


THÔNG TIN CHUNG CỦA BÁO CÁO

- Link YouTube video của báo cáo (tối đa 5 phút):
<https://youtu.be/RcQYk8fC1FU>
- Link slides (dạng .pdf đặt trên Github):
<https://github.com/leirisyue/CS2205.CH181.git>
- Mỗi thành viên của nhóm điền thông tin vào một dòng theo mẫu bên dưới
- Sau đó điền vào Đề cương nghiên cứu (tối đa 5 trang), rồi chọn Turn in

<ul style="list-style-type: none">● Họ và Tên: Lê Nguyễn Thanh Thùy● MSSV: 230201055 	<ul style="list-style-type: none">● Lớp: CS2205.CH181● Tự đánh giá (điểm tổng kết môn): 8.5/10● Số buổi vắng: 0● Số câu hỏi QT cá nhân:● Link Github: https://github.com/leirisyue/CS2205.CH181
--	---

ĐỀ CƯƠNG NGHIÊN CỨU

TÊN ĐỀ TÀI (IN HOA)

ỨNG DỤNG MÔ HÌNH LSTM DỰ BÁO CHỈ SỐ VIETNAM INDEX 30

TÊN ĐỀ TÀI TIẾNG ANH (IN HOA)

APPLY LSTM MODEL FOR FORECASTING VIETNAM INDEX 30

TÓM TẮT *(Tối đa 400 từ)*

Ngày nay, Thị trường chứng khoán là kênh đầu tư hấp dẫn đối với nhiều quỹ đầu trong và ngoài nước. Trong bối cảnh kinh tế thị trường ngày càng biến động và phức tạp, việc dự báo chính xác xu hướng của các chỉ số chứng khoán đang là bài toán vô cùng quan trọng đối với các nhà đầu tư vừa và nhỏ. Chỉ số VietNam Index 30 (hay còn gọi VN30), bao gồm 30 cổ phiếu hàng đầu trên sàn HOSE, không chỉ phản ánh tình hình hoạt động của các doanh nghiệp lớn mà còn phản ánh hưởng mạnh mẽ đến quyết định khả năng đầu tư của nhiều cá nhân và tổ chức. Nhằm nâng cao độ chính xác cho mô hình dự báo các chỉ số chứng khoán, nhiều nhà nghiên cứu cũng như các chuyên gia tài chính đã áp dụng các phương pháp tiên tiến, trong đó được kể đến mô hình Long Short-Term Memory (LSTM) nổi lên như một công cụ mạnh mẽ và hiệu quả. LSTM, một dạng cải tiến của mạng nơ-ron hồi quy (RNN), có khả năng ghi nhớ và xử lý các phụ thuộc dài hạn trong dữ liệu chuỗi thời gian. Ưu điểm này đặc biệt hữu ích trong việc phân tích và dự báo các biến động thị trường của VN30. Bởi lẽ, chúng không chỉ phụ thuộc vào các yếu tố ngắn hạn mà còn liên quan trực tiếp đến các sự kiện kinh tế vĩ mô đã diễn ra trong quá khứ. Trong nghiên cứu này tập trung vào việc ứng dụng mô hình LSTM trong dự báo chỉ số VietNam Index 30, từ đó khám phá các bước triển khai cũng như quá trình đánh giá hiệu quả của mô hình, và cuối cùng cung cấp cái nhìn toàn diện về tiềm năng của LSTM trong lĩnh vực dự báo tài chính.

GIỚI THIỆU *(Tối đa 1 trang A4)*

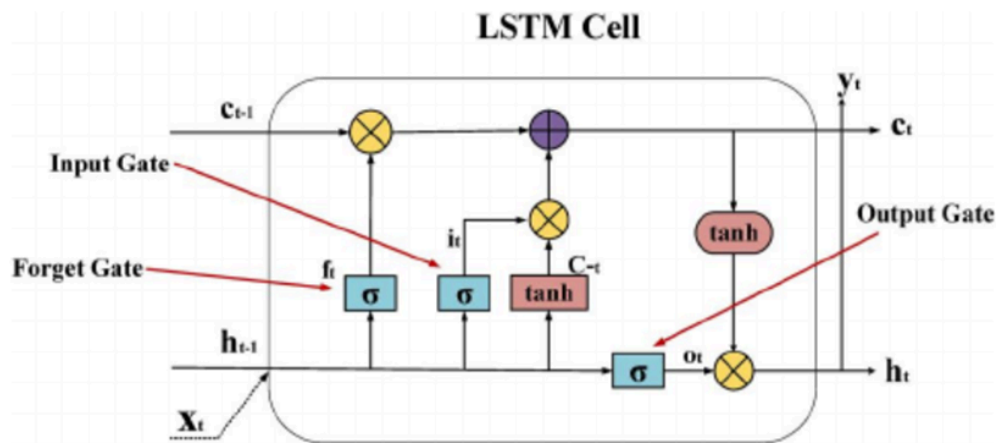
Mô hình Long Short-Term Memory (LSTM) được ứng dụng trong nghiên cứu này

