**MÔ TẢ PHÂN TÍCH DỮ LIỆU**

**Thông tin thành viên:**

Nguyễn Văn Vũ Em

Trường Xuân Cảnh

3122410090

3122410038

1. **Giới thiệu**

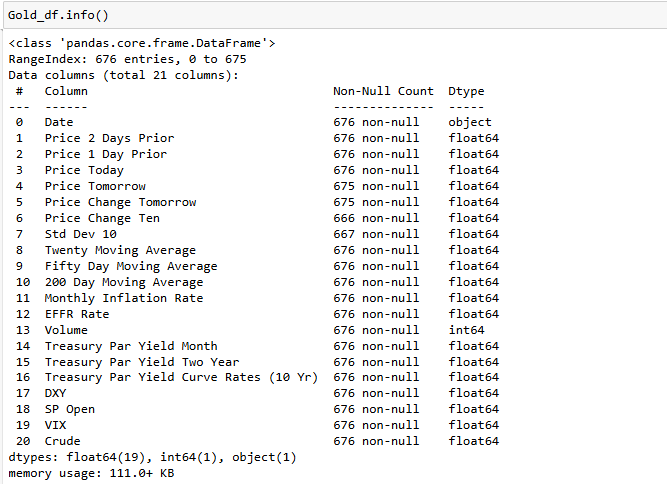
Trong báo cáo này, chúng tôi tiến hành phân tích tập dữ liệu "Dự đoán giá vàng" (GoldPricePrediction) với mục tiêu hiểu rõ hơn về cấu trúc dữ liệu, các thông tin chính liên quan đến biến động giá vàng, và các yếu tố kinh tế khác có ảnh hưởng đến giá vàng. Dữ liệu này được sử dụng để dự đoán giá vàng dựa trên các yếu tố kinh tế và tài chính khác nhau.

1. **Cấu trúc dữ liệu**
2. **Tổng quan về dữ liệu**

* Dữ liệu bao gồm 676 dòng và 21 cột, chi tiết từng cột như sau:
* **Date** (Ngày): Ngày ghi nhận giá vàng.
* **Price 2 Days Prior** (Giá trước 2 ngày): Giá vàng của 2 ngày trước ngày được ghi nhận.
* **Price 1 Day Prior** (Giá trước 1 ngày): Giá vàng của 1 ngày trước ngày được ghi nhận.
* **Price Today** (Giá hiện tại): Giá vàng trong ngày được ghi nhận.
* **Price Tomorrow** (Giá dự đoán ngày mai): Giá vàng dự đoán cho ngày hôm sau (có một số giá trị bị thiếu).
* **Price Change Tomorrow** (Mức thay đổi giá ngày mai): Mức độ thay đổi giá vàng dự đoán cho ngày hôm sau.
* **Price Change Ten** (Mức thay đổi giá trong 10 ngày): Mức độ thay đổi giá vàng trong 10 ngày tới (một số giá trị bị thiếu).
* **Std Dev 10** (Độ lệch chuẩn trong 10 ngày): Độ lệch chuẩn của giá vàng trong 10 ngày.
* **Twenty Moving Average** (Trung bình động 20 ngày): Giá vàng trung bình trong 20 ngày.
* **Fifty Day Moving Average** (Trung bình động 50 ngày): Giá vàng trung bình trong 50 ngày.
* **200 Day Moving Average** (Trung bình động 200 ngày): Giá vàng trung bình trong 200 ngày.
* **Monthly Inflation Rate** (Tỷ lệ lạm phát hàng tháng): Tỷ lệ lạm phát theo tháng.
* **EFFR** **Rate** (Lãi suất EFFR): Lãi suất quỹ liên bang (EFFR).
* **Volume** (Khối lượng giao dịch): Khối lượng giao dịch.
* **Treasury** **Par** **Yield** **Month** (Lợi suất trái phiếu kho bạc tháng): Lợi suất trái phiếu kho bạc tháng hiện tại.
* **Treasury** **Par Yield Two Year** (Lợi suất trái phiếu kho bạc 2 năm): Lợi suất trái phiếu kho bạc kỳ hạn 2 năm.
* **Treasury Par Yield Curve Rates** (10 Yr) (Lợi suất trái phiếu kho bạc 10 năm): Lợi suất trái phiếu kho bạc kỳ hạn 10 năm.
* **DXY** (Chỉ số DXY): Chỉ số đô la Mỹ (DXY).
* **SP Open** (Giá mở cửa của S&P 500): Giá mở cửa của chỉ số S&P 500.
* **VIX**: Chỉ số đo lường biến động (VIX).
* **Crude** (Giá dầu thô): Giá dầu thô.

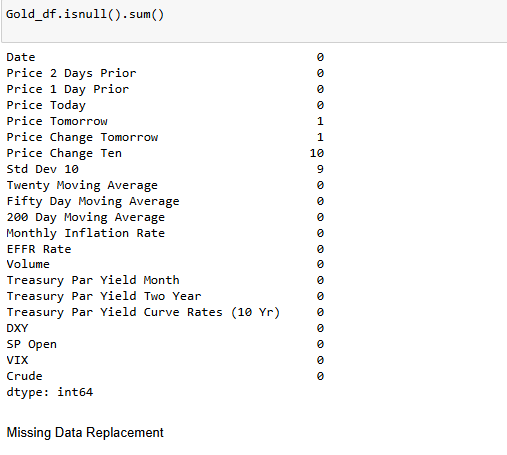
1. **Kiểu dữ liệu**

* Object (chuỗi ký tự): 1 cột
* Date (Ngày)
* Float64 (số thực): 19 cột
* Price 2 Days Prior
* Price 1 Day Prior
* Price Today
* Price Tomorrow
* Price Change Tomorrow
* Price Change Ten
* Std Dev 10
* Twenty Moving Average
* Fifty Day Moving Average
* 200 Day Moving Average
* Monthly Inflation Rate
* EFFR Rate
* Treasury Par Yield Month
* Treasury Par Yield Two Year
* Treasury Par Yield Curve Rates (10 Yr)
* DXY
* SP Open
* VIX
* Crude
* Int64 (số nguyên): 1 cột
* Volume (Khối lượng giao dịch)



