**ĐẠI HỌC QUỐC GIA HÀ NỘI**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ**

----------\*\*\*----------

**BÁO CÁO BÀI TẬP LỚN HỌC PHẦN CÁC VẤN ĐỀ HIỆN ĐẠI TRONG KHMT**



**Tên nhóm: Nhóm 10**

**Lớp: 2425II\_INT3011E\_2**

**Thành viên: Nguyễn Minh Hùng – 22028117 –** [**22028117@vnu.edu.vn**](mailto:22028117@vnu.edu.vn)

Hà Nội, 2025

**Tóm tắt (Abstract):**

Dự án này tập trung vào việc xây dựng một mô hình học máy để dự đoán giá đóng cửa của cổ phiếu Công ty Cổ phần Hoàng Anh Gia Lai (HAG) trong 3 năm tiếp theo. Dữ liệu lịch sử giá cổ phiếu (bao gồm giá mở cửa, cao nhất, thấp nhất, đóng cửa và khối lượng giao dịch) được thu thập từ file vdhd.xlsx. Các đặc trưng bổ sung như đường trung bình động giản đơn (SMA), đường trung bình động hàm mũ (EMA), chỉ số sức mạnh tương đối (RSI), phần trăm thay đổi giá và độ biến động đã được tính toán để làm giàu thông tin cho mô hình. Mô hình mạng Hồi quy Lưu trữ Dài-Ngắn hạn (LSTM) đã được lựa chọn và huấn luyện để nắm bắt các phụ thuộc tuần tự trong dữ liệu giá cổ phiếu. Sau khi huấn luyện, mô hình được sử dụng để dự đoán giá trên tập dữ liệu kiểm tra và sau đó là dự đoán cho 3 năm tới. Kết quả cho thấy tiềm năng của mô hình LSTM trong việc nắm bắt xu hướng giá, tuy nhiên, việc dự đoán dài hạn vẫn mang tính thách thức cao và cần được xem xét cẩn trọng.

**1. Giới thiệu:**

Thị trường chứng khoán luôn là một lĩnh vực hấp dẫn nhưng cũng đầy biến động, thu hút sự quan tâm của các nhà đầu tư. Việc dự đoán giá cổ phiếu là một bài toán phức tạp do bị ảnh hưởng bởi nhiều yếu tố kinh tế, chính trị, và tâm lý thị trường. Cổ phiếu HAG của Công ty Cổ phần Hoàng Anh Gia Lai là một trong những mã cổ phiếu được nhiều nhà đầu tư quan tâm tại Việt Nam.

Bài toán: Dự đoán xu hướng giá cổ phiếu HAG trong tương lai.

Mục tiêu: Xây dựng mô hình học máy có khả năng dự đoán giá đóng cửa của cổ phiếu HAG trong 3 năm tiếp theo, dựa trên dữ liệu lịch sử từ năm 2021 đến (ngày cuối cùng trong file dữ liệu của bạn).

Phạm vi:

- Sử dụng dữ liệu lịch sử giá cổ phiếu HAG (Giá mở cửa, cao nhất, thấp nhất, đóng cửa, khối lượng).

- Áp dụng các kỹ thuật tiền xử lý dữ liệu và kỹ thuật đặc trưng (feature engineering) để chuẩn bị dữ liệu đầu vào.

- Xây dựng và huấn luyện mô hình LSTM.

- Đánh giá hiệu suất mô hình và thực hiện dự đoán dài hạn.

**2. Phương pháp luận:**

**2.1. Thu thập dữ liệu:**

Dữ liệu được lấy từ file Excel có tên vdhd.xlsx, chứa thông tin giao dịch hàng ngày của cổ phiếu HAG. Các cột dữ liệu ban đầu bao gồm 'NGÀY', 'GIÁ MỞ CỬA', 'GIÁ CAO NHẤT', 'GIÁ THẤP NHẤT', 'GIÁ ĐÓNG CỬA', và 'KHỐI LƯỢNG'.