nhằm dự báo ngắn hạn chỉ số VietNam Index 30, một chỉ số tài chính quan trọng phản ánh tình hình kinh tế vĩ mô và tài chính của Việt Nam. LSTM là một biến thể tiên tiến của mạng nơ-ron hồi quy (Recurrent Neural Networks - RNNs), được đề xuất nhằm khắc phục các hạn chế của RNN trong việc xử lý các phụ thuộc dài hạn trong dữ liệu chuỗi thời gian.

Điểm nổi bật của LSTM so với các kiến trúc RNN truyền thống là khả năng trong việc lưu trữ thông tin quan trọng trong bộ nhớ nội bộ mà không bị mất mát qua thời gian, điều này có ý nghĩa quan trọng trong các ứng dụng dự báo. Cấu trúc của LSTM bao gồm các thành phần chính như một cổng đầu vào (input gate), một cổng quên (forget gate) và một cổng đầu ra (output gate), cùng với một trạng thái nội bộ (cell state) giữ vai trò như một bộ nhớ dài hạn, cho phép nó lưu giữ thông tin qua nhiều bước thời gian được thể hiện ở [hình 1].

Cổng đầu vào (input gate) quyết định thông tin nào sẽ được lưu trữ trong trạng thái cục bộ, cổng quên (forget gate) xác định thông tin nào sẽ bị loại bỏ khỏi trạng thái nội bộ, và cổng đầu ra (output gate) điều khiển thông tin từ trạng thái nội bộ đến đầu ra của mạng nơ-ron. Sự kết hợp của các cổng này cho phép mô hình LSTM hạn chế vấn đề biến mất gradient, thường gặp trong các mạng RNN thông thường, và duy trì khả năng học tập ổn định qua các chuỗi dài [1].

Nghiên cứu sử dụng mô hình LSTM này để phân tích và dự báo các biến động của chỉ số VietNam Index 30, với kỳ vọng kết quả của nghiên cứu sẽ cung cấp một công cụ hữu ích cho các nhà đầu tư cá nhân hoặc tổ chức tiếp cận rõ hơn về xu hướng thị trường trong ngắn hạn. Mô hình này không chỉ giúp dự báo các xu hướng (up trend hoặc down trend) mà còn có thể dự đoán các biến động tiềm ẩn của tình hình kinh tế.



Hình 1: Long short-term memory cell structure [2]

Input:

- Nguồn dữ liệu: Bao gồm dữ liệu thị trường chứng khoán do các website cộng đồng cung cấp (như giá cổ phiếu, tin tức tài chính, khối lượng giao dịch,...), dữ liệu lịch sử giao dịch của theo thời gian, và dữ liệu biểu đồ theo thời gian của các chỉ số tài chính từ các nền tảng chính thống như <https://www.blockchain.com>

Output:

- Đánh giá và dự báo của chỉ số VN30 cho các ngày, tuần, hoặc tháng trong tương lai dựa trên các dữ liệu có sẵn lịch sử.

MỤC TIÊU

(Viết trong vòng 3 mục tiêu, lưu ý về tính khả thi và có thể đánh giá được)

- Trình bày biểu đồ so sánh giữa giá trị tài chính trên thực tế của chỉ số VietNam Index 30 và giá trị dự báo mà mô hình LSTM tính toán để đánh giá độ tin cậy của mô hình.
- Phân tích chi tiết các vấn đề về hiệu suất của mô hình trong các giai đoạn khác nhau trong quá trình thực hiện.
- Cung cấp kênh thông tin giúp các nhà đầu tư hiểu rõ hơn về khả năng dự báo của mô hình LSTM đối với chỉ số VietNam index 30, từ đó đưa ra các quyết định danh mục đầu tư. Bên cạnh đó có thể làm cơ sở để các nhà nghiên cứu mở rộng mô hình hoặc kết hợp với những mô hình hiện đại hơn để độ chính xác.

