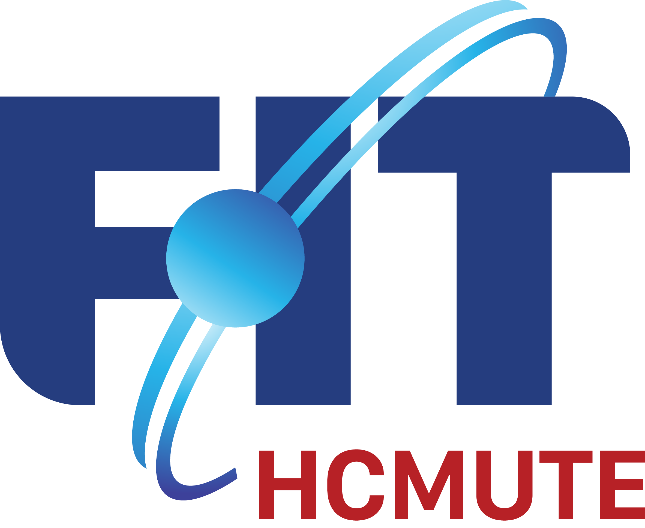
**TRƯỜNG ĐẠI HỌC SƯ PHẠM KỸ THUẬT TP.HCM**

**KHOA CÔNG NGHỆ THÔNG TIN**

**BỘ MÔN NHẬP MÔN KHOA HỌC DỮ LIỆU**



**NGUYỄN TRUNG HIẾU – 22110138**

**LÊ HOÀNG BẢO PHÚC – 22110200**

Đề Tài :

**PHÂN TÍCH RỦI RO ĐẦU TƯ CỔ PHIẾU TRONG CÁC CÔNG TY LỚN**

**ĐỒ ÁN CUỐI KỲ**

**GIÁO VIÊN HƯỚNG DẪN**

**ThS. LÊ MINH TÂN**

**KHÓA 2022 – 2026**

# LỜI CẢM ƠN

Nhóm chúng em xin gửi lời tri ân chân thành đến ThS. Lê Minh Tân, giảng viên môn nhập môn khoa học dữ liệu, người đã tận tâm hướng dẫn chúng em trong suốt hành trình học tập và hoàn thiện báo cáo cuối kỳ. Với kiến thức chuyên sâu và phong cách giảng dạy sinh động, thầy đã giúp chúng em nắm vững các khái niệm cốt lõi, từ thiết kế giao diện người dùng đến tích hợp các tính năng phức tạp như truy xuất dữ liệu và xử lý thời gian thực. Sự nhiệt huyết của thầy, qua những bài giảng rõ ràng và ví dụ thực tiễn, không chỉ giúp chúng em hiểu rõ lý thuyết mà còn khơi dậy niềm đam mê với lĩnh vực phát triển ứng dụng di động. Những buổi thảo luận sôi nổi do thầy dẫn dắt đã truyền cảm hứng, giúp chúng em tự tin khám phá các công nghệ mới.

Chúng em đặc biệt biết ơn sự tận tình của ThS. Lê Minh Tân trong vai trò người hướng dẫn dự án. Thầy đã kiên nhẫn giải đáp mọi thắc mắc, từ các lỗi lập trình nhỏ đến những vấn đề kỹ thuật phức tạp, đồng thời cung cấp những gợi ý quý báu để nhóm cải thiện chất lượng dự án. Sự hỗ trợ kịp thời của thầy, thông qua các buổi tư vấn và phản hồi chi tiết, đã giúp chúng em vượt qua những thử thách lớn trong quá trình phát triển dự án. Hơn thế, thầy luôn khuyến khích chúng em tư duy sáng tạo, thử nghiệm các giải pháp mới, từ đó không chỉ nâng cao kỹ năng lập trình mà còn rèn luyện khả năng phân tích và giải quyết vấn đề. Sự đồng hành của thầy là nguồn động lực lớn lao để chúng em hoàn thiện báo cáo một cách tốt nhất.

Cuối cùng, chúng em xin bày tỏ lòng tri ân sâu sắc đến ThS. Lê Minh Tân vì những bài học vượt ngoài khuôn khổ môn học. Qua những lời khuyên sâu sắc, thầy đã giúp chúng em nhận ra tầm quan trọng của sự kiên trì và tinh thần học hỏi không ngừng trong lĩnh vực công nghệ thông tin. Sự khích lệ của thầy đã xây dựng trong chúng em niềm tin vào khả năng của bản thân, thôi thúc chúng em tiếp tục theo đuổi đam mê phát triển ứng dụng di động. Những dấu ấn mà thầy để lại, từ kiến thức chuyên môn đến tinh thần tận tụy, sẽ là hành trang quý giá để chúng em vững bước trong các dự án tương lai. Một lần nữa, xin chân thành cảm ơn thầy vì tất cả những đóng góp không thể thay thế.

# MỤC LỤC

[**LỜI CẢM ƠN 1**](#_heading=h.syff2pnbug7g)

[**MỤC LỤC 2**](#_heading=h.bhekdhd6xmf3)

[**PHẦN 1: MỞ ĐẦU 3**](#_heading=h.zda43geunnh1)

[1. TÍNH CẤP THIẾT CỦA ĐỀ TÀI 3](#_heading=h.pt00wrzdn4p1)

[2. ĐỐI TƯỢNG, PHẠM VI NGHIÊN CỨU 3](#_heading=h.ete2l87colg8)

[3. PHÂN TÍCH HƯỚNG NGHIÊN CỨU LIÊN QUAN 4](#_heading=h.dgr7xotjp73t)

[4. KẾT QUẢ DỰ KIẾN 5](#_heading=h.s21lpbask45k)

[**PHẦN 2: NỘI DUNG 6**](#_heading=h.eczeevkm0jsn)

[**1. CHƯƠNG 1: TÌM HIỂU VỀ CỔ PHIẾU CỦA CÔNG TY LỚN 6**](#_heading=h.hzmkffxe2qwm)

[1. CÔNG TY APPLE 6](#_heading=h.b6lp37xmwnlf)

[1.1. Giới thiệu về tập dữ liệu 6](#_heading=h.uig2a4lh2qet)

[1.2. Phân tích về cổ phiếu 6](#_heading=h.h7kutkfmu5in)

[2. CÔNG TY GOOGLE 10](#_heading=h.5hpkyp73gukh)

[2.1. Giới thiệu về tập dữ liệu 10](#_heading=h.igjaudo7pv2j)

[2.2. Phân tích về cổ phiếu 11](#_heading=h.lwm1q394nevu)

[**2. CHƯƠNG 2: ĐÁNH GIÁ SỰ RỦI RO CỦA CỔ PHIẾU 24**](#_heading=h.6bxxfrdvj179)

[1. CƠ SỞ ĐÁNH GIÁ 24](#_heading=h.hjjl38am28oc)

[2. DẪN CHỨNG VỀ SỐ LIỆU CỔ PHIẾU 24](#_heading=h.u1k1l6dfbz6)

[3. ĐÁNH GIÁ VỀ RỦI RO CỔ PHIẾU 35](#_heading=h.pyyc7vq32mda)

[4. XÂY DỰNG MÔ HÌNH DỰ ĐOÁN XU HƯỚNG TĂNG GIẢM CỔ PHIẾU 38](#_heading=h.31j8433h2uzs)

[**PHẦN 3: KẾT LUẬN 40**](#_heading=h.ggquvb73hbr5)

[**TÀI LIỆU THAM KHẢO 42**](#_heading=h.opobzizefdzi)

# PHẦN 1: MỞ ĐẦU

## TÍNH CẤP THIẾT CỦA ĐỀ TÀI

Sự phát triển mạnh mẽ của thị trường chứng khoán toàn cầu, cùng với xu hướng hội nhập kinh tế, đã thúc đẩy sự quan tâm ngày càng lớn của nhà đầu tư Việt Nam đến cổ phiếu của các công ty hàng đầu như Google, Apple, Tesla và Amazon. Những công ty này, với vị thế dẫn đầu trong các lĩnh vực công nghệ, sản xuất và thương mại điện tử, mang lại tiềm năng sinh lời cao, nhưng đồng thời ẩn chứa nhiều rủi ro do biến động giá cổ phiếu, thay đổi chính sách kinh tế vĩ mô, hay tâm lý thị trường khó lường. Tại Việt Nam, phần lớn nhà đầu tư cá nhân thiếu kiến thức và công cụ chuyên sâu để phân tích các yếu tố rủi ro, dẫn đến các quyết định đầu tư dựa trên cảm tính hoặc thông tin không đầy đủ. Điều này không chỉ gây tổn thất tài chính mà còn hạn chế khả năng tận dụng cơ hội từ thị trường quốc tế, nơi các công ty lớn đóng vai trò chủ chốt trong việc định hình xu hướng đầu tư.

Hiện nay, các nghiên cứu và tài liệu tại Việt Nam về phân tích rủi ro đầu tư cổ phiếu vẫn còn khan hiếm, chưa đáp ứng được nhu cầu của nhà đầu tư và các tổ chức tài chính trong việc đưa ra quyết định chính xác. Đề tài phân tích rủi ro khi đầu tư cổ phiếu của các công ty lớn ra đời nhằm cung cấp các phương pháp đánh giá toàn diện, giúp nhận diện các yếu tố ảnh hưởng đến hiệu suất đầu tư, từ biến động thị trường đến các sự kiện doanh nghiệp cụ thể. Kết quả nghiên cứu không chỉ hỗ trợ nhà đầu tư tối ưu hóa lợi nhuận và giảm thiểu tổn thất mà còn góp phần nâng cao nhận thức tài chính trong cộng đồng. Hơn nữa, việc nghiên cứu này thúc đẩy sự phát triển bền vững của thị trường chứng khoán Việt Nam, tạo nền tảng cho các hoạt động giáo dục tài chính và hội nhập kinh tế quốc tế, mang lại giá trị kinh tế và xã hội lâu dài.

## ĐỐI TƯỢNG, PHẠM VI NGHIÊN CỨU

Đối tượng nghiên cứu của đề tài là các yếu tố rủi ro tài chính liên quan đến việc đầu tư vào cổ phiếu của các công ty lớn, bao gồm Google, Apple, Tesla, Amazon và một số doanh nghiệp hàng đầu khác trong các lĩnh vực công nghệ, sản xuất, và thương mại điện tử. Cụ thể, nghiên cứu tập trung vào các đặc điểm của giá cổ phiếu, như xu hướng biến động, mức độ dao động, và các yếu tố ảnh hưởng như kết quả kinh doanh, chính sách kinh tế vĩ mô, hay tâm lý thị trường. Đối tượng nghiên cứu còn bao gồm hành vi của nhà đầu tư cá nhân và tổ chức tại Việt Nam khi tham gia thị trường chứng khoán quốc tế, đặc biệt là cách họ nhận diện và quản lý rủi ro trong các quyết định đầu tư. Thông qua việc phân tích dữ liệu lịch sử và các chỉ số tài chính, đề tài hướng đến việc cung cấp cái nhìn toàn diện về mức độ rủi ro và cơ hội đầu tư, từ đó hỗ trợ nhà đầu tư tối ưu hóa danh mục đầu tư và giảm thiểu tổn thất tài chính.

Phạm vi nghiên cứu bao gồm việc phân tích dữ liệu lịch sử giá cổ phiếu và các yếu tố liên quan của các công ty lớn, tập trung vào thị trường chứng khoán quốc tế, với sự chú trọng đến bối cảnh Việt Nam. Nghiên cứu xem xét các khía cạnh như xu hướng giá, khối lượng giao dịch, và các chỉ số rủi ro tài chính, đồng thời đánh giá tác động của các sự kiện kinh tế và doanh nghiệp đến hiệu suất đầu tư. Về mặt không gian, đề tài tập trung vào các công ty niêm yết trên các sàn giao dịch lớn như NASDAQ và NYSE, nhưng kết quả nghiên cứu được định hướng để áp dụng cho nhà đầu tư Việt Nam. Về mặt thời gian, nghiên cứu phân tích dữ liệu trong một khoảng thời gian đủ dài để phản ánh các chu kỳ thị trường, bao gồm các giai đoạn tăng trưởng, suy thoái, và phục hồi. Đề tài cũng đề xuất các phương pháp phân tích rủi ro có thể áp dụng trong giáo dục tài chính, góp phần nâng cao nhận thức và kỹ năng quản lý tài sản cho cộng đồng, đặc biệt là thế hệ trẻ tại Việt Nam.

## PHÂN TÍCH HƯỚNG NGHIÊN CỨU LIÊN QUAN

Các nghiên cứu về phân tích rủi ro đầu tư cổ phiếu trên thị trường quốc tế đã đạt được nhiều thành tựu đáng kể, đặc biệt trong việc áp dụng các phương pháp thống kê và học máy để đánh giá biến động giá và dự đoán hiệu suất đầu tư. Các công trình tiêu biểu tập trung vào việc sử dụng các chỉ số như độ lệch chuẩn, giá trị rủi ro (VaR), và tỷ lệ Sharpe để đo lường mức độ rủi ro của cổ phiếu các công ty lớn như Google, Apple, Tesla, và Amazon. Ngoài ra, một số nghiên cứu đã phân tích tác động của các yếu tố kinh tế vĩ mô, như lãi suất, lạm phát, và tâm lý thị trường, cũng như các sự kiện doanh nghiệp, như chia tách cổ phiếu hay báo cáo tài chính, đến giá cổ phiếu. Những nghiên cứu này cung cấp nền tảng lý thuyết vững chắc và các công cụ phân tích hiệu quả, giúp nhà đầu tư quốc tế đưa ra quyết định dựa trên dữ liệu. Tuy nhiên, phần lớn các công trình tập trung vào thị trường phát triển như Mỹ và châu Âu, ít chú trọng đến bối cảnh các thị trường mới nổi như Việt Nam, nơi nhà đầu tư có đặc điểm và nhu cầu khác biệt.

Tại Việt Nam, các nghiên cứu về rủi ro đầu tư cổ phiếu vẫn còn hạn chế, chủ yếu tập trung vào thị trường nội địa và thiếu các phân tích chuyên sâu về cổ phiếu quốc tế. Một số công trình đã đề xuất các phương pháp đánh giá rủi ro dựa trên dữ liệu lịch sử, nhưng thường không tích hợp các yếu tố như hành vi nhà đầu tư cá nhân, vốn phổ biến ở Việt Nam, hoặc tác động của các sự kiện toàn cầu đến các công ty lớn. Đề tài này hướng đến việc lấp đầy khoảng trống nghiên cứu bằng cách phân tích rủi ro đầu tư cổ phiếu các công ty lớn, với trọng tâm là ứng dụng thực tiễn cho nhà đầu tư Việt Nam. Nghiên cứu không chỉ khai thác dữ liệu lịch sử để đánh giá xu hướng giá và các chỉ số rủi ro mà còn xem xét bối cảnh văn hóa và kinh tế Việt Nam, từ đó đề xuất các khuyến nghị phù hợp. Kết quả nghiên cứu góp phần nâng cao nhận thức tài chính, hỗ trợ giáo dục tài chính cho thế hệ trẻ, và thúc đẩy sự hội nhập của Việt Nam vào thị trường chứng khoán toàn cầu.

