Ảnh có chứa Phông chữ, Đồ họa, biểu tượng, thiết kế

Mô tả được tạo tự động

**BÁO CÁO Data Analyst**

**Project :Phân tích giá chứng khoán**

**GVHD: Thầy Đặng Trí Thanh**

**SVTH: Phạm Trọng Nhân**

Thủ Đức, ngày 4 tháng 11 năm 2023

# **Chương 1: Giới thiệu tổng quan**

## Giới thiệu đề tài

Chứng khoán không chỉ đơn thuần là biểu đồ con số, mà còn phản ánh nhịp điệu sôi động của thị trường tài chính. Trong số các tên tuổi nổi bật, Apple Inc. (AAPL) luôn thu hút sự quan tâm lớn từ các nhà đầu tư và người quan sát thị trường. Với sự ảnh hưởng vượt trội trong ngành công nghiệp công nghệ và sản phẩm tiêu dùng điện tử, việc phân tích giá cổ phiếu của Apple đóng vai trò quan trọng trong việc đưa ra quyết định đầu tư và dự báo xu hướng thị trường.

1. Lý do chọn đề tài

Dự án này tập trung vào việc phân tích biểu đồ giá chứng khoán của Apple (AAPL), nhằm cung cấp cái nhìn sâu hơn về sự biến động của cổ phiếu này trong một khoảng thời gian cụ thể. Chúng ta sẽ xem xét các chỉ số kỹ thuật, moving average, và các yếu tố kỹ thuật khác để đánh giá xu hướng và tín hiệu tiềm năng.

Thông qua việc nghiên cứu kỹ lưỡng và phân tích chính xác, chúng ta có thể nhận ra các điểm quan trọng trong biểu đồ giá của AAPL và từ đó, đưa ra quyết định đầu tư có căn cứ.

1. Các nội dung thực hiện

Trong quá trình thực hiện đề tài, các bước chính bao gồm thu thập dữ liệu, xử lý dữ liệu, chuẩn hóa dữ liệu, trực quan hóa dữ liệu thu thập được, chia tập dữ liệu thành tập huấn luyện và tập kiểm tra, xây dựng mô hình dự đoán sử dụng thuật toán hồi quy tuyến tính, và đánh giá mô hình sử dụng các độ đo như R^2 Score, Mean Absolute Error.

Chương 2: Các công việc đã làm

1. Thu thập dữ liệu từ yahoo finance và tính min,max,Standard deviation,median cho các cột

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Phông chữ

Mô tả được tạo tự động

* Sử dụng thư viện yfinance để thu thập dữ liệu và sử dụng hàm agg() để tính các giá trị min, max,median,variancce và Standard deviation của các cột trong dữ liệu

1. Tính độ tương quan giữa các cột dữ liệu và trực quan hóa bằng heatmap

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

* Sử dụng hàm corr() để tính độ tương quan giữa các cột sau đó dùng hàm heatmap cảu thư viện seaborn để trực quan hóa dữ liệu

1. Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

   Mô tả được tạo tự độngTrực quan hóa dữ liệu bằng pair plot
2. Vẽ biểu đồ đường trực quan hóa dữ liệu cho các cột

Ảnh có chứa văn bản, ảnh chụp màn hình, phần mềm, Hệ điều hành

Mô tả được tạo tự động

1. Vẽ biểu đồ biến thiên giữa High-Low và Open-Close

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

1. Vẽ biểu đồ đường cho giá đóng cửa và biến thiên High-Low

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ

Mô tả được tạo tự động

1. Xây dựng mô hình huấn luyện sử dụng linear regression và vẽ đồ thị cho mô hình đó.

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, phần mềm

Mô tả được tạo tự động

Ảnh có chứa văn bản, ảnh chụp màn hình

Mô tả được tạo tự động

Chương 3: Kết quả

1. Heatmap of Correlation Matrix

Ảnh có chứa văn bản, ảnh chụp màn hình, số

Mô tả được tạo tự động

Các giá trị trong ma trận là các hệ số tương quan, có giá trị từ -1 đến 1. Các giá trị gần 1 cho thấy có một mối tương quan dương mạnh giữa hai biến, tức là khi một biến tăng, biến còn lại cũng tăng. Ngược lại, các giá trị gần -1 cho thấy có mối tương quan âm mạnh, tức là khi một biến tăng, biến còn lại giảm. Các giá trị gần 0 cho thấy không có mối tương quan rõ rệt giữa các biến.

Cụ thể trong ma trận trên, có những điểm đáng chú ý như:

- Các biến "Open", "Close", "High", và "Adj Close" có mối tương quan dương mạnh với nhau, gần với 1, cho thấy chúng có xu hướng di chuyển cùng nhau.

