 nếu Giá(T+1) > Giá(T), 0 nếu Giá(T+1) <= Giá(T) (điều chỉnh logic tùy theo định nghĩa "Tăng/Giảm").
   * **Thuật toán:** Logistic Regression, SVM, Random Forest, Gradient Boosting (XGBoost, LightGBM), Neural Networks đơn giản.
   * **Đánh giá:** Accuracy, Precision (cho tín hiệu Mua/Bán), Recall, F1-Score, AUC. Tập trung vào Precision nếu muốn tín hiệu chắc chắn hơn.
2. **Model 2: Dự Đoán Giá Trị (Regression)**
   * **Mục tiêu:** Dự đoán giá đóng cửa cụ thể ngày T+1 (hoặc T+n), hoặc dự đoán % thay đổi giá.
   * **Target Variable:** Giá đóng cửa T+1, hoặc (Giá(T+1) - Giá(T)) / Giá(T).
   * **Thuật toán:** Linear Regression (ít hiệu quả), SVR, Random Forest Regressor, Gradient Boosting Regressor, LSTM (nếu muốn tập trung vào tính chuỗi thời gian).
   * **Đánh giá:** MAE (Sai số tuyệt đối trung bình), MSE/RMSE (Sai số bình phương trung bình/Căn bậc hai), R-squared.
3. **Model 3: Ước Tính/Dự Đoán Rủi Ro Sai Lệch Hướng Đi**
   * **Mục tiêu:** Ước tính mức lỗ tiềm năng (%) nếu quyết định Mua (dựa trên dự đoán "Tăng" của Model 1) hoặc Bán (dựa trên dự đoán "Giảm" của Model 1) nhưng thực tế giá đi ngược lại.
   * **Phương pháp chính (Dựa trên Lịch sử):**
     + **Bước 3.1:** Chạy backtest riêng cho Model 1 trên dữ liệu lịch sử (out-of-sample).
     + **Bước 3.2:** Lọc ra các trường hợp Model 1 dự đoán sai:
       - Sai khi dự đoán Tăng: Model 1 dự đoán "Tăng", nhưng Giá(T+1) <= Giá(T). Tính mức lỗ thực tế (Giá(T+1) - Giá(T)) / Giá(T) (sẽ là số âm hoặc 0).
       - Sai khi dự đoán Giảm: Model 1 dự đoán "Giảm", nhưng Giá(T+1) > Giá(T). Tính mức "lỗ" cơ hội hoặc mức tăng thực tế (Giá(T+1) - Giá(T)) / Giá(T) (sẽ là số dương). *Lưu ý: Nếu bạn short sell thì đây mới là lỗ thực sự.* Nếu chỉ không mua/bán thì đây là "lỗ cơ hội". Cần xác định rõ bạn đo lường cái gì. Giả sử bạn đo lỗ khi mua sai và lỗ khi bán khống sai.
     + **Bước 3.3:** Phân tích phân phối các mức lỗ này: Tính toán Mức lỗ trung bình khi sai, Mức lỗ tối đa khi sai, Phân vị lỗ khi sai (ví dụ: lỗ ở phân vị 95%).
   * **(Nâng cao) Phương pháp Dự Đoán:**
     + Sử dụng các mức lỗ tính ở Bước 3.3 làm target variable cho một mô hình hồi quy (Model 3).
     + Features cho Model 3: Xác suất/độ tin cậy của Model 1, các chỉ báo biến động (ATR), trạng thái thị trường...
     + Mục tiêu là dự đoán *trước* mức lỗ tiềm năng nếu Model 1 sai trong lần dự đoán *tiếp theo*.