## KẾT QUẢ DỰ KIẾN

Nghiên cứu phân tích rủi ro khi đầu tư cổ phiếu của các công ty lớn như Google, Apple, Tesla, và Amazon dự kiến mang lại một hệ thống các chỉ số và phương pháp đánh giá rủi ro tài chính toàn diện, giúp nhà đầu tư Việt Nam đưa ra quyết định đầu tư chính xác hơn. Cụ thể, nghiên cứu sẽ cung cấp các phân tích chi tiết về xu hướng giá cổ phiếu, mức độ biến động, và các yếu tố ảnh hưởng như kết quả kinh doanh, chính sách kinh tế vĩ mô, hoặc tâm lý thị trường. Kết quả bao gồm các chỉ số rủi ro như độ lệch chuẩn, giá trị rủi ro (VaR), và tỷ lệ Sharpe, được trình bày dưới dạng bảng biểu và đồ thị trực quan, giúp nhà đầu tư dễ dàng nhận diện mức độ rủi ro và cơ hội sinh lời của từng cổ phiếu. Ngoài ra, nghiên cứu sẽ xác định các giai đoạn thị trường có rủi ro cao, chẳng hạn như trong các cuộc khủng hoảng kinh tế hoặc sự kiện doanh nghiệp lớn, từ đó hỗ trợ nhà đầu tư xây dựng chiến lược quản lý danh mục đầu tư hiệu quả.

Bên cạnh đó, nghiên cứu dự kiến đóng góp vào việc nâng cao nhận thức và giáo dục tài chính tại Việt Nam, đặc biệt cho thế hệ trẻ và nhà đầu tư cá nhân. Các kết quả nghiên cứu sẽ được tổng hợp thành tài liệu tham khảo, cung cấp kiến thức về cách phân tích rủi ro và đánh giá hiệu suất đầu tư, phù hợp để sử dụng trong các chương trình đào tạo tài chính tại trường học và tổ chức. Đề tài cũng kỳ vọng đề xuất các khuyến nghị thực tiễn, chẳng hạn như cách đa dạng hóa danh mục đầu tư hoặc thời điểm mua/bán cổ phiếu, nhằm tối ưu hóa lợi nhuận và giảm thiểu tổn thất. Những kết quả này không chỉ hỗ trợ nhà đầu tư Việt Nam tham gia thị trường chứng khoán quốc tế mà còn thúc đẩy sự phát triển bền vững của thị trường nội địa, góp phần vào quá trình hội nhập kinh tế toàn cầu và nâng cao năng lực tài chính của cộng đồng.

# PHẦN 2: NỘI DUNG

# CHƯƠNG 1: TÌM HIỂU VỀ CỔ PHIẾU CỦA CÔNG TY LỚN

## CÔNG TY APPLE

### Giới thiệu về tập dữ liệu

Link dataset: <https://www.kaggle.com/datasets/suyashlakhani/apple-stock-prices-20152020>

Điểm Usability: 10.00

Tập dữ liệu Apple Stock Prices (2015-2020) là tập dữ liệu thông tin về giá cổ phiếu của Apple Inc trong khoảng thời gian từ năm 2015 đến năm 2020. Đây là tập dữ liệu giúp phân tích xu hướng giá cổ phiếu, dự đoán thị trường và kiểm tra hiệu suất của cổ phiếu Apple qua các năm.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột thuộc tính** | **Kiểu dữ liệu** | **Ý nghĩa** | **Kiểu dữ liệu** |
| Số thứ tự | int | Số thứ tự của các cột dữ liệu | Rời rạc |
| Symbol | string | Mã chứng khoán của Apple | Rời rạc |
| Date | datetime | Ngày ghi nhận giá cổ phiếu | Liên tục |
| Close | float | Giá đóng cửa của cổ phiếu trong ngày | Liên tục |
| High | float | Giá cao nhất của cổ phiếu trong ngày | Liên tục |
| Low | float | Giá thấp nhất của cổ phiếu trong ngày | Liên tục |
| Open | float | Giá mở cửa của cổ phiếu trong ngày | Liên tục |
| Volume | int | Khối lượng cổ phiếu được giao dịch trong ngày | Liên tục |
| AdjClose | float | Giá đóng cửa đã được điều chỉnh theo các yếu tố như cổ tức và chia tách cổ phiếu | Liên tục |
| AdjHigh | float | Giá cao nhất đã được điều chỉnh | Liên tục |
| AdjLow | float | Giá thấp nhất đã được điều chỉnh. | Liên tục |
| AdjOpen | float | Giá mở cửa đã được điều chỉnh. | Liên tục |
| AdjVolume | int | Khối lượng giao dịch đã được điều chỉnh. | Liên tục |
| DivCash | float | Cổ tức tiền mặt trả cho cổ đông trên mỗi cổ phiếu. | Liên tục |
| SplitFactor | int | Hệ số chia tách cổ phiếu | Liên tục |

### Phân tích về cổ phiếu

Dưới đây là tổng quan của tập dữ liệu:



**Số Lượng (Count):** Số lượng bản ghi: 1258 cho mỗi cột, cho thấy không có giá trị bị thiếu trong bộ dữ liệu.

**Độ Lệch Chuẩn (Standard Deviation)**

* Độ lệch chuẩn của giá đóng cửa (close): 56.85 USD, cho thấy sự biến động đáng kể trong giá cổ phiếu.
* Độ lệch chuẩn của khối lượng giao dịch (volume): 1.729e+07 (17.29 triệu), thể hiện sự biến động trong khối lượng giao dịch.
* Độ lệch chuẩn của giá đóng cửa điều chỉnh (adjClose): 58.73 USD, cũng cho thấy sự biến động đáng kể trong giá điều chỉnh.
* Độ lệch chuẩn của cổ tức tiền mặt (divCash): 0.03366 USD, cho thấy sự biến động trong việc trả cổ tức.

**Xu hướng giá:** Giá trị trung bình của giá đóng cửa, giá cao nhất và giá thấp nhất cho thấy xu hướng tăng trưởng ổn định từ 2015 đến 2020.Độ lệch chuẩn cao của giá đóng cửa, giá cao nhất và giá thấp nhất cho thấy sự biến động đáng kể trong các phiên giao dịch hàng ngày.

**Khối lượng giao dịch:** Khối lượng giao dịch trung bình là 35 triệu, nhưng giá trị tối đa lên đến 162.2 triệu, cho thấy có những phiên giao dịch đặc biệt có khối lượng giao dịch rất lớn.

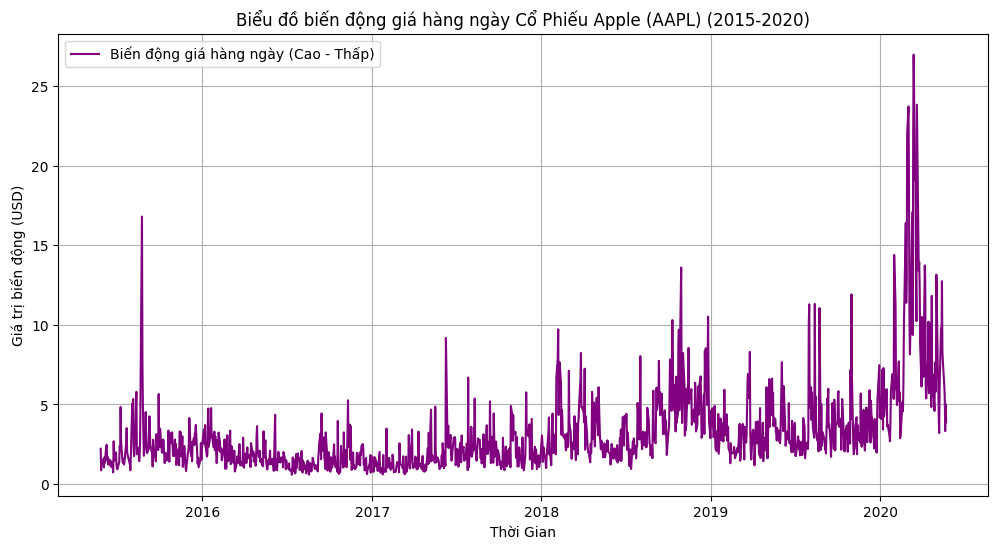
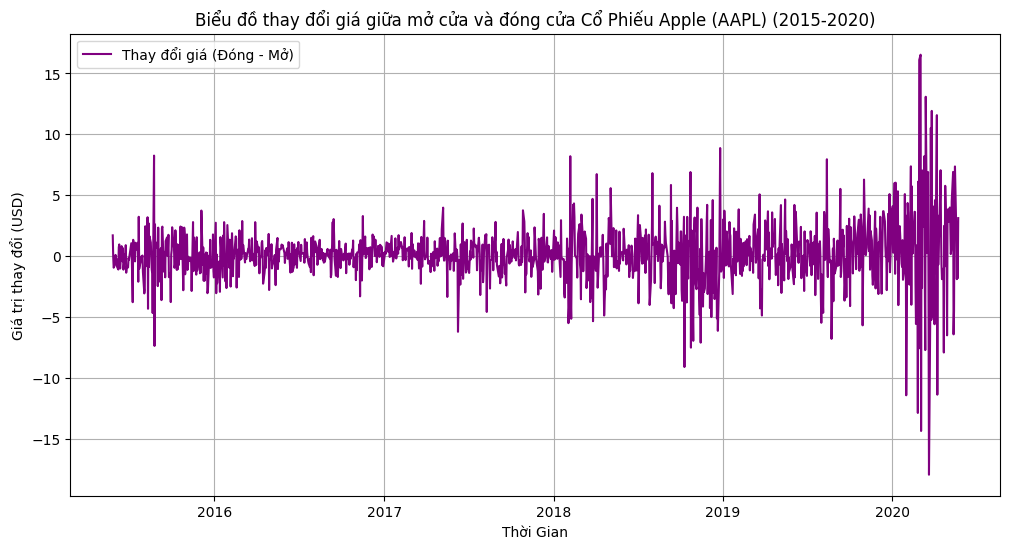
**Cổ tức:** Cổ tức tiền mặt trung bình là rất nhỏ (0.010477 USD) và giá trị tối đa là 0.82 USD, cho thấy không phải ngày nào cũng trả cổ tức, và khi trả thì giá trị cũng không đồng đều.

**Phân phối dữ liệu:** Biểu đồ hộp cho thấy sự phân phối của các giá trị và các giá trị ngoại lệ (outliers). Giá trị trung vị (median) gần với giá trị trung bình (mean) cho thấy phân phối gần đối xứng. Ngoài ra giá trị tứ phân vị cho phép đánh giá sự tập trung của dữ liệu xung quanh giá trị trung vị.

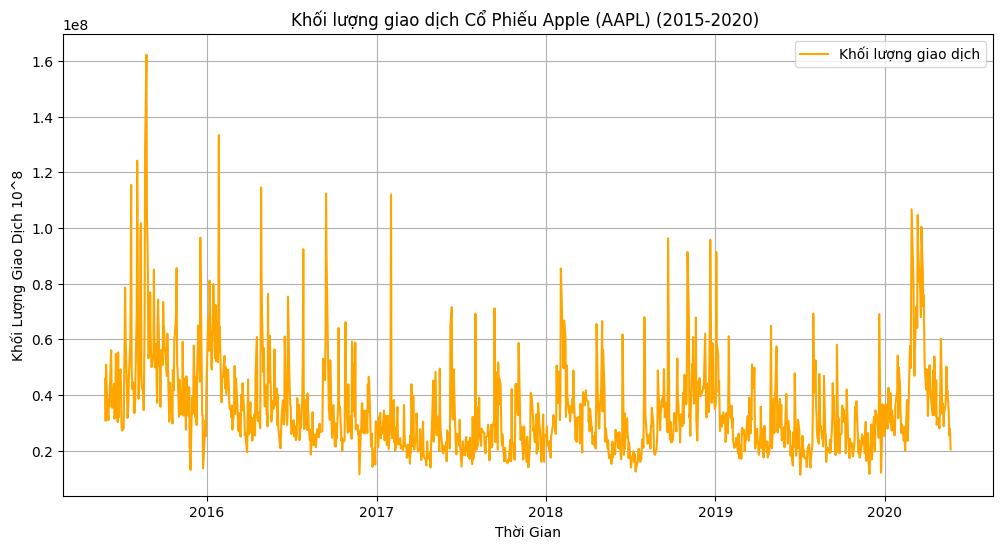
**Xu Hướng Tổng Quan**

Tăng Trưởng: Giá cổ phiếu Apple có xu hướng tăng trưởng mạnh mẽ trong giai đoạn 2015 - 2020. Từ khoảng 132 USD vào tháng 5 năm 2015, giá cổ phiếu đã tăng lên trên 320 USD vào tháng 5 năm 2020.

**Biến Động Giá**

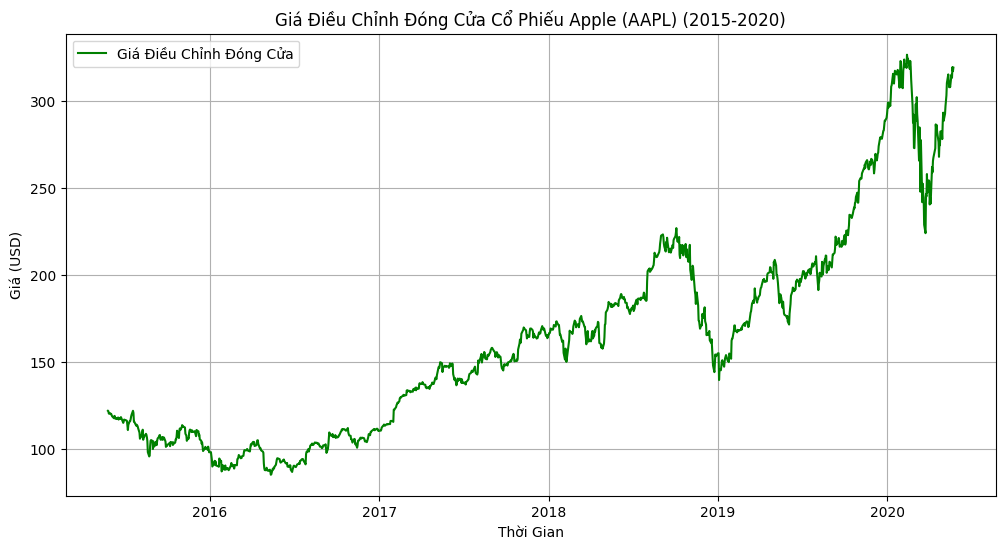


Có thể thấy rõ sự biến động giá hàng ngày qua các cột high và low. Sự khác biệt giữa giá cao nhất và thấp nhất trong một ngày cho thấy mức độ biến động trong phiên giao dịch đó. Ngoài ra, biểu đồ về sự thay đổi giữa giá mở cửa và đóng cửa hàng ngày cho thấy tâm lý thị trường, với giá trị dương thường chỉ ra tâm lý tích cực và giá trị âm thể hiện tâm lý tiêu cực. Biểu đồ biến động giá hàng ngày (giá cao nhất trừ giá thấp nhất) đo lường mức độ rủi ro trong mỗi phiên giao dịch.