**2.2. Tiền xử lý dữ liệu:**

- Đổi tên cột: Các tên cột tiếng Việt được đổi sang tiếng Anh ('Date', 'Open', 'High', 'Low', 'Close', 'Volume') để thuận tiện cho việc xử lý.

- Chuyển đổi kiểu dữ liệu ngày: Cột 'Date' được chuyển đổi sang định dạng datetime và đặt làm chỉ số (index) cho DataFrame. Dữ liệu được sắp xếp theo thứ tự thời gian.

- Lựa chọn đặc trưng cơ bản: Các cột 'Open', 'High', 'Low', 'Close', 'Volume' được chọn làm đặc trưng đầu vào ban đầu cho mô hình.

- Xử lý giá trị thiếu: Các hàng có giá trị thiếu (NaN) được loại bỏ sau khi tính toán các chỉ báo kỹ thuật.

**2.3. Kỹ thuật đặc trưng (Feature Engineering):**

- Để cung cấp thêm thông tin cho mô hình, các chỉ báo kỹ thuật sau đã được tính toán từ giá đóng cửa ('Close'):

**SMA\_14**: Đường trung bình động giản đơn 14 ngày.

**EMA\_14**: Đường trung bình động hàm mũ 14 ngày.

**RSI**: Chỉ số sức mạnh tương đối 14 ngày.

**Price\_Change**: Phần trăm thay đổi giá so với ngày trước đó.

**Volatility**: Độ biến động, được tính bằng độ lệch chuẩn của giá đóng cửa trong 14 ngày.

**2.4. Chuẩn hóa dữ liệu:**

Toàn bộ dữ liệu đặc trưng (bao gồm giá OHLC, Volume và các chỉ báo kỹ thuật mới tạo) được chuẩn hóa về khoảng giá trị [0, 1] bằng cách sử dụng MinMaxScaler từ thư viện sklearn. Điều này giúp mô hình LSTM hội tụ nhanh hơn và ổn định hơn.

**2.5. Chuẩn bị dữ liệu chuỗi thời gian:**

Dữ liệu được chuyển đổi thành các chuỗi (sequences) để phù hợp với đầu vào của mô hình LSTM. Mỗi chuỗi đầu vào X bao gồm dữ liệu của 60 ngày liên tiếp (sequence\_length = 60), và giá trị mục tiêu y là giá đóng cửa của ngày tiếp theo.

**2.6. Phân chia dữ liệu:**

Dữ liệu chuỗi đã tạo được chia thành tập huấn luyện (80%) và tập kiểm tra (20%).

**2.7. Xây dựng mô hình LSTM:**

- Mô hình LSTM được xây dựng bằng thư viện Keras với kiến trúc tuần tự (Sequential) bao gồm các lớp sau:

Lớp LSTM thứ nhất: 64 đơn vị, return\_sequences=True (để truyền chuỗi đầu ra cho lớp LSTM tiếp theo).

Lớp Dropout: Tỷ lệ 0.2 (giúp giảm overfitting).

Lớp LSTM thứ hai: 64 đơn vị.

Lớp Dropout: Tỷ lệ 0.2.

Lớp Dense (Output): 1 đơn vị (dự đoán giá đóng cửa).

Mô hình được biên dịch với optimizer=’adam’ và hàm mất mát là ‘mean\_squared\_error’.

**2.8. Huấn luyện mô hình:**

- Mô hình được huấn luyện trên tập huấn luyện với các thông số:

epochs=100

batch\_size=32

Sử dụng EarlyStopping để theo dõi val\_loss (loss trên tập kiểm tra) và dừng huấn luyện sớm nếu không có cải thiện sau 10 epochs (patience=10), đồng thời khôi phục lại trọng số tốt nhất (restore\_best\_weights=True).

**2.9. Dự đoán và Đánh giá:**

Dự đoán trên tập kiểm tra: Mô hình được sử dụng để dự đoán giá trên tập kiểm tra. Các giá trị dự đoán (đã được chuẩn hóa) sau đó được biến đổi ngược lại về thang giá gốc.