**Bước 3: Kết Hợp Các Mô Hình và Xây Dựng Logic Ra Quyết Định**

1. **Chạy Đồng Thời:** Với mỗi điểm dữ liệu mới (ví dụ: cuối ngày T), chạy cả 3 mô hình:
   * Model 1: Dự đoán hướng (Tăng/Giảm).
   * Model 2: Dự đoán giá T+1 (hoặc % thay đổi).
   * Model 3: Ước tính/Dự đoán mức lỗ tiềm năng nếu Model 1 sai (dựa trên phân tích lịch sử hoặc mô hình dự đoán rủi ro).
2. **Thiết Kế Logic Quyết Định (Ví dụ):**
   * **Tín Hiệu Mua:**
     + Model 1 dự đoán "Tăng" (có thể yêu cầu xác suất > ngưỡng, ví dụ > 60%).
     + Model 2 dự đoán % tăng > ngưỡng lợi nhuận tối thiểu (ví dụ > 1%, sau khi trừ phí).
     + Model 3 cho thấy Mức lỗ tiềm năng nếu sai < Ngưỡng chấp nhận rủi ro của bạn (ví dụ: < 3%).
     + => **Quyết định:** Mua.
   * **Tín Hiệu Bán (hoặc Bán Khống):**
     + Model 1 dự đoán "Giảm" (xác suất > ngưỡng).
     + Model 2 dự đoán % giảm > ngưỡng lợi nhuận tối thiểu.
     + Model 3 cho thấy Mức lỗ tiềm năng (nếu bán khống sai) < Ngưỡng chấp nhận rủi ro.
     + => **Quyết định:** Bán/Bán khống.
   * **Không Giao Dịch:** Nếu bất kỳ điều kiện nào ở trên không được thỏa mãn.
3. **Điều Chỉnh Quyết Định Nâng Cao:**
   * **Quản Lý Vốn:** Sử dụng output của Model 3 (mức lỗ tiềm năng khi sai) để điều chỉnh khối lượng giao dịch. Rủi ro tiềm năng cao => Khối lượng nhỏ hơn.
   * **Đặt Cắt Lỗ:** Sử dụng Mức lỗ tối đa/phân vị lỗ khi sai từ Model 3 làm cơ sở tham khảo để đặt lệnh stop-loss ban đầu.

**Bước 4: Backtesting và Tinh Chỉnh**

1. **Backtest Toàn Bộ Hệ Thống:** Áp dụng logic quyết định ở Bước 3 trên dữ liệu lịch sử (out-of-sample).
2. **Mô Phỏng Giao Dịch:** Tính toán chi phí giao dịch (phí, thuế), trượt giá (slippage) ước tính.
3. **Đánh Giá Hiệu Quả Tổng Thể:**
   * Lợi nhuận/Thua lỗ ròng.
   * Tỷ lệ thắng (Win Rate).
   * Tỷ lệ Risk/Reward trung bình.
   * Drawdown tối đa (Mức sụt giảm tài khoản lớn nhất).
   * Sharpe Ratio, Sortino Ratio (Đo lường lợi nhuận đã điều chỉnh theo rủi ro).
4. **Phân Tích Kết Quả:** Xem xét các giao dịch thua lỗ lớn, các giai đoạn hệ thống hoạt động kém hiệu quả.
5. **Tinh Chỉnh:** Quay lại Bước 2 và 3 để:
   * Cải thiện các mô hình riêng lẻ (thêm features, đổi thuật toán, tinh chỉnh tham số).
   * Điều chỉnh logic kết hợp, các ngưỡng quyết định.
   * Lặp lại quá trình backtesting.

**Bước 5: Triển Khai và Theo Dõi (Nếu muốn tự động hoặc bán tự động)**

1. **Triển Khai:** Đưa các mô hình và logic vào hoạt động (có thể là script chạy hàng ngày, hoặc hệ thống phức tạp hơn).
2. **Theo Dõi Liên Tục:** Giám sát hiệu quả của hệ thống trên dữ liệu thực tế. Thị trường luôn thay đổi, mô hình cần được đánh giá lại và huấn luyện lại định kỳ (retrain).

**Lưu Ý Quan Trọng:**

* **Bắt đầu đơn giản:** Có thể bắt đầu chỉ với Model 1 và Model 3 (phân tích lịch sử), sau đó thêm Model 2.
* **Tránh Overfitting:** Đảm bảo các mô hình hoạt động tốt trên dữ liệu chưa thấy (test set), không chỉ trên dữ liệu huấn luyện. Sử dụng cross-validation.
* **Không có "Chén Thánh":** Mục tiêu là xây dựng một hệ thống có lợi thế thống kê, không phải dự đoán đúng 100%.
* **Quản Lý Rủi Ro Thực Tế:** Luôn tuân thủ các quy tắc quản lý vốn nghiêm ngặt, không bao giờ mạo hiểm quá nhiều trên một giao dịch, bất kể mô hình nói gì.

Hướng đi này cung cấp một khung sườn chi tiết để bạn xây dựng hệ thống dự đoán của mình, tập trung vào việc kết hợp hiệu quả dự đoán hướng đi, mức độ và đặc biệt là loại rủi ro cụ thể mà bạn quan tâm. Chúc bạn thành công!