Khối lượng giao dịch: Khối lượng giao dịch (volume) cũng biến động theo thời gian. Có những ngày khối lượng giao dịch tăng đột biến, thường trùng với các sự kiện quan trọng hoặc thông tin mới về công ty, các sự kiện ra mắt sản phẩm mới.

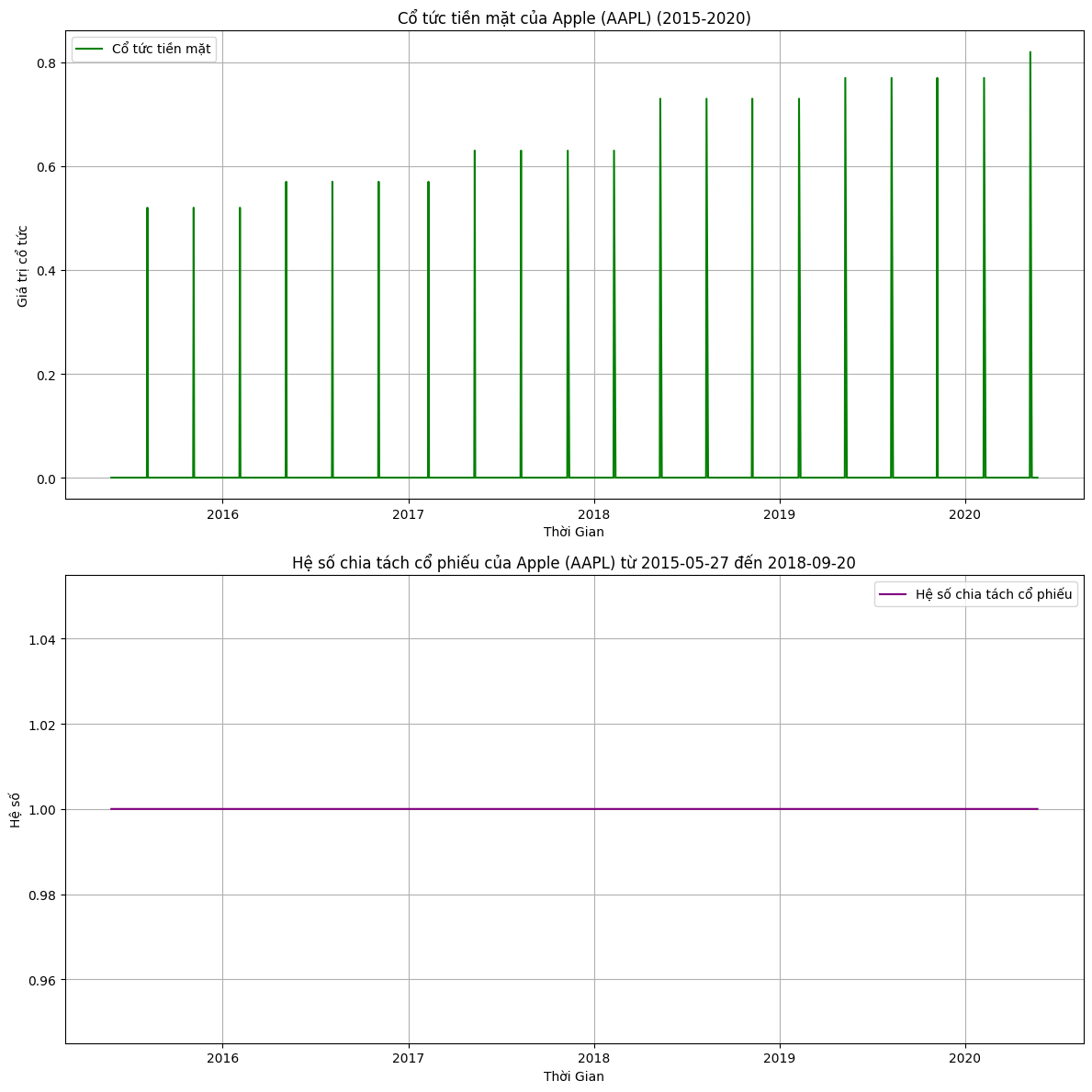
**Giá Đóng Cửa Điều Chỉnh (adjClose)**

****

Tính toán giá điều chỉnh: Giá đóng cửa điều chỉnh (adjClose) cung cấp cái nhìn chính xác hơn về hiệu suất cổ phiếu theo thời gian, vì nó đã loại bỏ ảnh hưởng của các sự kiện như chia tách cổ phiếu và cổ tức.

Xu hướng: Xu hướng chung của adjClose ở các giai đoạn là tăng, nhưng sự điều chỉnh giúp so sánh giá trị cổ phiếu một cách chính xác hơn qua các năm.

**Cổ Tức và Chia Tách Cổ Phiếu**



Cổ tức (divCash): Đa số các giá trị trong cột này là 0, cho thấy không phải ngày nào Apple cũng trả cổ tức.

Hệ số chia tách (splitFactor): Hầu hết giá trị là 1, cho thấy không có chia tách cổ phiếu lớn nào trong giai đoạn này.

**Các Giai Đoạn Quan Trọng**

* Giai đoạn 2015-2016: Giá cổ phiếu có nhiều sự thay đổi, không ổn định.
* Giai đoạn 2017-2018: Giá cổ phiếu tăng trưởng mạnh và ổn định hơn, đặc biệt là trong năm 2018.
* Giai đoạn 2019-2020: Giá cổ phiếu tăng nhanh nhưng có nhiều biến động trong các tháng đầu năm 2019 và 2020.

## CÔNG TY GOOGLE

### Giới thiệu về tập dữ liệu

Link dataset: <https://www.kaggle.com/datasets/yaminh/google-historical-stock-prices>

Điểm Usability: 8.24

Tập dữ liệu Google Historical Stock Prices, cung cấp thông tin lịch sử toàn diện về giá cổ phiếu Alphabet Inc. (mã GOOGL) trên sàn NasdaqGS. Bộ dữ liệu bao gồm các chỉ số chính như ngày giao dịch, giá mở cửa, đóng cửa, cao nhất, thấp nhất, khối lượng giao dịch, giá điều chỉnh, cổ tức, và hệ số chia tách. Với dữ liệu phong phú này, nhà phân tích có thể khám phá xu hướng giá, biến động thị trường, và tác động của các hành động doanh nghiệp.

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột thuộc tính** | **Kiểu dữ liệu** | **Ý nghĩa** | **Kiểu dữ liệu** |
| date | datetime | Ngày giao dịch | Rời rạc |
| close | float | Giá đóng cửa của cổ phiếu vào ngày giao dịch | Liên tục |
| high | float | Mức giá cao nhất của cổ phiếu đạt được trong ngày giao dịch | Liên tục |
| low | float | Mức giá thấp nhất của cổ phiếu đạt được trong ngày giao dịch | Liên tục |
| open | float | Giá mở cửa của cổ phiếu vào ngày giao dịch | Liên tục |
| volume | int | Tổng số cổ phiếu được giao dịch trong ngày | Rời rạc |
| adjClose | float | Giá đóng cửa của cổ phiếu đã được điều chỉnh bởi doanh nghiệp | Liên tục |
| adjHigh | float | Mức giá cao nhất của cổ phiếu đã được điều chỉnh bởi doanh nghiệp | Liên tục |
| adjLow | float | Mức giá thấp nhất của cổ phiếu đã được điều chỉnh bởi doanh nghiệp | Liên tục |
| adjOpen | float | Giá mở cửa của cổ phiếu đã được điều chỉnh theo các hành động của doanh nghiệp | Liên tục |
| adjVolume | int | Tổng khối lượng cổ phiếu giao dịch đã được điều chỉnh bởi doanh nghiệp | Rời rạc |
| divCash | float | Số tiền cổ tức bằng tiền mặt được chi trả trên mỗi cổ phiếu | Liên tục |
| splitFactor | int | Hệ số chia tách cổ phiếu | Rời rạc |

### Phân tích về cổ phiếu

Ta bắt đầu với các chỉ số thống kê cơ bản của hai cột chính là close (giá đóng cửa) và volume (khối lượng giao dịch). Dữ liệu này là giá cổ phiếu Google từ 2016 đến 2021. Kết quả thống kê cho thấy trung bình giá đóng cửa là 1216.32 USD, trung vị 1132.46 USD, độ lệch chuẩn 383.33 USD, giá thấp nhất 668.26 USD và cao nhất 2521.60 USD. Đối với khối lượng giao dịch, trung bình mỗi ngày có khoảng 1.6 triệu cổ phiếu được giao dịch, trung vị 1.41 triệu, độ lệch chuẩn 696017.23, thấp nhất 346753 cổ phiếu và cao nhất 6.2 triệu cổ phiếu.

Khi phân tích giá đóng cửa, ta thấy trung bình 1216.32 USD cao hơn trung vị 1132.46 USD, điều này cho thấy dữ liệu bị lệch phải, tức là có một số ngày giá cổ phiếu tăng cao bất thường, kéo trung bình lên. Độ lệch chuẩn 383.33 USD là khá lớn so với trung bình, chứng tỏ giá cổ phiếu Google biến động mạnh trong giai đoạn này. Giá dao động từ 668.26 USD đến 2521.60 USD, biên độ rất rộng, điều này dễ hiểu với một cổ phiếu công nghệ lớn như Google, đặc biệt trong giai đoạn 2016-2021 khi thị trường chứng khoán có nhiều biến động lớn, như khủng hoảng năm 2020 hay sự tăng trưởng mạnh của ngành công nghệ sau đó.

Còn về khối lượng giao dịch, trung bình 1.6 triệu cổ phiếu mỗi ngày nhưng trung vị chỉ 1.41 triệu, cũng cho thấy dữ liệu bị lệch phải. Có một số ngày khối lượng giao dịch tăng đột biến, ví dụ cao nhất lên tới 6.2 triệu cổ phiếu, trong khi ngày thấp nhất chỉ 346753 cổ phiếu. Độ lệch chuẩn 696017.23 cho thấy sự dao động rất lớn trong khối lượng giao dịch, thường liên quan đến các sự kiện lớn như công bố báo cáo tài chính, tin tức quan trọng về Google, hoặc biến động chung của thị trường.

Tiếp theo, ta kiểm tra xem dữ liệu có bị thiếu không và có giá trị bất thường nào cần xử lý không, để đảm bảo chất lượng dữ liệu trước khi phân tích sâu hơn. Kết quả kiểm tra cho thấy không có giá trị nào bị thiếu trong cả close và volume, tức là 0 giá trị thiếu. Đây là một điểm cộng lớn, vì dữ liệu đầy đủ sẽ giúp các bước phân tích và mô hình hóa sau này chính xác hơn, không cần phải điền giá trị hay xử lý thêm.

Sau đó, ta dùng phương pháp IQR để tìm các giá trị bất thường. Kết quả cho thấy có 90 giá trị bất thường trong close và 74 giá trị trong volume. Với close, các giá trị bất thường nằm ở mức cao như 2521.60 USD hay 2513.93 USD, hoặc thấp như 2024.17 USD và 2026.71 USD. Đây có thể là những ngày thị trường biến động mạnh, ví dụ khi Google công bố lợi nhuận vượt kỳ vọng hoặc khi thị trường chung giảm điểm, như giai đoạn đầu dịch COVID-19. Với volume, các giá trị bất thường như 6207027 hay 5130576 cho thấy có những ngày giao dịch sôi động bất thường, có thể do các sự kiện lớn như chia tách cổ phiếu, tin tức quan trọng, hoặc hành động mua bán lớn từ các nhà đầu tư tổ chức.

Để xử lý các giá trị bất thường này, ta đã áp dụng phương pháp Winsorizing, tức là thay thế các giá trị nằm ngoài ngưỡng IQR bằng ngưỡng trên và ngưỡng dưới. Sau khi xử lý, ta kiểm tra lại và thấy số lượng giá trị bất thường trong cả close và volume đều giảm về 0. Điều này cho thấy dữ liệu đã được làm sạch, không còn các giá trị cực đoan có thể gây ảnh hưởng lớn đến các mô hình phân tích sau này.

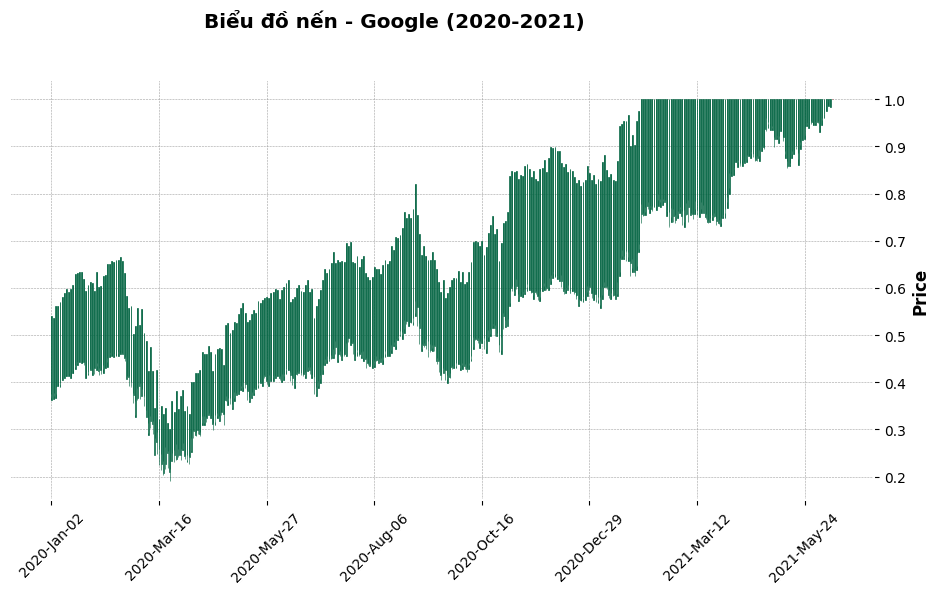
Sau khi xử lý outliers, ta tính các chỉ số thống kê mô tả cho cột close để xem dữ liệu thay đổi thế nào. Kết quả cho thấy trung bình là 1197.44 USD, phương sai 111862.25, độ lệch 0.83, và độ nhọn 0.005.