Đánh giá: Hiệu suất mô hình trên tập kiểm tra được đánh giá bằng Root Mean Squared Error (RMSE) giữa giá thực tế và giá dự đoán.

- Dự đoán tương lai (3 năm):

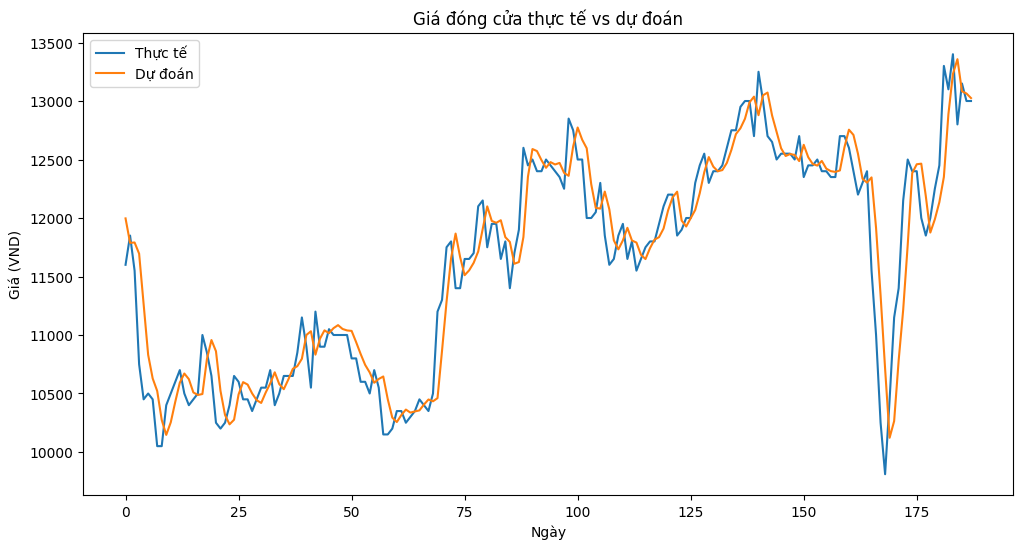
Để dự đoán cho 3 năm tới (khoảng 3 \* 252 ngày giao dịch), mô hình lấy 60 ngày dữ liệu cuối cùng từ tập dữ liệu gốc (đã chuẩn hóa) làm đầu vào ban đầu.

Sau đó, mô hình dự đoán giá cho ngày tiếp theo. Giá trị dự đoán này được thêm vào cuối chuỗi đầu vào, và giá trị cũ nhất trong chuỗi được loại bỏ để duy trì độ dài chuỗi là 60.

Quá trình này được lặp lại cho đến khi đủ số ngày dự đoán.

Các giá trị dự đoán cuối cùng cũng được biến đổi ngược lại về thang giá gốc.

**3. Kết quả và Thảo luận:**



**3.1. Hiệu suất mô hình trên tập kiểm tra:**

Mô hình LSTM đã được huấn luyện và đánh giá trên tập dữ liệu kiểm tra.