PHẠM VI

Phạm vi nghiên cứu này sử dụng các dữ liệu tài chính công khai được cung cấp từ các nguồn uy tín kết hợp các chỉ số đặc thù phân tích tài chính để tạo mô hình LSTM có khả năng dự báo chỉ số thị trường VietNam Index 30 dựa vào các sự kiện lịch sử có sẵn.

NỘI DUNG VÀ PHƯƠNG PHÁP

(Viết nội dung và phương pháp thực hiện để đạt được các mục tiêu đã nêu)

NỘI DUNG :

- Chuẩn hóa dữ liệu
 - Thu thập dữ liệu
 - Tiền xử lý dữ liệu
 - Tạo các chuỗi thời gian
- Xây dựng và huấn luyện mô hình LSTM
 - Train model, validation model và test model.
 - Đánh giá mô hình
- Tính khả thi của mô hình dự báo và trực quan hóa
 - Sử dụng mô hình đã huấn luyện để dự báo chỉ số VN30
 - Trực quan hóa kết quả dự báo so với dữ liệu thực tế.
- Nâng cao hiệu năng sử dụng hệ thống

PHƯƠNG PHÁP:

1. Phương pháp thực Chuẩn hóa dữ liệu

- Phương pháp thu thập số liệu: Sử dụng API từ các sàn giao dịch chứng khoán hoặc các trang tin tức tài chính, và các công cụ phân tích mạng xã hội. Cụ thể, các giá trị cổ phiếu được lấy từ Yahoo Finance, đây là thư viện do python cung cấp, và được sử dụng rộng rãi trong nghiên cứu dự báo giá cổ phiếu.
- Phương pháp Min-Max Scaling (Điều chỉnh tỷ lệ): đây một phương pháp chuẩn hóa dữ liệu trong quá trình xây dựng mô hình máy học. Mục tiêu của việc sử dụng phương pháp này là đưa dữ liệu về một khoảng cụ thể, thường là $[0, 1]$. Dưới đây là cách thức thực hiện Min-Max Scaling

- Phương pháp Sliding Window: Một cửa sổ trượt kích thước 7 ngày được sử dụng để tạo ra các tập dữ liệu đầu vào cho mô hình. Ý nghĩa là mô hình sử dụng dữ liệu của 7 ngày trước để dự đoán giá cổ phiếu của 7 ngày tiếp theo. Kích thước cửa sổ trượt này giúp mô hình có thể học các mẫu thời gian ngắn hạn trong dữ liệu [3].

2. Xây dựng mô hình LSTM:

- Phương pháp phân tách dữ liệu: Dữ liệu được chia thành hai tập: 80% dữ liệu dùng để huấn luyện mô hình và 20% còn lại dùng để kiểm tra mô hình.

- Sử dụng thư viện TensorFlow kết hợp Keras để xây dựng mô hình LSTM. Cấu trúc cơ bản bao gồm các lớp LSTM và Dense.

- Phương pháp phân tích kỹ thuật: Bên cạnh đó kết hợp với khái niệm phổ biến trong phân tích kỹ thuật, đặc biệt trong lĩnh vực đầu tư và giao dịch chứng khoán: đường trung bình đơn giản (SMA), đường trung bình hội tụ phân kỳ (MACD), và chỉ số sức mạnh tương đối (RSI) [2]