Khi phân tích các chỉ số này, ta thấy trung bình 1197.44 USD giảm nhẹ so với trước khi xử lý là 1216.32 USD. Điều này cho thấy các giá trị cao bất thường như 2521.60 USD đã được thay thế bằng ngưỡng trên của IQR, làm trung bình giảm xuống, phản ánh đúng hơn giá trị trung bình thực tế của dữ liệu. Phương sai 111862.25 vẫn khá lớn, cho thấy giá cổ phiếu Google vẫn có độ biến động đáng kể ngay cả sau khi xử lý outliers, điều này phù hợp với bản chất của một cổ phiếu công nghệ lớn, vốn thường chịu ảnh hưởng từ nhiều yếu tố như thị trường, tin tức, và kỳ vọng của nhà đầu tư.

Độ lệch 0.83 là một giá trị dương, nghĩa là dữ liệu vẫn lệch phải, nhưng mức độ lệch đã giảm so với trước khi xử lý, do ảnh hưởng của các giá trị cực đại đã được giảm bớt. Dữ liệu lệch phải cho thấy giá cổ phiếu có xu hướng tăng nhiều hơn giảm, phù hợp với giai đoạn tăng trưởng của Google từ 2016 đến 2021. Còn độ nhọn 0.005 rất gần 0, cho thấy phân phối của close khá giống phân phối chuẩn, không có đuôi quá dày hay quá mỏng. Điều này có nghĩa là sau khi xử lý outliers, dữ liệu đã trở nên cân đối hơn, ít bị ảnh hưởng bởi các giá trị cực đoan.

Tóm lại, sau khi xử lý outliers, dữ liệu close trở nên ổn định hơn, với các chỉ số thống kê phản ánh đúng hơn đặc điểm của giá cổ phiếu Google trong giai đoạn 2016-2021. Giá trị trung bình giảm nhẹ, độ lệch phải vẫn còn nhưng ở mức hợp lý, và phân phối của dữ liệu gần giống phân phối chuẩn hơn. Điều này sẽ giúp các bước phân tích tiếp theo như vẽ biểu đồ hay xây dựng mô hình cho kết quả đáng tin cậy hơn, không bị ảnh hưởng quá mức bởi các giá trị bất thường.

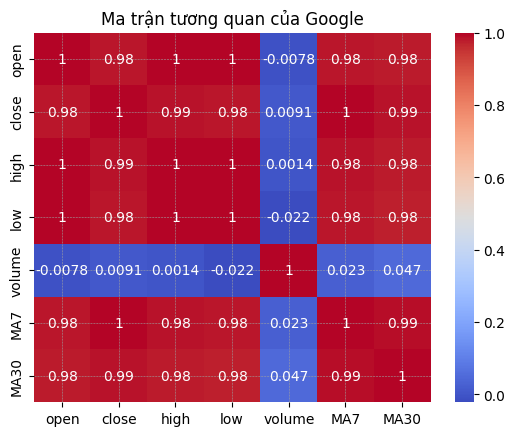
Tiếp theo, ta sẽ phân tích trực quan hơn thông qua các biểu đồ biểu diễn các đặc trưng ảnh hưởng trực tiếp đến cổ phiếu.



Biểu đồ nến trước mặt thể hiện giá cổ phiếu Google từ tháng 1/2020 đến tháng 5/2021, với giá trị chuẩn hóa từ 0 đến 1 qua MinMaxScaler, trục ngang là thời gian (các mốc như 1/2020, 3/2020, 5/2021), trục dọc là giá chuẩn hóa, và các thanh nến xanh lá biểu thị giá mở cửa, đóng cửa, cao nhất, thấp nhất hàng ngày. Nhìn tổng thể, giá có xu hướng tăng dài hạn, từ mức 0.2 đầu năm 2020 lên gần 1 vào giữa năm 2021, nhưng không thiếu biến động, đặc biệt giai đoạn đầu năm 2020 do đại dịch COVID-19, trước khi phục hồi mạnh nhờ nhu cầu công nghệ tăng cao.

Giá khởi điểm khoảng 0.2 vào tháng 1/2020, rồi giảm mạnh xuống gần 0 vào tháng 3/2020, trùng với đáy thị trường khi COVID-19 bùng phát, gây bán tháo cổ phiếu công nghệ. Từ tháng 4/2020, giá phục hồi, đạt 0.4 giữa năm 2020, dù biến động nhấp nhô, đến tháng 8/2020 tăng vượt 0.6, nhờ sự lạc quan từ các biện pháp kích thích kinh tế và nhu cầu dịch vụ Google như YouTube, Google Meet. Sang năm 2021, từ tháng 3 đến tháng 5, giá tăng ổn định, chạm mức 1, cho thấy đỉnh giá, có thể do báo cáo tài chính tích cực, nhưng cuối biểu đồ, thanh nến thu hẹp, báo hiệu khả năng điều chỉnh.

Biểu đồ phản ánh tác động của yếu tố vĩ mô như đại dịch, với giá giảm mạnh đầu 2020 nhưng phục hồi nhanh nhờ sức mạnh nội tại của Google trong xu hướng số hóa. Biến động lớn ở giai đoạn đầu (tháng 3-4/2020) cho thấy rủi ro cao, trong khi xu hướng tăng ổn định năm 2021 thể hiện niềm tin dài hạn vào công ty. Tuy nhiên, giá chuẩn hóa chạm 1 vào tháng 5/2021 đặt câu hỏi về bong bóng giá, và thanh nến thu hẹp có thể báo hiệu điều chỉnh. Biểu đồ này giúp ta nhận diện xu hướng và rủi ro, hỗ trợ phân tích sâu hơn bằng các chỉ số như trung bình động hoặc xây dựng chiến lược đầu tư.



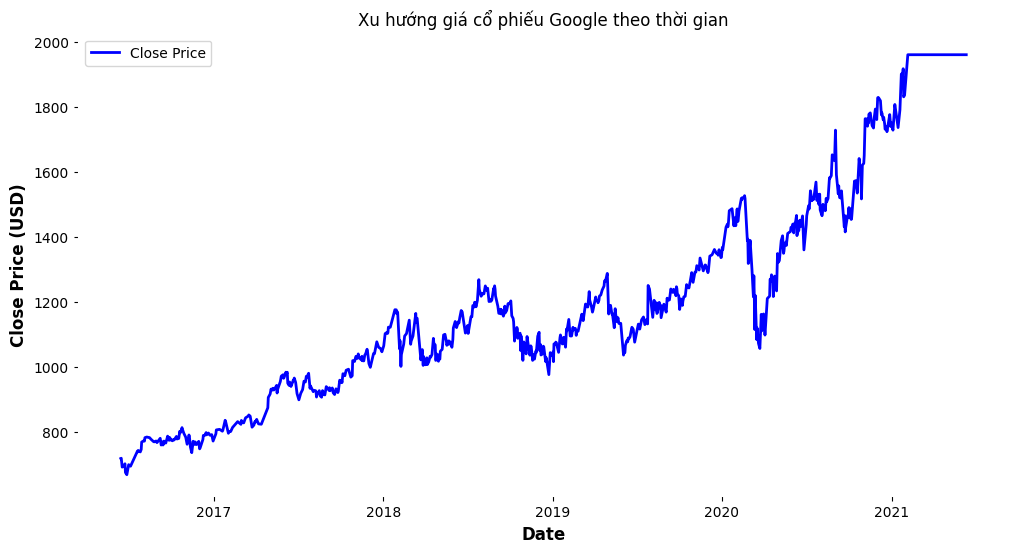
Biểu đồ tiếp theo là ma trận tương quan, phác họa mối quan hệ giữa các chỉ số tài chính của Google, bao gồm giá mở cửa (open), giá đóng cửa (close), giá cao nhất (high), giá thấp nhất (low), khối lượng giao dịch (volume), trung bình động 7 ngày (MA7), và trung bình động 30 ngày (MA30). Ma trận sử dụng bảng màu, với ô đỏ đậm thể hiện tương quan dương gần 1, ô xanh dương biểu thị tương quan âm gần -1, và các giá trị trung gian được mã hóa từ -1 đến 1. Dựa trên ngữ cảnh dữ liệu từ 2016 đến 2021, biểu đồ này giúp ta khám phá cách các yếu tố này tương tác trong giai đoạn đó. Nhìn chung, các chỉ số giá (open, close, high, low) gắn bó chặt chẽ với nhau, trong khi volume và trung bình động (MA7, MA30) lại ít liên quan hoặc ngược chiều, phản ánh sự khác biệt giữa giá và khối lượng, cũng như vai trò làm mịn của trung bình động.

Khi quan sát các chỉ số giá, ta thấy mối quan hệ giữa open, close, high, và low cực kỳ chặt chẽ, với tương quan dao động từ 0.98 đến 1. Chẳng hạn, open và close đạt 0.98, còn high và low lên tới 0.99, cho thấy giá trong ngày thay đổi đồng bộ, phản ánh xu hướng chung của thị trường. Điều này hợp lý với đặc điểm dữ liệu cổ phiếu, khi các mức giá thường biến động trong phạm vi hẹp hàng ngày. Tuy nhiên, khi chuyển sang volume, ta nhận thấy tương quan với các chỉ số giá rất yếu, dao động từ -0.078 đến -0.222, như volume và close chỉ -0.001, hay với low là -0.222. Mối quan hệ âm nhẹ này gợi ý rằng khối lượng giao dịch đôi khi tăng khi giá giảm, có thể do bán tháo trong biến động mạnh, hoặc giá tăng mà không có sự hỗ trợ từ khối lượng, báo hiệu xu hướng yếu.

Về phần MA7 và MA30, cả hai có tương quan cao với các chỉ số giá, từ 0.95 đến 0.99, nhưng lại ngược chiều với volume, với giá trị -0.022 và -0.047 lần lượt. Điều này cho thấy trung bình động, dựa trên giá đóng cửa qua 7 hoặc 30 ngày, phản ánh xu hướng giá tốt, trong khi không liên quan nhiều đến biến động khối lượng. Quan hệ giữa MA7 và MA30 đạt 0.99, rất cao, do cả hai đều phụ thuộc vào giá đóng cửa với thời gian tính toán khác nhau. Ma trận này chứng tỏ dữ liệu đã được xử lý kỹ lưỡng, với các giá trị tương quan nhất quán, nhưng sự tách biệt của volume nhắc nhở ta rằng khối lượng là yếu tố độc lập, cần phân tích riêng để hiểu hành vi thị trường.

Từ đây, ta nhận thấy các chỉ số giá gần như tuyến tính, cho phép sử dụng bất kỳ chỉ số nào để phân tích xu hướng chung của Google, với sự khác biệt giữa chúng rất nhỏ. Tuy nhiên, tương quan âm nhẹ của volume với giá là điểm đáng chú ý, có thể xuất hiện khi giá giảm mạnh kéo theo khối lượng tăng, hoặc khi giá tăng mà không có sự hậu thuẫn từ giao dịch, báo hiệu xu hướng không bền vững. Mối quan hệ ngược giữa volume và trung bình động (-0.022, -0.047) cho thấy khối lượng không luôn hỗ trợ xu hướng giá dài hạn, trong khi tương quan cao của MA7 và MA30 với giá (0.95-0.99) khẳng định vai trò quan trọng của chúng trong dự đoán xu hướng, đặc biệt cho chiến lược đầu tư dài hạn.

Biểu đồ này không chỉ cung cấp bức tranh rõ ràng về tương quan giữa các chỉ số, mà còn mở ra cơ hội ứng dụng thực tiễn. Ta có thể tập trung vào close để đơn giản hóa mô hình, đồng thời theo dõi volume để phát hiện các điểm đảo chiều hoặc biến động bất thường. Kết hợp với trung bình động, biểu đồ hỗ trợ dự đoán xu hướng, đặc biệt trong các giai đoạn biến động như 2020-2021. Dữ liệu này còn cho phép so sánh với các cổ phiếu khác, giúp xác định vị thế riêng của Google trong ngành công nghệ, từ đó xây dựng chiến lược đầu tư hiệu quả hơn.



Tiếp theo là biểu đồ đường thể hiện xu hướng giá đóng cửa của cổ phiếu Google từ năm 2016 đến năm 2021. Trục hoành là thời gian, kéo dài từ đầu năm 2016 đến giữa năm 2021, với các mốc chính được đánh dấu là 2017, 2018, 2019, 2020, và 2021. Trục tung là giá đóng cửa, tính bằng USD, dao động từ khoảng 600 USD đến hơn 2000 USD. Đường màu xanh dương đại diện cho giá đóng cửa (Close Price), được vẽ liên tục để cho thấy sự thay đổi giá qua từng ngày giao dịch. Biểu đồ này giúp ta nắm bắt xu hướng dài hạn của giá cổ phiếu Google, đồng thời nhận diện các giai đoạn biến động đáng chú ý trong khoảng thời gian 5 năm này.

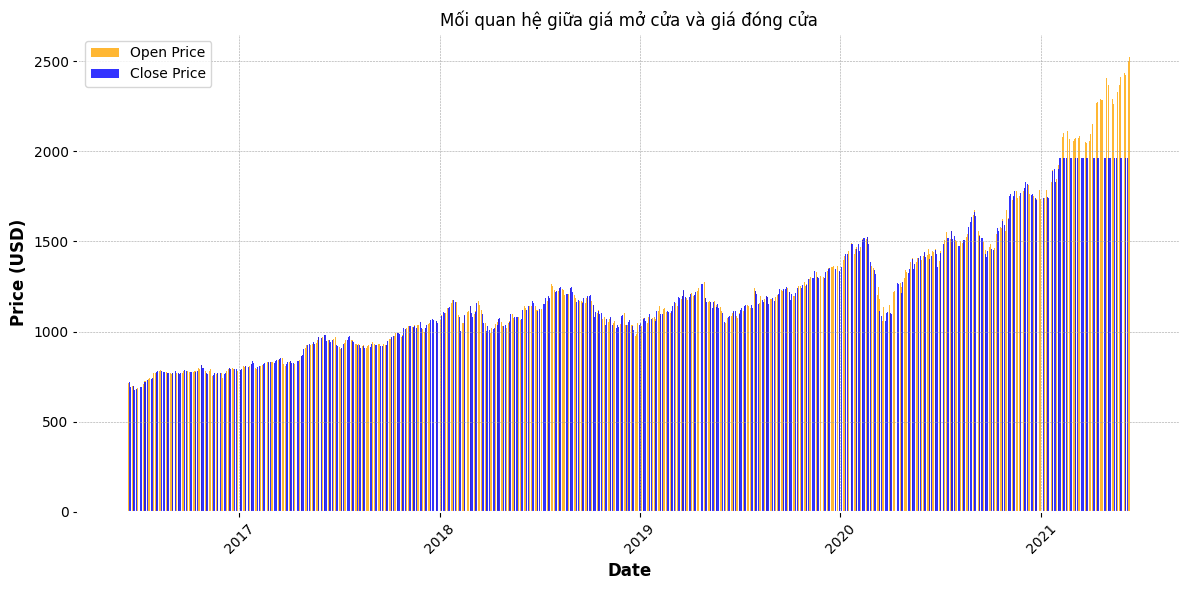
Nhìn tổng thể, giá cổ phiếu Google có xu hướng tăng mạnh từ 2016 đến 2021, với một số giai đoạn điều chỉnh ngắn hạn. Giá bắt đầu từ mức khoảng 700 USD vào đầu năm 2016 và tăng dần lên hơn 2000 USD vào giữa năm 2021, cho thấy mức tăng trưởng ấn tượng hơn 180%. Xu hướng này phản ánh sự phát triển mạnh mẽ của Google trong giai đoạn bùng nổ công nghệ, cùng với sự phục hồi kinh tế sau đại dịch COVID-19.