1. **Dữ liệu bị thiếu**

* Trong tập dữ liệu, một số cột chứa giá trị NaN (dữ liệu bị thiếu). Cụ thể:
* Price Tomorrow: 1 giá trị NaN.
* Price Change Tomorrow: 1 giá trị NaN.
* Price Change Ten: 10 giá trị NaN.
* Std Dev 10: 9 giá trị NaN.
* Các giá trị này cần được xử lý trước khi tiến hành các phân tích sâu hơn.



1. **Phân tích chi tiết**

* Dữ liệu giá vàng: Các cột "Price 2 Days Prior", "Price 1 Day Prior", "Price Today", và "Price Tomorrow" cho thấy xu hướng giá vàng trong một khoảng thời gian ngắn (2 ngày trước, 1 ngày trước và hôm nay). Ngoài ra, các cột "Twenty Moving Average", "Fifty Day Moving Average", và "200 Day Moving Average" cung cấp các xu hướng dài hạn hơn của giá vàng.
* Dự báo giá: "Price Tomorrow" và "Price Change Tomorrow" giúp dự báo giá vàng vào ngày tiếp theo và mức thay đổi dự kiến. Tuy nhiên, một số giá trị bị thiếu trong cột "Price Tomorrow".
* Lạm phát và lãi suất: "Monthly Inflation Rate" và "EFFR Rate" cung cấp thông tin về tỷ lệ lạm phát và lãi suất, hai yếu tố quan trọng có thể ảnh hưởng đến giá vàng, bởi vàng thường được coi là tài sản phòng vệ trong thời kỳ lạm phát cao hoặc lãi suất thấp.
* Khối lượng giao dịch: Cột "Volume" cho biết số lượng giao dịch vàng trong mỗi ngày, giúp phân tích mức độ quan tâm và biến động của thị trường.
* Lợi suất trái phiếu: Các cột về "Treasury Par Yield" cho thấy mức lợi suất của trái phiếu kho bạc ở các kỳ hạn khác nhau (tháng, 2 năm, 10 năm), có thể ảnh hưởng đến giá vàng vì lợi suất trái phiếu thường được xem là thước đo của nền kinh tế.
* Chỉ số thị trường: Các chỉ số "DXY", "SP Open", "VIX", và "Crude" cho phép so sánh giá vàng với các chỉ số quan trọng khác như chỉ số đô la Mỹ (DXY), giá mở cửa của S&P 500, chỉ số biến động (VIX), và giá dầu thô.

1. **Nhận xét và hướng xử lý tiếp theo**

Dữ liệu cung cấp đầy đủ các thông tin liên quan đến giá vàng, khối lượng giao dịch, và các yếu tố kinh tế khác. Tuy nhiên, có một số cột chứa dữ liệu thiếu (NaN), cần xử lý trước khi thực hiện các phân tích dự báo chính xác.

Để tiếp tục phân tích, chúng tôi sẽ xử lý các giá trị NaN, áp dụng các phương pháp thống kê để tìm hiểu sự tương quan giữa các yếu tố kinh tế và giá vàng, từ đó xây dựng mô hình dự đoán giá vàng chính xác hơn.

1. **Kết luận**

Báo cáo này đã trình bày tổng quan về cấu trúc dữ liệu, các biến chính, và tình trạng dữ liệu thiếu. Sau khi xử lý dữ liệu, chúng tôi sẽ tiến hành các phân tích và dự đoán chi tiết hơn về biến động giá vàng, bao gồm sử dụng các phương pháp học máy (machine learning) để dự đoán xu hướng giá.

1. **Nguồn tài liệu**
2. **Nguồn dữ liệu:**

**Tập dữ liệu được lấy từ trang web**

<https://www.kaggle.com/datasets>

**Tập dữ liệu chính :**

<https://www.kaggle.com/datasets/cvergnolle/gold-price-and-relevant-metrics/data>

1. **Nguồn dữ liệu khác**

<https://www.usagold.com/daily-gold-price-history/>

<https://www.usinflationcalculator.com/inflation/historical-inflation-rates/>

<https://www.newyorkfed.org/markets/reference-rates/effr>

<https://www.newyorkfed.org/markets/reference-rates/effr>

<https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value_month=202408>

<https://home.treasury.gov/resource-center/data-chart-center/interest-rates/TextView?type=daily_treasury_yield_curve&field_tdr_date_value_month=202408>

<https://www.investing.com/indices/usdollar-historical-data>

<https://www.nasdaq.com/market-activity/index/spx/historical?page=1&rows_per_page=10&timeline=y5>

<https://finance.yahoo.com/quote/%5EVIX/history/?period1=1640995200&period2=1723573736>

<https://finance.yahoo.com/quote/CL%3DF/history/?period1=1640995200&period2=1723573988>