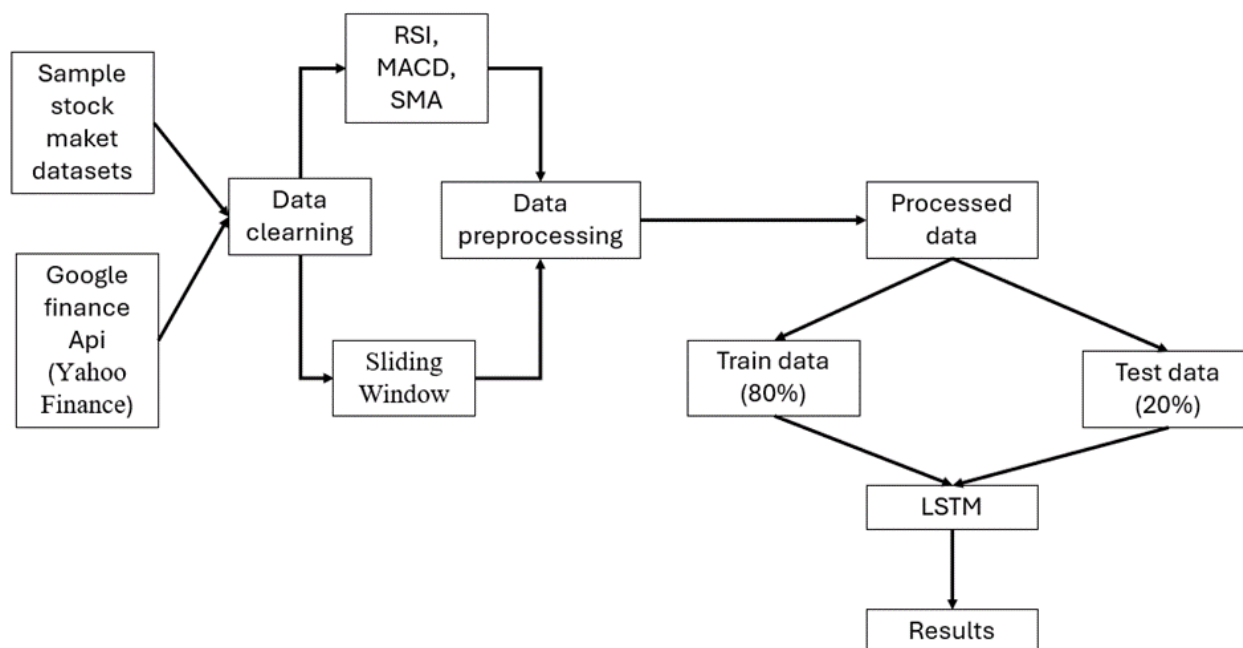
3. Phương pháp thực nghiệm, so sánh, phân tích đánh giá:

- Phương pháp so sánh các mô hình với số lớp LSTM khác nhau: Các mô hình với 2, 3, và 4 lớp LSTM. Có nhiều chỉ số dùng để đo lường, đánh giá và kiểm định chất lượng của một mô hình. Trong nghiên cứu này sử dụng MAPE (Mean absolute percentage error) hay phần trăm sai số trung bình tuyệt đối của mô hình dự báo. MSE (Mean squared error) là trung bình bình phương của các sai số của mô hình. Và cuối cùng là RMSE (Root Mean Square Error) nhằm đánh chất lượng mô hình.

- Phương pháp so sánh hiệu quả trên các cổ phiếu khác nhau: Hiệu quả của mô hình được đánh giá trên ba cổ phiếu khác nhau (có thể tùy chọn 3 cổ phiếu quan tâm trong danh mục đầu tư cá nhân) để kiểm tra tính tổng quát của mô hình.

- Sử dụng mô hình đã huấn luyện để dự báo chỉ số VN30.

- Trực quan hóa kết quả dự báo so với dữ liệu thực tế



Hình 2: lưu đồ các bước thực hiện

KẾT QUẢ MONG ĐỢI

(Viết kết quả phù hợp với mục tiêu đặt ra, trên cơ sở nội dung nghiên cứu ở trên)

- Chứng minh tính khả thi của việc sử dụng mô hình LSTM trong dự báo chỉ số VN30, góp phần hỗ trợ quyết định đầu tư và quản lý rủi ro cho các nhà đầu tư.
- Đưa ra các dự báo đáng tin cậy có thể sử dụng để lập kế hoạch đầu tư ngắn hạn và dài hạn.
- Thông qua nghiên cứu này có thể mở rộng và phát triển mô hình vào việc ứng dụng thực tế, từ đó tạo ra giá trị thực tiễn cho lĩnh vực tài chính và đầu tư.

TÀI LIỆU THAM KHẢO (Định dạng DBLP)

[1] Aadhitya A, Rajapriya R, Vineetha R S, Anurag M Bagde. "Predicting Stock Market time-series data using CNN-LSTM Neural Network model".

arXiv:2305:14378v1, 2023.

[2] Tran Phuoc, Pham Thi Kim Anh, Phan Huy Tam, Chien V. Nguyen. "Applying machine learning algorithms to predict the stock price trend in the stock market – The case of Vietnam". Humanities And Social Sciences Communications, 2024: 11, 393.

[3] Sakshi Kulshreshtha, Vijayalakshmi A. “An ARIMA- LSTM Hybrid Model for Stock Market Prediction Using Live Data”. School of Computer Science and Engineering, 2020: 13(4), 117 – 123.