Vào năm 2016 và 2017, giá cổ phiếu Google dao động quanh mức 700 USD đến 900 USD, với tốc độ tăng trưởng khá ổn định. Giai đoạn này không có biến động lớn, cho thấy cổ phiếu Google đang trong giai đoạn tích lũy, khi công ty tiếp tục củng cố vị thế trong lĩnh vực tìm kiếm và quảng cáo trực tuyến. Đến năm 2018, giá bắt đầu tăng nhanh hơn, đạt mức khoảng 1100 USD vào giữa năm, nhưng sau đó có một đợt điều chỉnh nhẹ vào cuối năm 2018, khi giá giảm xuống còn khoảng 1000 USD. Đây có thể là ảnh hưởng từ biến động chung của thị trường chứng khoán, khi chỉ số S&P 500 và nhiều cổ phiếu công nghệ khác cũng giảm mạnh vào cuối năm 2018 do lo ngại về lãi suất và chiến tranh thương mại Mỹ-Trung.

Sang năm 2019, giá cổ phiếu phục hồi và tiếp tục tăng, dao động quanh mức 1100 USD đến 1300 USD, với xu hướng đi ngang nhẹ trong phần lớn thời gian. Điều này cho thấy thị trường đang chờ đợi thêm tín hiệu tích cực từ Google, có thể liên quan đến các sản phẩm mới hoặc báo cáo tài chính. Đến đầu năm 2020, ta nhận thấy một đợt sụt giảm mạnh vào khoảng tháng 3, khi giá giảm từ 1400 USD xuống còn khoảng 1100 USD. Đây là thời điểm đại dịch COVID-19 bùng phát, gây ra cú sốc lớn cho thị trường chứng khoán toàn cầu. Tuy nhiên, giá nhanh chóng phục hồi từ tháng 4/2020, tăng mạnh lên 1500 USD vào giữa năm 2020, phản ánh sự lạc quan của nhà đầu tư khi Google hưởng lợi từ xu hướng làm việc từ xa và nhu cầu sử dụng các dịch vụ như YouTube, Google Meet, và Google Cloud.

Giai đoạn từ cuối năm 2020 đến giữa năm 2021 là thời kỳ tăng trưởng mạnh nhất, với giá cổ phiếu tăng từ 1500 USD lên hơn 2000 USD, đạt đỉnh vào khoảng tháng 5/2021. Sự tăng trưởng này cho thấy niềm tin lớn vào tương lai của Google, đặc biệt khi công ty tiếp tục mở rộng các mảng kinh doanh như điện toán đám mây và công nghệ AI, đồng thời thị trường chứng khoán Mỹ nói chung cũng đạt đỉnh mới nhờ các gói kích thích kinh tế.

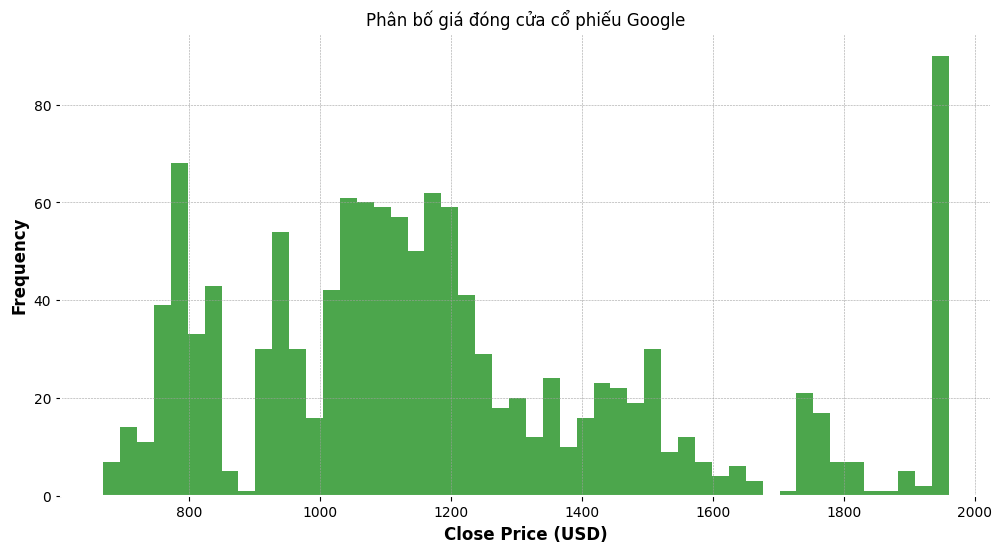
Biểu đồ cho thấy giá cổ phiếu Google có xu hướng tăng dài hạn rõ rệt, dù không tránh khỏi các đợt điều chỉnh ngắn hạn do ảnh hưởng của thị trường. Giai đoạn biến động mạnh nhất là đầu năm 2020, khi đại dịch COVID-19 gây ra cú sụt giảm lớn, nhưng sự phục hồi nhanh chóng sau đó chứng tỏ sức mạnh nội tại của Google. Công ty đã tận dụng tốt các cơ hội từ xu hướng số hóa và làm việc từ xa, giúp giá cổ phiếu tăng trưởng vượt bậc trong giai đoạn 2020-2021. Tuy nhiên, mức giá hơn 2000 USD vào giữa năm 2021 cũng đặt ra câu hỏi về khả năng định giá quá cao, vì tốc độ tăng trưởng nhanh có thể dẫn đến rủi ro điều chỉnh nếu thị trường đảo chiều.



Biểu đồ cột này thể hiện sự so sánh giữa giá mở cửa (màu cam) và giá đóng cửa (màu xanh dương) của cổ phiếu Google từ 2016 đến giữa năm 2021. Trục ngang là thời gian, với các mốc chính là 2017, 2018, 2019, 2020, và 2021, còn trục dọc là giá trị USD, từ 0 đến 2500 USD. Các cột xen kẽ giúp ta quan sát sự khác biệt giữa hai chỉ số giá hàng ngày, với tiêu đề "Mối quan hệ giữa giá mở cửa và giá đóng cửa" nhấn mạnh mục tiêu phân tích xu hướng và biến động qua thời gian. Nhìn chung, cả giá mở cửa và giá đóng cửa đều tăng mạnh, từ mức 700-800 USD năm 2016 lên trên 2000 USD vào giữa năm 2021, phản ánh sự phát triển vượt bậc của Google trong giai đoạn công nghệ bùng nổ và thị trường chứng khoán Mỹ đi lên, dù không thiếu các biến động ngắn hạn.

Từ 2016 đến đầu 2018, giá dao động từ 700 USD đến 1100 USD, với chênh lệch giữa giá mở cửa và đóng cửa khá nhỏ, cho thấy thị trường ổn định và Google đang trong giai đoạn tích lũy. Đến cuối 2018, một số ngày có sự chênh lệch rõ rệt, thường là giá đóng cửa thấp hơn giá mở cửa, trùng với thời điểm thị trường điều chỉnh do lo ngại về lãi suất và căng thẳng thương mại Mỹ-Trung. Sang năm 2019, giá tăng lên 1100-1300 USD, với hai chỉ số giá khá tương đồng, thể hiện thị trường cân bằng. Đầu năm 2020, giá sụt mạnh vào tháng 3, từ 1400 USD xuống 1100 USD, do ảnh hưởng của đại dịch COVID-19, nhưng phục hồi nhanh từ tháng 4, đạt 1500 USD giữa năm 2020, nhờ Google hưởng lợi từ xu hướng làm việc trực tuyến. Giai đoạn cuối 2020 đến giữa 2021, giá tăng vọt lên trên 2000 USD, đạt đỉnh vào tháng 5/2021, nhưng có một số ngày chênh lệch lớn, đặc biệt ở đoạn cuối, cho thấy biến động mạnh, có thể do tin tức hoặc giao dịch lớn.

Biểu đồ cho thấy giá mở cửa và đóng cửa thường đi cùng chiều, khẳng định xu hướng tăng dài hạn ổn định từ 2016 đến 2021, nhưng những ngày chênh lệch lớn vào cuối 2018 và giữa 2021 cho thấy biến động ngắn hạn, có thể do sự kiện tài chính hoặc biến động thị trường toàn cầu. Giai đoạn giảm mạnh tháng 3/2020 chứng minh tác động đồng đều của đại dịch lên cả hai chỉ số, nhưng sự phục hồi nhanh sau đó cho thấy Google tận dụng tốt xu hướng số hóa. Đỉnh giá trên 2000 USD vào tháng 5/2021, với chênh lệch tăng, gợi ý nguy cơ bong bóng giá, đòi hỏi ta theo dõi thêm các yếu tố kinh tế để dự đoán điều chỉnh. Biểu đồ này hữu ích để phân tích hành vi giao dịch, kết hợp với trung bình động hoặc ma trận tương quan, nhằm dự đoán xu hướng hoặc so sánh với các cổ phiếu khác trong ngành.



Biểu đồ trước mắt chúng ta là một histogram thể hiện sự phân bố của giá đóng cửa cổ phiếu Google, với trục ngang là giá trị giá đóng cửa tính bằng USD, dao động từ khoảng 600 USD đến hơn 2000 USD, và trục dọc là tần suất xuất hiện của các mức giá này. Histogram được chia thành 50 khoảng (bins), mỗi khoảng biểu thị một dải giá, và chiều cao của cột cho biết số lần giá đóng cửa rơi vào dải đó. Tiêu đề "Phân bố giá đóng cửa cổ phiếu Google" đã nói lên mục đích của biểu đồ: cung cấp cái nhìn về sự tập trung và phân tán của giá cổ phiếu trong suốt giai đoạn được khảo sát, có lẽ từ 2016 đến 2021 như ngữ cảnh trước đó.

Nhìn tổng quan, biểu đồ cho thấy giá đóng cửa không phân bố đều mà có xu hướng tập trung ở một số mức giá nhất định, với vài đỉnh nổi bật và một đuôi kéo dài về phía bên phải. Điều này gợi ý rằng giá cổ phiếu Google có sự biến thiên lớn, với phần lớn thời gian giá nằm ở mức trung bình thấp, nhưng cũng có những giai đoạn giá đạt mức cao bất thường, phản ánh sự tăng trưởng mạnh trong một số thời điểm.

Khi quan sát kỹ hơn, chúng ta thấy giá đóng cửa có tần suất cao nhất ở khoảng từ 1100 USD đến 1300 USD, nơi các cột đạt chiều cao khoảng 60-70. Đây là khu vực tập trung lớn nhất, cho thấy trong nhiều ngày giao dịch, giá cổ phiếu Google dao động quanh mức này, có thể tương ứng với giai đoạn ổn định từ 2018 đến 2019. Một đỉnh khác đáng chú ý nằm ở khoảng 700 USD đến 900 USD, với tần suất khoảng 50-60, có lẽ phản ánh giai đoạn đầu của dữ liệu, từ 2016 đến 2017, khi giá cổ phiếu còn ở mức thấp hơn.

Bên cạnh đó, histogram cũng có một số đỉnh nhỏ hơn ở các mức giá cao hơn, như khoảng 1500 USD đến 1600 USD và 1800 USD đến 1900 USD, với tần suất lần lượt khoảng 20-30. Những mức giá này có thể đại diện cho giai đoạn tăng trưởng mạnh từ 2020 đến 2021, khi Google hưởng lợi từ xu hướng số hóa và phục hồi kinh tế sau đại dịch. Đặc biệt, phần đuôi bên phải kéo dài đến hơn 2000 USD, với một cột cao ở khoảng 2000 USD, cho thấy có một số ngày giá đóng cửa đạt mức đỉnh, tương ứng với thời điểm giữa năm 2021 khi cổ phiếu Google đạt giá trị cao nhất trong giai đoạn khảo sát.

Hình dạng của histogram không đối xứng mà nghiêng về bên phải, với đuôi dài kéo về các mức giá cao. Điều này cho thấy phân bố giá đóng cửa có tính lệch phải, nghĩa là có nhiều ngày giá ở mức thấp hoặc trung bình, nhưng cũng có một số ngày giá tăng vọt, kéo dài đuôi phân bố. Sự lệch này phù hợp với xu hướng tăng trưởng dài hạn của Google, khi giá cổ phiếu tăng dần qua các năm, nhưng vẫn có những giai đoạn giá dao động ở mức thấp trong thời gian dài.

Histogram này mang đến một góc nhìn trực quan về sự phân tán của giá cổ phiếu Google, giúp chúng ta hiểu rõ hơn về mức độ biến thiên của giá trong giai đoạn 2016-2021. Khu vực tập trung cao ở mức 1100-1300 USD cho thấy giá cổ phiếu có xu hướng ổn định trong một thời gian dài, trong khi các mức giá cao hơn như 1800-2000 USD phản ánh những giai đoạn tăng trưởng mạnh, có thể do các yếu tố như báo cáo tài chính tích cực hoặc sự tăng trưởng của thị trường công nghệ. Sự lệch phải của phân bố cũng khẳng định rằng Google là một cổ phiếu có tiềm năng tăng trưởng lớn, nhưng đồng thời đi kèm với biến động đáng kể.

Phần cuối, ta sẽ đánh giá các số liệu quan trọng của hai mô hình hồi quy và áp dụng các tham số cần thiết để đánh giá được mức độ rủi ro của cổ phiếu công ty Google.