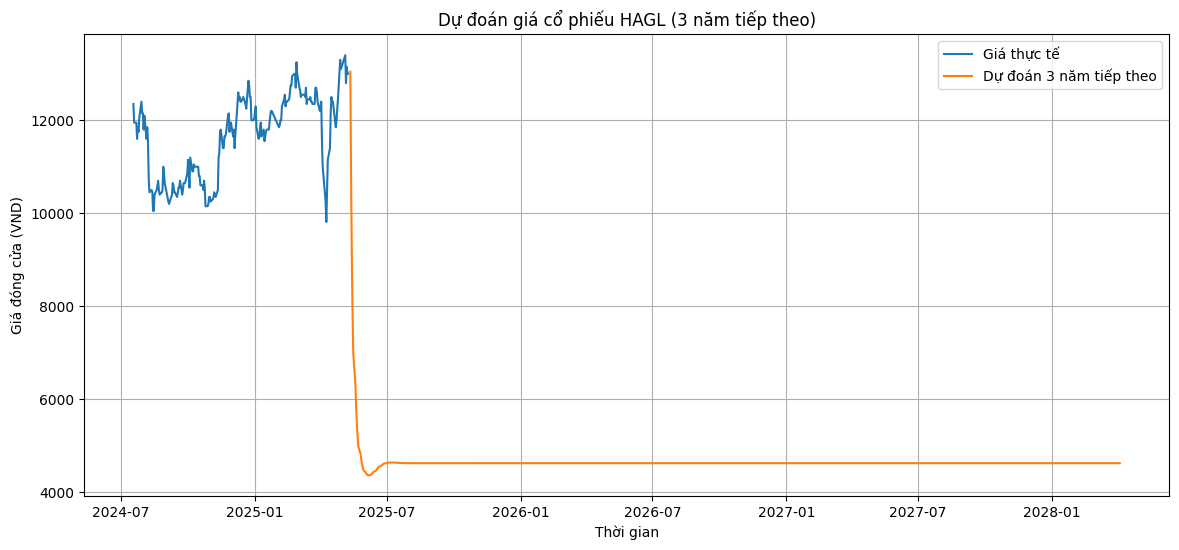
RMSE: (Điền giá trị RMSE bạn nhận được ở đây). Giá trị RMSE cho biết mức độ sai lệch trung bình giữa giá dự đoán và giá thực tế.

Trực quan hóa: (Mô tả biểu đồ so sánh giá thực tế và giá dự đoán trên tập kiểm tra). Biểu đồ này giúp đánh giá trực quan khả năng của mô hình trong việc bắt kịp xu hướng và biến động của giá cổ phiếu.

3.2. Dự đoán cho 3 năm tới:

Mô hình đã được sử dụng để tạo ra dự đoán giá cổ phiếu HAG cho 3 năm tiếp theo.

Trực quan hóa:



**3.3. Thảo luận:**

Điểm mạnh: Mô hình LSTM với các đặc trưng kỹ thuật cho thấy khả năng học hỏi các mẫu trong dữ liệu giá cổ phiếu lịch sử. Việc sử dụng MinMaxScaler và Dropout là các thực hành tốt trong việc xây dựng mô hình mạng neuron.

Hạn chế và Thách thức:

Độ chính xác của dự đoán dài hạn: Dự đoán giá cổ phiếu, đặc biệt là trong khoảng thời gian dài như 3 năm, là một nhiệm vụ cực kỳ khó khăn. Thị trường chịu ảnh hưởng của nhiều yếu tố phức tạp và khó lường (kinh tế, chính trị, sự kiện bất ngờ, tâm lý nhà đầu tư) mà mô hình dựa trên dữ liệu lịch sử khó có thể nắm bắt hết được.

Lỗi tích lũy: Trong quá trình dự đoán lặp đi lặp lại cho tương lai, sai số ở mỗi bước dự đoán có thể tích lũy, làm giảm độ chính xác của các dự đoán xa hơn.

Tính không dừng của dữ liệu tài chính: Giá cổ phiếu thường không ổn định theo thời gian, các mô hình có thể gặp khó khăn khi các mẫu hình thị trường thay đổi.

Phụ thuộc vào dữ liệu lịch sử: Mô hình chỉ học từ những gì đã xảy ra trong quá khứ và không thể dự đoán các "thiên nga đen" hoặc các thay đổi cấu trúc thị trường đột ngột.

Phạm vi dữ liệu huấn luyện: Hiệu suất của mô hình phụ thuộc rất nhiều vào chất lượng và phạm vi của dữ liệu huấn luyện. Nếu dữ liệu huấn luyện (trong trường hợp này là "2021-2025") không bao gồm đủ các trạng thái thị trường khác nhau, mô hình có thể hoạt động không tốt trong các điều kiện thị trường mới. Cần làm rõ ngày kết thúc chính xác của dữ liệu vdhd.xlsx để hiểu rõ hơn về khoảng thời gian dự đoán.