- Các biến "Low" có mối tương quan dương với "High" và "Adj Close" (gần 1), nhưng mối tương quan với "Open" và "Close" lại thấp hơn (0.98 và 0.98).

- Các biến "Volume" có mối tương quan âm với các biến còn lại, gần -0.07, cho thấy không có mối tương quan mạnh với các biến giá.

2. Biểu đồ Pair Plot

Ảnh có chứa văn bản, hàng, Song song, biểu đồ

Mô tả được tạo tự động

* Mục tiêu: Biểu đồ pair plot hiển thị các scatter plots của các cặp biến trong tập dữ liệu. Điều này giúp trực quan hóa mối tương quan giữa các biến và cho phép kiểm tra các biến có tương quan tuyến tính hay không.
* Trình bày: Pair plot tạo ra một ma trận các scatter plots. Trên đường chéo chính, nó hiển thị các biểu đồ phân phối của từng biến. Trên các ô còn lại, nó hiển thị scatter plots cho cặp biến tương ứng.

3. Biểu đồ đường trực quan hóa dữ liệu của các cột

Ảnh có chứa văn bản, biểu đồ, hàng, Sơ đồ

Mô tả được tạo tự động

4. Biểu đồ Biến thiên Open-Close và biến thiên High-Low

Ảnh có chứa văn bản, ảnh chụp màn hình, Sơ đồ, hàng

Mô tả được tạo tự động

5. Biểu đồ giá đóng cửa và High-Low

Ảnh có chứa văn bản, hàng, Sơ đồ, biểu đồ

Mô tả được tạo tự động

* Moving Average 20 (còn được gọi là MA20) là một loại đặc biệt của moving average được tính bằng cách lấy giá trị trung bình của 20 điểm dữ liệu gần nhất trong một chuỗi thời gian. Điều này có thể áp dụng cho bất kỳ loại dữ liệu nào, nhưng thường được sử dụng trong ngành tài chính, đặc biệt trong phân tích kỹ thuật và giao dịch chứng khoán.

6. Biều đồ dự đoán của mô hình huấn luyện

Ảnh có chứa văn bản, hàng, Sơ đồ, biểu đồ

Mô tả được tạo tự động

R-squared: 0.9983006598491878

Mean Absolute Error: 0.6338607764047026

* Dựa trên kết quả đánh giá
* Giá trị R^2 score trong trường hợp này là 0.9998, cho thấy mô hình giải thích được khoảng 99.83% sự biến thiên của giá đóng cửa bằng các biến được sử dụng trong mô hình. Đây là mức độ tương quan cao, mô hình tương đối tốt trong việc dự đoán giá cổ phiếu.
* Giá trị MAE là 0.6338607764047026. MAE đo lường độ lớn trung bình của sai số tuyệt đối giữa các giá trị dự đoán và giá trị thực tế. Trong trường hợp này, sai số trung bình là khoảng 0.6338, ngụ ý rằng dự đoán của mô hình có thể sai lệch trung bình là 0.6338 đơn vị so với giá thực tế.
* Tóm lại, mô hình của bạn có một mức độ dự đoán tương đối cao, được thể hiện bởi giá trị R-squared gần 1. Tuy nhiên, việc sai số trung bình (MAE) vẫn còn khá lớn, cho thấy mô hình vẫn còn phạm sai lệch đáng kể trong việc dự đoán giá cổ phiếu.
* Dựa vào đồ thị mô hình
* Dựa vào đồ thị mô hình, ta có thể thấy các điểm dữ liệu được phân tán xung quanh đường thẳng chéo từ góc dưới bên trái đến góc trên cùng bên phải. Điều này cho thấy mô hình hồi quy tuyến tính đã dự đoán tương đối chính xác các giá chứng khoán.
* Đồ thị cũng cho thấy sự tương quan tuyến tính giữa giá trị thực tế (trục x) và giá trị dự đoán (trục y). Nếu các điểm dữ liệu nằm gần đường thẳng, thì mô hình đã dự đoán chính xác các giá trị. Các điểm dữ liệu gần đường thẳng chéo cho thấy mô hình dự đoán tốt.
* Tuy nhiên, cũng có một số điểm dữ liệu có sai số lớn hơn và không nằm gần đường thẳng. Điều này có thể đại diện cho các trường hợp mà mô hình không dự đoán chính xác.
* Dựa vào đồ thị mô hình, ta có thể kết luận rằng mô hình hồi quy tuyến tính đã đạt được một mức độ dự đoán tương đối tốt với việc dự đoán giá chứng khoán, nhưng cũng cần xem xét các điểm dữ liệu có sai số lớn hơn để tăng cường hiệu suất của mô hình.