Ta sẽ tiếp tục với kết quả của hai mô hình hồi quy được áp dụng trên dữ liệu cổ phiếu Google. Trước hết, mô hình hồi quy tuyến tính được dùng để dự đoán giá đóng cửa (Close), và kết quả cho thấy Mean Squared Error (MSE) là 0.0, còn R2 Score đạt 0.99. MSE bằng 0.0 là một con số rất ấn tượng, cho thấy sai số bình phương trung bình giữa giá trị dự đoán và giá trị thực tế gần như không có, nghĩa là mô hình dự đoán cực kỳ chính xác trên tập dữ liệu kiểm tra. R2 Score 0.99 cũng khẳng định rằng mô hình giải thích được 99% sự biến thiên của giá đóng cửa, chứng tỏ các đặc trưng được chọn như giá mở cửa, khối lượng giao dịch, và trung bình động 7 ngày (MA7) có mối quan hệ rất mạnh với giá đóng cửa. Tuy nhiên, kết quả gần hoàn hảo như vậy cũng đặt ra câu hỏi liệu mô hình có bị quá khớp (overfitting) hay không, vì dữ liệu thực tế thường có nhiều nhiễu hơn. Ta cần kiểm tra thêm trên một tập dữ liệu mới hoặc áp dụng kỹ thuật như cross-validation để đảm bảo tính tổng quát của mô hình.

Tiếp theo, mô hình hồi quy logistic được sử dụng để dự đoán xu hướng giá (tăng hoặc giảm), và độ chính xác đạt 53.17%. Con số này cho thấy mô hình chỉ đúng trong khoảng hơn một nửa số trường hợp, không quá ấn tượng, gần như tương đương với việc đoán ngẫu nhiên (50%). Điều này có thể do các đặc trưng như giá mở cửa, khối lượng giao dịch, và MA7 chưa đủ để dự đoán chính xác xu hướng, hoặc dữ liệu có nhiều biến động phức tạp mà mô hình logistic đơn giản không thể nắm bắt. Ngoài ra, việc giá cổ phiếu Google có xu hướng tăng dài hạn từ 2016 đến 2021 có thể khiến dữ liệu bị mất cân bằng, với nhiều ngày tăng hơn giảm, làm giảm hiệu quả của mô hình. Để cải thiện, ta có thể thử thêm các đặc trưng khác như chỉ số biến động hoặc dữ liệu vĩ mô, hoặc sử dụng các mô hình phức tạp hơn như Random Forest hay mạng nơ-ron để dự đoán xu hướng. Nhìn chung, mô hình hồi quy tuyến tính cho kết quả rất tốt trong việc dự đoán giá trị cụ thể, nhưng dự đoán xu hướng bằng hồi quy logistic vẫn cần được cải thiện đáng kể.

Bây giờ ta chuyển sang phân tích các chỉ số rủi ro của cổ phiếu Google, được tính dựa trên tỷ suất lợi nhuận hàng ngày (daily return). Độ lệch chuẩn (Std) là 1.63%, cho thấy mức độ biến động của lợi nhuận hàng ngày không quá lớn, tương đối ổn định so với một cổ phiếu công nghệ vốn thường có biến động cao. Giá trị VaR (95%) là -2.49%, nghĩa là trong 95% trường hợp, lợi nhuận hàng ngày sẽ không giảm quá 2.49%, một mức rủi ro khá thấp, phù hợp với xu hướng tăng trưởng dài hạn của Google. Tuy nhiên, CVaR (95%) là -4.03%, cho biết trong 5% trường hợp xấu nhất, lợi nhuận trung bình có thể giảm tới 4.03%, phản ánh rủi ro tiềm ẩn trong những ngày thị trường biến động mạnh, như giai đoạn đầu đại dịch COVID-19 vào tháng 3/2020.

Sharpe Ratio đạt 0.06, một con số khá thấp, cho thấy tỷ lệ lợi nhuận trên mỗi đơn vị rủi ro không cao, có thể do lợi nhuận trung bình hàng ngày nhỏ so với độ biến động. Điều này gợi ý rằng mặc dù Google có xu hướng tăng dài hạn, nhưng lợi nhuận hàng ngày không đủ lớn để bù đắp cho rủi ro, đặc biệt nếu so sánh với các tài sản ít rủi ro hơn. Cuối cùng, Maximum Drawdown là -30.79%, nghĩa là trong giai đoạn khảo sát, giá cổ phiếu Google từng giảm tối đa 30.79% từ đỉnh xuống đáy, có thể xảy ra vào đợt sụt giảm năm 2020 do đại dịch. Con số này cho thấy cổ phiếu Google có thể chịu rủi ro giảm giá đáng kể trong ngắn hạn, đặc biệt trong các giai đoạn khủng hoảng.

# CHƯƠNG 2: ĐÁNH GIÁ SỰ RỦI RO CỦA CỔ PHIẾU

## CƠ SỞ ĐÁNH GIÁ

Link tải: <https://www.kaggle.com/datasets/sunnykusawa/stock-price-dataset-eod>

Điểm Usability: 9.41

"Stock Price Dataset EOD" chứa giá cổ phiếu cuối ngày (End of Day - EOD) của 16 công ty thuộc các sàn giao dịch khác nhau.

Các công ty bao gồm: Google (GOOGL), Apple (AAPL), Amazon (AMZN), Microsoft (MSFT), Facebook (FB), Tesla (TSLA), Walmart (WMT), Alibaba (BABA), JPMorgan (JPM), Johnson & Johnson (JNJ), Mastercard (MA), Disney (DIS), NVIDIA (NVDA), PayPal (PYPL), Netflix (NFLX), Abbott Laboratories (ABT).

Các cột dữ liệu trong dataset:

|  |  |  |  |
| --- | --- | --- | --- |
| **Tên cột thuộc tính** | **Kiểu dữ liệu** | **Ý nghĩa** | **Kiểu dữ liệu** |
| open | float | Giá mở bán của cổ phiếu trong ngày | Liên tục |
| high | float | Giá cao nhất của cổ phiếu trong ngày | Liên tục |
| low | float | Giá thấp nhất của cổ phiếu trong ngày | Liên tục |
| close | float | Giá đóng cửa của cổ phiếu trong ngày | Liên tục |
| volume | float | Số lượng cổ phiếu được giao dịch | Liên tục |
| adj-high | float | Giá mở cửa của cổ phiếu đã điều chỉnh | Liên tục |
| adj-low | float | Giá cao nhất của cổ phiếu đã điều chỉnh | Liên tục |
| adj-close | float | Giá thấp nhất của cổ phiếu đã điều chỉnh | Liên tục |
| adj-open | float | Giá đóng cửa của cổ phiếu đã điều chỉnh | Liên tục |
| adj-volume | float | Khối lượng giao dịch đã điều chỉnh | Liên tục |
| symbol | string | Mã chứng khoán của công ty | Rời rạc |
| exchange | string | Tên sàn giao dịch | Rời rạc |
| date | string | Ngày giao dịch | Rời rạc |

## DẪN CHỨNG VỀ SỐ LIỆU CỔ PHIẾU

Đầu tiên, ta sẽ thống kê cơ bản cho Thuộc tính Close, Volume của toàn bộ cổ phiếu của các công ty, bao gồm: giá trị lớn nhất, giá trị nhỏ nhất, giá trị trung bình, trung vị, độ lệch chuẩn.

Kết quả thống kê như sau:

A table of numbers and letters

AI-generated content may be incorrect.

Chúng ta có thể phân tích các giá trị từ bảng thống kê như sau:

**Cổ phiếu có giá trung bình cao nhất:** AMZN (3198.74).

Amazon là cổ phiếu đắt giá nhất danh sách, phản ánh giá trị thị trường lớn và kì vọng cao của nhà đầu tư đối với công ty, liên hệ thực tế ta có thể thấy, vào cuối năm 2020 Amazon được hưởng lợi từ sự tăng trưởng mạnh mẽ trong thương mại điện tử, đặc biệt trong bối cảnh đại dịch COVID-19, khi nhu cầu mua sắm trực tuyến tăng vọt. Giá cao cũng đi kèm với quy mô vốn hóa lớn và niềm tin của thị trường vào khả năng tăng trưởng dài hạn của Amazon, đặc biệt trong lĩnh vực điện toán đám mây(AWS) và các mảng kinh doanh khác

**Cổ phiếu có giá trung bình thấp nhất:** ABT (107.13 USD).

Abbott Laboratories (ABT) có giá trung bình thấp nhất, điều này không có nghĩa công ty ABT là một công ty nhỏ hay ít giá trị, mà chỉ đơn giản là giá cổ phiếu của họ thấp hơn so với các công ty khác trong danh sách. ABT hoạt động trong lĩnh vực y tế(Thiết bị y tế, dược phẩm), vốn thường có giá cổ phiếu ổn định và không biến động như các công ty công nghệ. Có thể giá thấp hơn là phản ánh chiến lược chia tách cổ phiếu hoặc giá trị vốn hóa thị trường nhỏ hơn so với các công ty công nghệ lớn. ABT thường được xem là một cổ phiếu phòng thủ(Defensive stock), hấp dẫn với các nhà đầu tư tìm kiếm sự ổn định.

**Cổ phiếu có biến động giá cao nhất**: TSLA (507.35 USD).

Tesla(TSLA) có biến động cho giá đóng cửa cao nhất(Độ lệch chuẩn). Cho thấy giá cổ phiếu của TSLA biến động mạnh mẽ trong khoảng thời gian từ tháng 8 đến tháng 12 năm 2020. Điều này phản ánh sự không chắc chắn và tâm lý thị trường dao động mạnh liên qua đến TSLA, vốn là một cổ phiếu rất được quan tâm vào thời điểm đó. Vào năm 2020, cổ phiếu TSLA chứng kiến sự tăng trưởng bùng nổ được thúc đẩy bởi niềm tin vào tương lai xe điện, các thông báo tích cực về Elon Musk. Tuy nhiên biến động cao cũng cho thấy rủi ro lớn, vì giá có thể giảm mạnh nếu tâm lý thị trường thay đổi.

**Cổ phiếu có biến động giá thấp nhất:** ABT (3.27 USD).

ABT lại có biến động giá thấp nhất cho thấy giá cổ phiếu của ABT rất ổn định, ít biến động trong khoảng thời gian này. ABT là cổ phiếu trong ngành y tế, vốn ít chịu ảnh hưởng bởi tâm lý thị trường ngắn hạn. Vào năm 2020, ABT cũng được hưởng lợi từ nhu cầu xét nghiệm COVID-19, nhưng ngành y tế có xu hướng ổn định hơn so với các ngành công nghệ. Điều này làm ABT trở thành lựa chọn tốt cho các nhà đầu tư muốn tránh rủi ro

**Cổ phiếu có khối lượng giao dịch trung bình cao nhất:** AAPL (121047765.22 USD).

Apple có khối lượng giao dịch trung bình cao nhất điều đó thể hiện cổ phiếu của Apple thu hút sự quan tâm lớn từ nhà đầu tư cá nhân và tổ chức. Apple là một trong những công ty có vốn hóa lớn nhất thế giới vào năm 2020, sự phổ biến của cổ phiếu này đến từ tính thanh khoản cao, niềm tin vào thương hiệu Apple và các sản phẩm chủ lực như iPhone, iPad. Ngoài ra, AAPL đã chia tách cổ phiếu 4:1 vào ngày 31/8/2020. Điều này làm tăng tính thanh khoản và thu hút thêm nhà đầu tư nhỏ lẻ, dẫn đến khối lượng giao dịch lớn.

**Cổ phiếu có khối lượng giao dịch trung bình thấp nhất**: GOOGL (1,676,304.48 cổ phiếu).

Cổ phiếu Google ít được giao dịch hơn so với các cổ phiếu khác trong dataset. Giá cổ phiếu của Google khá cao, điều này có thể làm giảm tính thanh khoản, vì nhà đầu tư nhỏ lẻ ít có khả năng mua số lượng lớn. Ngoài ra, GOOGL không chia tách cổ phiếu thường xuyên, dẫn đến khối lượng giao dịch thấp hơn so với các cổ phiếu khác.

**Cổ phiếu có biến động khối lượng giao dịch cao nhất:** AAPL (58382674.37 cổ phiếu).

AAPL không chỉ có khối lượng giao dịch trung bình cao mà còn có độ lệch chuẩn khối lượng giao dịch lớn nhất. Điều này cho thấy khối lượng giao dịch của AAPL dao động mạnh qua các ngày, có những ngày giao dịch rất sôi động nhưng cũng có những ngày trầm lắng. Sự biến động này liên quan đến các sự kiện lớn của Apple như thông báo chia tách cổ phiếu, ra mắt sản phẩm mới(iPhone 12 vào tháng 10/2020), hoặc các báo cáo tài chính. Những sự kiện này thu hút nhiều sự chú ý dẫn đến khối lượng giao dịch tăng đột biến.

**Cổ phiếu có phạm vi khối lượng giao dịch thấp nhất:** GOOGL (4530559.0 cổ phiếu).

GOOGL có độ lệch chuẩn khối lượng giao dịch thấp nhất cho thấy sự ổn định của khối lượng giao dịch qua các ngày. Điều này phù hợp với khối lượng giao dịch trung bình thấp GOOGL, và GOOGL không có các sự kiện lớn gây đột biến như chia tách cổ phiếu.

Tiếp theo, ta phân tích các giá trị bất thường(Outliner) của dataset, ta dùng phương pháp IQR để phát hiện outliers:

A screenshot of a computer screen

AI-generated content may be incorrect.

**Phân tích giá trị bất thường trong giá đóng cửa:**

Ta thấy, AAPL và TSLA cùng với MA đều có nhiều outliers trong giá đóng cửa:

* AAPL đã trải qua các biến động giá lớn liên quan đến sự kiện chia tách cổ phiếu 4:1 vào ngày 31/08/2020 và sự ra mắt iPhone 12 vào tháng 10/2020.
* TSLA cũng có biến động giá mạnh, điều này liên quan đến việc tâm lý thị trường sôi động sung quanh ngành xe điện.
* MA(Mastercard Incorporated) có biến động cũng nhiều đa số ảnh hưởng vì sự thay đổi hành vi tiêu dùng, cụ thể là người dùng chuyển sang thanh toán không dùng tiền mặt nhiều hơn.
* GOOGL, WMT, BABA, JPM, NFLX: Không có outliers trong giá đóng cửa, Vì các cổ phiếu này ổn định trong khoảng thời gian được phân tích, không có biến động bất thường vượt quá ngưỡng.

**Phân tích giá trị bất thường trong khối lượng giao dịch:**

* Baba: khối lượng giao dịch dao động bất thường nhiều lần, do các tin tức liên quan đến Alibaba, áp lực từ chính phủ Trung Quốc và báo cáo tài chính.
* Dis: khối lượng giao dịch dao động bất thường nhiều lần, do Disney đã trải qua các ngày giao dịch sôi động do các thông báo về dịch vụ phát trực tuyến Disney+ và các tin tức liên quan đén ngành giải trí.
* MSFT: Khối lượng giao dịch tăng đột biến do các báo cáo tài chính và tin tức về mảng điện toán đám mây(Azure).
* NFLX: Khối lượng giao dịch tăng mạnh do báo cáo tài chính và các thông báo về nội dung.