**4. Kết luận và Hướng phát triển tương lai:**

**4.1. Kết luận:**

Dự án đã thành công trong việc triển khai một mô hình LSTM để dự đoán giá cổ phiếu HAG, sử dụng dữ liệu lịch sử và các chỉ báo kỹ thuật. Mô hình cho thấy khả năng dự đoán trên tập kiểm tra (cần nêu rõ mức độ dựa trên RMSE). Tuy nhiên, các kết quả dự đoán dài hạn (3 năm) cần được xem xét với sự thận trọng cao do những thách thức và hạn chế cố hữu của việc dự đoán thị trường tài chính trong thời gian dài.

**4.2. Hướng phát triển tương lai:**

Thử nghiệm thêm đặc trưng: Bổ sung các đặc trưng khác như chỉ báo khối lượng (ví dụ: OBV - On-Balance Volume), các chỉ báo kinh tế vĩ mô, hoặc phân tích cảm tính từ tin tức liên quan đến HAG.

Tinh chỉnh siêu tham số: Sử dụng các kỹ thuật như Grid Search hoặc Random Search để tìm bộ siêu tham số tối ưu cho mô hình LSTM (ví dụ: số lớp, số units mỗi lớp, learning rate).

So sánh với các mô hình khác: Thử nghiệm các mô hình dự đoán chuỗi thời gian khác như ARIMA, Prophet, hoặc các kiến trúc mạng neuron khác (ví dụ: GRU, Transformer) để so sánh hiệu suất.

Đánh giá chi tiết hơn: Sử dụng thêm các thước đo đánh giá như MAE, MAPE, và độ chính xác về hướng (directional accuracy).

Xem xét khoảng thời gian dự đoán ngắn hơn: Tập trung vào dự đoán ngắn hạn (ví dụ: vài ngày hoặc vài tuần) có thể mang lại kết quả đáng tin cậy hơn.

Phân tích độ nhạy của mô hình: Kiểm tra xem mô hình hoạt động như thế nào với các cửa sổ thời gian đầu vào (sequence length) khác nhau.

Kết hợp dữ liệu bên ngoài: Nếu có thể, tích hợp dữ liệu từ các nguồn bên ngoài (ví dụ: báo cáo tài chính của công ty, tin tức thị trường) có thể cải thiện độ chính xác của mô hình.

**Tài liệu tham khảo:**

1. **Python LSTM (Long Short-Term Memory Network) for Stock Predictions – DataCamp** - Một hướng dẫn chi tiết về cách sử dụng LSTM bằng Python để dự đoán giá cổ phiếu, bao gồm các bước tiền xử lý dữ liệu, xây dựng mô hình và trực quan hóa kết quả.

<https://www.datacamp.com/tutorial/lstm-python-stock-market>

2. **Stock Market Prediction using Machine Learning in 2025 -Simplilearn.com**   - Bài viết này cung cấp cái nhìn tổng quan về dự đoán giá cổ phiếu bằng học máy và đi sâu vào việc sử dụng LSTM, bao gồm các bước từ thu thập dữ liệu đến kiểm tra mô hình.

<https://www.simplilearn.com/tutorials/machine-learning-tutorial/stock-price-prediction-using-machine-learning>

3. **Stock Price Prediction Project using TensorFlow – GeeksforGeeks** -Một hướng dẫn từng bước về cách xây dựng mô hình dự đoán giá cổ phiếu bằng TensorFlow và LSTM, bao gồm cả code mẫu.

<https://www.geeksforgeeks.org/stock-price-prediction-project-using-tensorflow/>

4. **Stock Price Prediction using LSTM and its Implementation - Analytics Vidhya** - Giải thích về kiến trúc LSTM và cách nó giải quyết các thách thức trong dự đoán chuỗi thời gian, cùng với việc triển khai thực tế trên dữ liệu chứng khoán.

<https://www.analyticsvidhya.com/blog/2021/12/stock-price-prediction-using-lstm/>