Các cổ phiếu có ít outliers là do có khối lượng giao dịch ổn định và không biến động mạnh.

Ta có biểu đồ thể hiện xu hướng biến động giá đóng cửa của tất cả các công ty như sau:

A graph of a graph

AI-generated content may be incorrect.

Từ biểu đồ, ta có thể phân tích được:

**Cổ phiếu có giá cao nhất(AMZN, GOOGL, TSLA):**

* AMZN: Có giá trị cao nhất cũng như có xu hướng ổn định trong số các cổ phiếu, điều này cho thấy Amazon là cổ phiếu có giá trị cao và ổn định.
* GOOGLE: Có giá trị cao và ổn định, phản ánh tính chất bền vững của Google.
* TSLA: Giá tăng mạnh vào lúc đầu, sau đó giảm mạnh và ổn định lại. Sự tăng đột biến liên quan đến việc Tesla chia tách cổ phiếu 5:1 vào ngày 31/08/2020, khiến giá tăng mạnh do tâm lý thị trường, sau đó giá giảm do điềm chỉnh sau chia tách.

**Cổ phiếu có giá trung bình(NVDA, NFLX, BABA, MA):**

* NVDA: Giá trung bình, có một số biến động nhỏ do NVDA hưởng lợi từ nhu cầu GPU tăng cao trong dịch(gaming, AI, làm việc từ xa).
* NLFX: Giá trung bình, tăng nhẹ vào tháng 9, giảm vào tháng 10 và tăng trở lại vào tháng 12. Netflix có thể đã chịu ảnh hưởng từ biến động trong ngành giải trí(giảm giá khi công bố báo cáo tài chính không đạt kỳ vọng vào tháng 10 nhưng tăng trở lại nhờ nhu cầu xem phim tại nhà).
* BABA: Giá trung bình, giá tăng từ tháng 8 đến tháng 11, nhưng giảm mạnh vào cuối tháng 12 do áp lực từ chính phủ Trung Quốc đối với Alibaba.
* MA: Giá trung bình, ổn định với xu hướng tăng nhẹ, Mastercard hưởng lợi từ xu hướng thanh toán không dùng tiền mặt trong đại dịch.

**Cổ phiếu có giá thấp hơn(AAPL, MSFT, FB, WMT, JPM, JNJ, ABT, DIS, PYPL)**

* AAPL: Giá biến tăng đều từ tháng 8 đến tháng 12/2020 do chia tách cổ phiếu và sự ra mắt iphone12.
* MSFT: Tăng nhẹ và ổn định do điện toán đám mây và nhu cầu làm việc từ xa.
* FB: Tăng từ tháng 8 đến tháng 11 nhưng giảm nhẹ vào tháng 12, do các quy định và cạnh tranh trong quảng cáo.
* DIS: Giá tăng mạnh từ tháng 11 đến tháng 12 do các tin tức tích cực về Disney+ và sự phục hồi của ngành giải trí.
* PYPL: Giá tăng đều từ tháng 8 đến tháng 12/2020 do xu hướng thanh toán trực tuyến.
* WMT, JPM, JNJ, APT: Xu hướng tăng nhẹ, giá ổn định ít biến động vì là cổ phiếu phòng thủ (WMT: bán lẻ, JPM: ngân hàng, JNJ và ABT: y tế)

Tương tự ta cũng có biểu đồ về khối lượng giao dịch của các công ty:

A graph of different colored lines

AI-generated content may be incorrect.

**Cổ phiếu có khối lượng giao dịch cao nhất (AAPL, TSLA):**

* AAPL: Có nhiều ngày khối lượng giao dịch tăng đột biến do sự kiện chia tách cổ phiếu và sự ra mắt iPhone 12 cũng như tâm lý thị trường tích cực vào cuối năm.
* TSLA: Có nhiều ngày khối lượng giao dịch tăng đột biến do Tesla được thêm vào chỉ số S&P 500 vào ngày 21/12/2020 và sự chia tách cổ phiếu ngày 31/08/2020.

**Cổ phiếu có khối lượng giao dịch trung bình (MSFT, FB, BABA, DIS):**

* MSFT: có một số ngày tăng đột biến do báo cáo tài chính và tin tức về mảng điện toán đám mây.
* FB: Khối lượng giao dịch có một số ngày tăng đột biến(Tháng 9 và tháng 12/2020) do báo cáo tài chính và tin tức về ngành quảng cáo.
* BABA: Khối lượng giao dịch có nhiều ngày tăng đột biến do áp lực từ chính phủ Trung Quốc(Cuộc điều tra chống độc quyền đối với Alibaba).
* DIS: Khối lượng giao dịch có nhiều ngày tăng đột biến do tin tức tích cực về Disney+ và sự phục hồi của ngành giải trí.

**Cổ phiếu có khối lượng giao dịch thấp (GOOGL, AMZN, MA, NFLX, ABT):**

* GOOGL: Cổ phiếu khá ổn định với một số ngày tăng nhẹ.
* AMZN: Khối lượng giao dịch ổn định, thấp một phần do giá cổ phiếu cao.
* MA: Có một số ngày tăng đột biến do tin tức và báo cáo tài chính.
* NFLX và ABT: Cũng có một số ngày tăng đột biến nhưng nhìn chung là khá ổn định.

Từ giá đóng cửa và số lượng giao dịch, ta có thể tính được tổng số giao dịch của các công ty trong bộ dữ liệu:

A graph with colorful lines

AI-generated content may be incorrect.

Với T có nghĩa là nghìn tỉ.

**Nhóm cổ phiếu có tổng số giao dịch cao nhất**

TSLA dẫn đầu với tổng số giao dịch hơn 2.74 nghìn tỷ USD, cho thấy thanh khoản cực kì cao. Lý do là vì sau sự kiện chia tách cổ phiếu theo tỉ lệ 5:1(31/08/2020) cùng với việc được thêm vào danh sách S&P 500 gây ra giao dịch bùng nổ.

AAPL đứng thứ 2 với tổng số giao dịch hơn 1.67 nghìn tỷ USD đứng thứ 2. Lý do lớn nhất là vì sự ra mắt Iphone 12(vào tháng 10/2020) kèm thêm với xu hướng làm việc từ xa trong dịch Covid-19 làm tăng thêm nhu cầu thiết bị Apple. Ngoài ra còn sau sự kiện chia tách 5:1(31/08/2020). Cổ phiếu AAPL có thanh khoản cao.

AMZN có tổng số giao dịch hơn 1.46 nghìn tỷ USD, đứng thứ 3. Lý do là vì AMZN cũng hưởng lợi lớn từ nhu cầu thương mại điện tử trong đại dịch. Một phần làm AMZN có tổng số giao dịch cao đến vậy là do giá bán của AMZN cao.

MSFT có tổng số giao dịch khoảng 641 tỷ USD đứng thứ tư. MSFT hưởng lợi từ xu hướng làm việc từ xa(Microsoft Team, Azure).

NVDA và FB: Có tổng số giao dịch gần bằng nhau, khoảng hơn 550 tỷ USD. Tương tự như AMZN và MSFT, 2 cổ phiếu này tăng do nhu cầu chip và sự phổ biến của mạng xã hội trong mùa dịch.

BABA, NFLX, GOOGL, PyPL: Đây là nhóm cổ phiếu có tổng số giao dịch trung bình.

Ta có thể vẽ biểu đồ nến cho giá cổ phiếu của AAPL với các thông tin: Giá mở cửa, giá đóng cửa, giá cao nhất, giá thấp nhất trong ngày(Từ 24/11/2020 đến 31/12/2020).

A graph with different colored rectangles

AI-generated content may be incorrect.

**Giai đoạn 24/11/2020 - 04/12/2020:**

24/11/2020 - 27/11/2020:Giá giảm xuống mức thấp nhất vào ngày 25/11 sau đó giá phục hồi nhẹ vào ngày 27/11. Cho thấp áp lực bán trong những ngày đầu nhưng lực mua đã đẩy giá trở lại.

30/11/2020 - 04/12/2020: Giá bắt đầu tăng vào ngày 03/12 cho thấy lực mua mạnh và ít kháng cự, nhưng vào ngày 04/12 xuất hiện nến đỏ nhỏ cho thấy có một chút điều chỉnh.

**Giai đoạn 07/12/2020 - 14/12/2020:**

Giá dao động với nhiều nến nhỏ xen kẽ đỏ và xanh cho thấy thị trường đang trong giai đoạn tích lũy, không có xu hướng rõ ràng. Ngày 10/12 có nến đỏ bóng trên dài tức là giá đã tăng nhưng bị đẩy xuống thể hiện áp lực bán tại mức kháng cự.

**Giai đoạn 15/12/2020 - 24/12/2020**

15/12/2020 - 17/12/2020: Giá tăng mạnh cho thấy lực mua áp đảo, bóng nến ngắn cho thấy ít áp lực bán.

8/12/2020 - 24/12/2020: Có sự điều chỉnh nhẹ sau đà tăng mạnh, nhưng xu hướng tăng vẫn được duy trì

**Giai đoạn 28/12/2020 - 03/01/2021**

Giá tiếp tục tăng và đạt đỉnh, nhưng có áp lực bán mạnh tại mức đỉnh, đây là dấu hiệu của sự đảo chiều ngắn hạn, giảm nhẹ vào 03/01/2021.

**Mức hỗ trợ:** Khoảng 120–122 USD (giai đoạn tích lũy 07/12–14/12), và mức 114–116 USD (giai đoạn đầu).

**Mức kháng cự:** Khoảng 124 USD (trước breakout ngày 16/12) và 137 USD (mức đỉnh ngày 29/12). Giá không vượt qua được mức 137 USD và điều chỉnh sau đó.

Ta vẽ biểu đồ Box Diagram cho Daily Return, Daily Return là tỉ lệ thay đổi phần trăm của giá đóng cửa giữa 2 ngày liên tiếp, biểu thị mức độ tăng hoặc giảm giá hằng ngày.

A diagram of symbols and numbers

AI-generated content may be incorrect.

Phân tích từ biểu đồ:

Cổ phiếu công nghệ tăng trưởng cao (TSLA, AAPL, DIS, PYPL, NFLX) có lợi nhuận hằng ngày trung bình cao nhưng cũng có biến động mạnh.

Cổ phiếu phòng thủ(JNJ, ABT, WMT, JPM): Lợi nhuận hàng ngày nhỏ và ổn định phù hợp với vai trò của chúng trong các ngành ít chịu ảnh hưởng từ biến động thị trường.

Cổ phiếu công nghệ ổn định (GOOGL, AMZN, MSFT): Lợi nhuận hàng ngày phân phối khá cân đối, cho thấy sự ổn định tương đối dù thuộc ngành công nghệ.

A red and blue grid with white text

AI-generated content may be incorrect.

Nhóm giá (Open, High, Low, Close) có độ tương quan gần bằng 1 → Các giá trị này luôn có xu hướng biến động cùng nhau. Trên thực tế là hợp lý, bởi vì giá mở cửa, cao nhất, thấp nhất và đóng cửa của cổ phiếu thường dao động theo xu hướng chung trong ngày. Vậy nên các giá trị này luôn có xu hướng biến động cùng nhau. Xét trên thực tế, điều này hợp lý vì giá mở cửa, cao nhất, thấp nhất và đóng cửa của cổ phiếu thường dao động theo xu hướng chung trong ngày.

Tương quan cao giữa Volume và Adj\_Volume (~0.91) → Điều này cho thấy hai giá trị này gần như phản ánh cùng một thông tin. Đối với các giá trị âm ở adj\_volume, khối lượng giao dịch tăng thì giá có xu hướng giảm, có thể do áp lực bán tăng.

Adj\_Open, Adj\_Close, Adj\_High, Adj\_Low có tương quan cao (~0.98) với các giá tương ứng → Giá điều chỉnh thường rất sát với giá gốc.

Sơ đồ mạng:

A network of lines and dots

AI-generated content may be incorrect.

Cụm ngành công nghệ: GOOGL, MSFT, FB, AMZN, NFLX, NVDA, AAPL, PYPL, DIS có tương quan cao với nhau cho thấy chúng chịu ảnh hưởng chung từ các yếu tố như tâm lý thị trường, nhu cầu công nghệ trong đại dịch, và sự tăng trưởng của thương mại điện tử, giải trí trực tuyến.

Cụm ngành tài chính và kinh tế: MA, JPM, JNJ, BABA có tương quan cao với nhau, cho thấy chúng chịu ảnh hưởng từ các yếu tố như chi tiêu tiêu dùng, lãi suất, và tính chất phòng thủ trong giai đoạn biến động.

Cổ phiếu độc lập (TSLA): có tương quan âm với nhiều cổ phiếu, cho thấy nó di chuyển độc lập và chịu ảnh hưởng mạnh từ các yếu tố riêng (như chia tách cổ phiếu, tin tức về ngành xe điện, và động thái của Elon Musk).

WMT và ABT: ít tương quan với các cổ phiếu công nghệ và có tương quan âm với TSLA, cho thấy chúng là các lựa chọn an toàn trong giai đoạn thị trường biến động.

A graph with a line

AI-generated content may be incorrect.

Từ mô hình Linaer Regression ta có thể kết luận mô hình hoạt động khác tốt đối với dữ liệu với giá thực tế gần giống với giá dự đoán với MSE thấp.

## ĐÁNH GIÁ VỀ RỦI RO CỔ PHIẾU

Ta phân tích rủi ro của các công ty dựa vào các chỉ số sau:

* Std (Độ lệch chuẩn): Đo lường mức độ biến động của lợi nhuận hàng ngày. Giá trị cao hơn cho thấy cổ phiếu có rủi ro cao hơn.
* VaR (95%) (Value at Risk): Mức lỗ tối đa dự kiến ở mức tin cậy 95% trong một ngày. Giá trị âm lớn hơn (ví dụ: -0.0767) cho thấy rủi ro lỗ lớn hơn.
* CVaR (95%) (Conditional Value at Risk): Mức lỗ trung bình trong 5% trường hợp xấu nhất. Giá trị âm lớn hơn cho thấy rủi ro lỗ nghiêm trọng hơn.
* Sharpe Ratio: Tỷ lệ lợi nhuận điều chỉnh theo rủi ro, tính bằng lợi nhuận trung bình chia cho độ lệch chuẩn (trừ đi lãi suất phi rủi ro, nếu có). Giá trị cao hơn cho thấy hiệu suất tốt hơn trên mỗi đơn vị rủi ro.
* Maximum Drawdown: Mức sụt giảm lớn nhất từ đỉnh đến đáy trong giai đoạn, phản ánh rủi ro giảm giá tối đa.

Ta có bảng số liệu cùng với các biểu đồ:

A table of numbers with black text

AI-generated content may be incorrect.

Từ đây, ta sẽ bắt đầu phân tích:

**Cổ phiếu có Sharpe Ration dương(Hiệu suất tốt trên mỗi đơn vị rủi ro)**

* **AAPL (Sharpe Ratio = 1.5167):** Có chỉ số Sharpe Ratio cao nhất, cho thấy hiệu suất vượt trội trên mỗi đơn vị rủi ro, tuy nhiên lại có chỉ số Std cao thứ 2 phản ánh biến động lớn, Var và Cvar ở mức trung bình. Maximum Drawdown = -0.2184 (sụt giảm tối đa 21.84%), chấp nhận được so với các cổ phiếu rủi ro cao khác. Liên hệ thực tế AAPL là cổ phiếu có tiềm năng tăng trưởng dài hạn nhờ hệ sinh thái mạnh. Đây là cổ phiếu đáng đầu tư ngắn hạn, có tiềm năng đầu tư dài hạn nhưng cần phải theo dõi biến động giá do Std cao.
* **TSLA (Sharpe Ratio = 1.2558):** Có chỉ số Sharpe Ratio cao thứ hai, cho thấy lợi nhuận lớn trên mỗi đơn vị rủi ro, tuy nhiên lại có chỉ số Std cao nhất phản ánh biến động mạnh, chỉ số Var và Cvar cao nhất cho thấy rủi ro lỗ lớn, Maximum Drawdown cao nhất. Liên hệ với thực tế TSLA là cổ phiếu tăng trưởng dẫn đầu ngành xe điện nhưng định giá cao và biến động lớn khiến nó tồn tại nhiều rủi ro. Đây là cổ phiếu đáng đầu tư cho nhà đầu tư ngắn hạn chấp nhận rủi ro cao, muốn tận dụng đà tăng trưởng nhưng cần rất thận trọng với biến động mạnh và khả năng điều chỉnh giá.
* **BABA (Sharpe Ratio = 0.5802):** Chỉ số Sharpe Ratio dương, cao thứ 3 nhưng thấp hơn nhiều so với TSLA và APPL, cho thấy lợi nhuận vừa phải trên mỗi đơn vị rủi ro. BABA có chỉ số std thấp cho thấy biến động nhỏ, các chỉ số var, cVar cho thấy rủi ro lỗ ở mức trung bình. Maximum Drwadown = -0.2167. Liên hệ với thực tế BABA là công ty thương mại điện tử hàng đầu Trung Quốc, có tiềm năng dài hạn nếu vượt qua các rủi ro quy định nhưng trong khoảng thời gian này hiệu suất không nổi bật. Nên có thể đầu tư cho nhà đầu tư dài hạn và muốn đa dạng hóa sang thị trường Trung Quốc, nhưng cần theo dõi sát các quy định và tin tức.

**Cổ phiếu có Sharpe Ratio âm (Hiệu suất không vượt trội)**

* **AMZN(Sharpe Ratio = -0.2340):** Chỉ số Sharpe Ratio âm cho thấy lợi nhuận không bù đắp được rủi ro. Std thấp cho thấy biến động nhỏ, Var và Cvar cho thấy rủi ro lỗ ở mức trung bình, Maximum Drawdown = -0.1419 (sụt giảm tối đa 14.19%), thấp, cho thấy AMZN khá ổn định. AMZN là công ty công nghệ hàng đầu, dẫn đầu trong thương mại điện tử và điện toán đám mây (AWS). Dù hiệu suất ngắn hạn không vượt trội, AMZN có tiềm năng dài hạn mạnh. Đây là một lựa chọn đáng cân nhắc cho các nhà đầu tư dài hạn nhờ tiềm năng tăng trưởng và rủi ro thấp. Tuy nhiên không phù hợp cho đầu tư ngắn hạn trong giai đoạn này.
* **MSFT (Sharpe Ratio = -0.6145)**: Sharpe Ratio âm cho thấy hiệu suất không vượt trội. Std rất thấp, cho thấy biến động nhỏ, cổ phiếu có rủi ro lỗ thấp cùng với Maximum Drawdown = -0.1220 (sụt giảm tối đa 12.2%), thấp nhất, cho thấy MSFT rất ổn định. MSFT là cổ phiếu công nghệ ổn định, dẫn đầu trong phần mềm và điện toán đám mây, có tiềm năng dài hạn mạnh. Đây là sự lựa chọn đáng cân nhắc cho nhà đầu tư dài hạn muốn sự ổn định, không phù hợp cho đầu tư ngắn hạn trong giai đoạn này.

**Cổ phiếu phòng thủ (Rủi ro thấp):**

* **JNJ (Sharpe Ratio = -0.8712):** Sharpe Ratio âm cho thấy lợi nhuận không vượt trội, tuy nhiên chỉ số Std lại thấp nhấp, var cùng với Cvar cũng thấp nhất kèm thêm Maximum Drawdown = -0.1214 (sụt giảm tối đa 12.14%), rất thấp. Thực tế cho thấy JNJ có tiềm năng dài hạn nhờ vị thế trong ngành y tế. Đây là cổ phiếu đáng đầu tư cho nhà đầu tư dài hạn muốn an toàn, nhưng nếu đầu tư ngắn hạn thì lợi nhuận thấp.
* **WMT (Sharpe Ratio = -1.1242):** Sharpe Ratio âm phản ánh hiệu suất thấp. chỉ số std cùng với var và Maximum drawdown ở mức thấp. Trên thực tế WMT là cổ phiếu phòng thủ (bán lẻ), hưởng lợi từ nhu cầu tiêu dùng trong đại dịch, nhưng không tăng mạnh. WMT ổn định, phù hợp với nhà đầu tư muốn an toàn. Đây là cổ phiếu đáng đầu tư cho nhà đầu tư dài hạn muốn an toàn, nhưng không phù hợp cho đầu tư ngắn hạn.

**Cổ phiếu cần tránh**

* **GOOGL (Sharpe Ratio = -1.3288):** Sharpe Ratio âm lớn cho thấy hiệu suất không vượt trội. Chỉ số std cùng với rủi ro lỗ thấp. Tuy nhiên Maximum Drwadown = -0.2277 (sụt giảm tối đa 22.77%) có giá trị trung bình. Nếu cân nhắc so với các cổ phiếu khác thì không đáng đầu tư trong giai đoạn này, vì hiệu suất không vượt trội và rủi ro không được bù đắp.
* **JPM (Sharpe Ratio = -1.3890)**: Sharpe Ratio âm lớn cho thấy hiệu suất không vượt trội. Cổ phiếu có chỉ số std thấp, tuy nhiên rủi ro lỗ trung bình và Maximum Drawdown = -0.2608 (sụt giảm tối đa 26.08%), khá cao. Bối cảnh JPM chịu áp lực từ lãi suất thấp và lo ngại suy thoái trong đại dịch, nên không đáng đầu tư trong giai đoạn này.

## XÂY DỰNG MÔ HÌNH DỰ ĐOÁN XU HƯỚNG TĂNG GIẢM CỔ PHIẾU

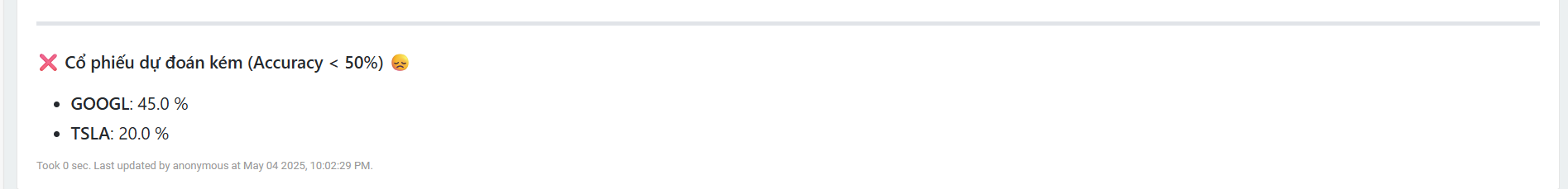
Ta có thể xây dựng một mô hình Random Forest có khả năng dự đoán xu hướng cổ phiếu dựa trên dữ liệu lịch sử từ tập dữ liệu, với các quy trình:

* Chuẩn bị dữ liệu: Là dữ liệu gốc từ file csv, thêm tỉ lệ thay đổi giá hàng ngày (daily\_return) và gán nhãn xu hướng 1 là tăng còn 0 là giảm hoặc không đổi (trend).
* Tạo thêm đặc trưng: Thêm các đặc trưng như Moving Average (ma\_5), chính là trung bình cộng giá đóng cửa (close) của 5 ngày gần nhất. Lagged price (lag\_1) chính là giá đóng cửa của ngày trước đó. RSI (Relative Strength index) để đo lường độ mạnh/yếu của giá đóng cửa trong 14 ngày gần nhất.
* Loại bỏ nhiễu: Thay thế outliers trong cột close và volume bằng trung bị của từng cột theo từng cổ phiếu.
* Chuẩn hóa dữ liệu: Đưa các đặc trưng về cùng thang đo để mô hình hoạt động hiệu quả.
* Chia tập dữ liệu thành tập huấn luyện và kiểm tra, xây dựng và huấn luyện Random forest.
* Dự đoán và đánh giá.

Sau khi dự đoán và đánh giá xong, ta có kết quả:

A screenshot of a computer

AI-generated content may be incorrect.



Để chính xác hơn, ta chỉ tiến hành dự đoán với các cổ phiếu dự đoán tốt. Có 2 chỉ số ta thêm vào, đó chính là Predicted\_Trend (Xu hướng tăng hoặc giảm của mô hình dự đoán) và Confidence (Tỉ lệ phần trăm thể hiện mức độ tin cây của mô hình đối với dự đoán, giá trị càng cao thì thể hiện mức độ đồng thuận giữa các cây trong Random forest càng cao.

Sau khi dự đoán, ta được bảng:

A screenshot of a computer

AI-generated content may be incorrect.

# PHẦN 3: KẾT LUẬN

Đề tài “Phân tích rủi ro đầu tư cổ phiếu công ty lớn” đã đạt được những thành tựu đáng kể trong việc đánh giá rủi ro tài chính khi đầu tư vào cổ phiếu của các công ty hàng đầu thuộc các lĩnh vực công nghệ, tài chính, y tế, và bán lẻ. Sử dụng bộ dữ liệu giá cổ phiếu cuối ngày, nghiên cứu đã phân tích các chỉ số như độ lệch chuẩn, giá trị rủi ro (VaR), giá trị rủi ro có điều kiện (CVaR), tỷ lệ Sharpe, và mức sụt giảm tối đa để đánh giá biến động giá và hiệu suất đầu tư. Kết quả cho thấy sự khác biệt rõ rệt giữa các cổ phiếu tăng trưởng cao với biến động lớn và các cổ phiếu phòng thủ với tính ổn định cao, cung cấp cơ sở khoa học để nhà đầu tư nhận diện rủi ro và cơ hội sinh lời. Những phân tích này hỗ trợ nhà đầu tư Việt Nam đưa ra quyết định đầu tư hiệu quả, đặc biệt trong bối cảnh hội nhập với thị trường chứng khoán quốc tế.

Về mặt học thuật, đề tài mang lại giá trị lớn trong việc củng cố kiến thức của nhóm sinh viên về khoa học dữ liệu và phân tích tài chính. Quá trình nghiên cứu đã giúp nhóm làm quen với các kỹ thuật như thống kê mô tả, phát hiện giá trị bất thường bằng phương pháp IQR, xây dựng mô hình hồi quy tuyến tính, và trực quan hóa dữ liệu qua biểu đồ nến, hộp, và mạng tương quan. Những kỹ năng này không chỉ hữu ích trong học tập mà còn là nền tảng cho các dự án phân tích dữ liệu tương lai. Đề tài mở ra cơ hội khám phá các phương pháp tiên tiến như học máy hoặc phân tích chuỗi thời gian để nâng cao độ chính xác của dự đoán tài chính. Việc áp dụng các nguyên tắc phân tích dữ liệu và diễn giải chỉ số rủi ro đã giúp nhóm phát triển tư duy logic và hiểu biết sâu sắc về thị trường chứng khoán.

Về mặt thực tiễn, nghiên cứu mang lại lợi ích thiết thực cho nhà đầu tư cá nhân và tổ chức tại Việt Nam. Các phân tích về biến động giá, khối lượng giao dịch, và chỉ số rủi ro cung cấp thông tin quan trọng để lựa chọn cổ phiếu phù hợp với mục tiêu tài chính, từ đầu tư ngắn hạn chấp nhận rủi ro cao đến đầu tư dài hạn ưu tiên an toàn. Các khuyến nghị về chiến lược đầu tư, như ưu tiên cổ phiếu tăng trưởng hoặc đa dạng hóa danh mục, giúp nhà đầu tư tối ưu hóa lợi nhuận và giảm thiểu tổn thất. Kết quả nghiên cứu cũng có thể được áp dụng trong giáo dục tài chính, hỗ trợ sinh viên, giáo viên, và cộng đồng hiểu rõ hơn về rủi ro và cơ hội trên thị trường chứng khoán quốc tế. Với tính ứng dụng cao, nghiên cứu góp phần nâng cao nhận thức tài chính và thúc đẩy sự phát triển của thị trường chứng khoán Việt Nam.

Trong tương lai, nhóm dự kiến phát triển thêm các hướng nghiên cứu để khắc phục hạn chế và nâng cao chất lượng. Trước tiên, việc tích hợp các mô hình học máy tiên tiến sẽ được ưu tiên để cải thiện khả năng dự đoán biến động giá và rủi ro. Thứ hai, nhóm sẽ mở rộng phạm vi phân tích, xem xét thêm các yếu tố kinh tế vĩ mô và sự kiện doanh nghiệp để đánh giá toàn diện hơn. Ngoài ra, việc phát triển các công cụ trực quan hóa dữ liệu tương tác sẽ tăng tính thực tiễn, giúp nhà đầu tư dễ dàng theo dõi xu hướng thị trường. Cuối cùng, nhóm dự kiến xây dựng tài liệu giáo dục dựa trên kết quả nghiên cứu, hỗ trợ các chương trình đào tạo tài chính và khoa học dữ liệu. Những cải tiến này sẽ giúp nghiên cứu đáp ứng tốt hơn nhu cầu hiện tại và trở thành nguồn tài nguyên giá trị cho cộng đồng đầu tư và học thuật.

# TÀI LIỆU THAM KHẢO

[1] Peng, R. D. (2016). *R Programming for Data Science*. Leanpub.

[2] Provost, F., & Fawcett, T. (2013). *Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking*. O'Reilly Media.