

REPORT FINPROS TEST

Chủ đề: xây dựng mô hình dự đoán biến động của cổ phiếu sau N ngày

Thực hiện: **NGUYỄN CAO QUỐC**

Vị trí: **AI Engineer Intern – Khu vực Hồ Chí Minh**

Email: caoquoc.job@gmail.com

MỤC LỤC

1.	GIỚI THIỆU:.....	2
2.	HƯỚNG GIẢI QUYẾT:	2
2.1.	GIỚI THIỆU DỮ LIỆU:	2
2.2.	XỬ LÝ DỮ LIỆU:	2
2.3.	GIỚI THIỆU DEEP LEARNING VÀ LSTM:	3
2.3.1.	GIỚI THIỆU DEEP LEARNING:.....	3
2.3.2.	GIỚI THIỆU LSTM:	3
3.	ÁP DỤNG:.....	4
4.	LỜI KẾT:.....	5

1. GIỚI THIỆU:

Trong thời kì phát triển của Trí Tuệ Nhân Tạo (AI) thì việc tận dụng các kỹ thuật và phương pháp của công nghệ này cho các bài toán dự đoán đang trở nên phổ biến hơn, và dự đoán giá trị của cổ phiếu cũng không phải ngoại lệ. Trong phần báo cáo này, em sẽ sử dụng các công nghệ của AI để giải quyết cho bài toán dự đoán sự biến động của cổ phiếu, bao gồm cả phân tích dữ liệu và vận dụng thuật toán.

2. HƯỚNG GIẢI QUYẾT:

2.1. GIỚI THIỆU DỮ LIỆU:

- Bộ dữ liệu được cung cấp bởi công ty cổ phần đầu tư FINPROS.
- Bộ dữ liệu bao gồm bốn tập về bốn mã chứng khoán khác nhau:
 - Công ty cổ phần FPT (mã FPT)
 - Công ty cổ phần tập đoàn Masan (MSN)
 - Công ty cổ phần vàng bạc đá quý Phú Nhuận (PNJ)
 - Tập đoàn Vingroup (VIC).
- Mỗi tập dữ liệu sẽ là các chỉ số chứng khoán của công ty từ năm 2017 tới năm 2020, bao gồm:
 - Ticker: Mã cổ phiếu của công ty
 - Date/Time: Ngày và giờ cụ thể
 - Open: Giá mở cửa của cổ phiếu
 - High: Giá trần trong ngày
 - Low: Giá sàn trong ngày
 - Close: Giá đóng cửa phiên giao dịch
 - Volume: Khối lượng cổ phiếu giao dịch tại thời điểm “Date/Time”.

2.2. XỬ LÝ DỮ LIỆU:

Để xử lý các tập dữ liệu trên, em sẽ đi qua từng bước cụ thể:

1. Do là bốn tập dữ liệu riêng lẻ nên em sử dụng hàm `concat()` trong thư viện `pandas` để nối tất cả lại thành một bộ dữ liệu chung.

2. Tách cột đặc trưng “Date/Time” thành bốn cột đặc trưng khác nhau bao gồm “Year”, “Month”, “Day” và “Time” để sử dụng cho phân tích.
3. Bỏ đi hai cột đặc trưng là “Date/Time” (do đã tách thành 4 cột riêng lẻ) và “Open Interest” (toàn bộ là giá trị 0, không ảnh hưởng lên bộ dữ liệu).
4. Sắp xếp lại các dữ liệu theo thứ tự tăng dần của cột đặc trưng “Year” và “Month”, mục đích là để thuận tiện cho việc trực quan hóa sau này.
5. Ánh xạ các dữ liệu tại cột “Ticker” từ các mã cổ phiếu có kiểu là string sang dạng int từ 0 tới 3 (mục đích là để huấn luyện mô hình).
6. Kiểm tra xem có dữ liệu bị thiếu hoặc sai sót hay không bằng hàm isnull(), nếu có thì sẽ bổ sung. Trong trường hợp này là không.

2.3. GIỚI THIỆU DEEP LEARNING VÀ LSTM:

2.3.1. GIỚI THIỆU DEEP LEARNING:

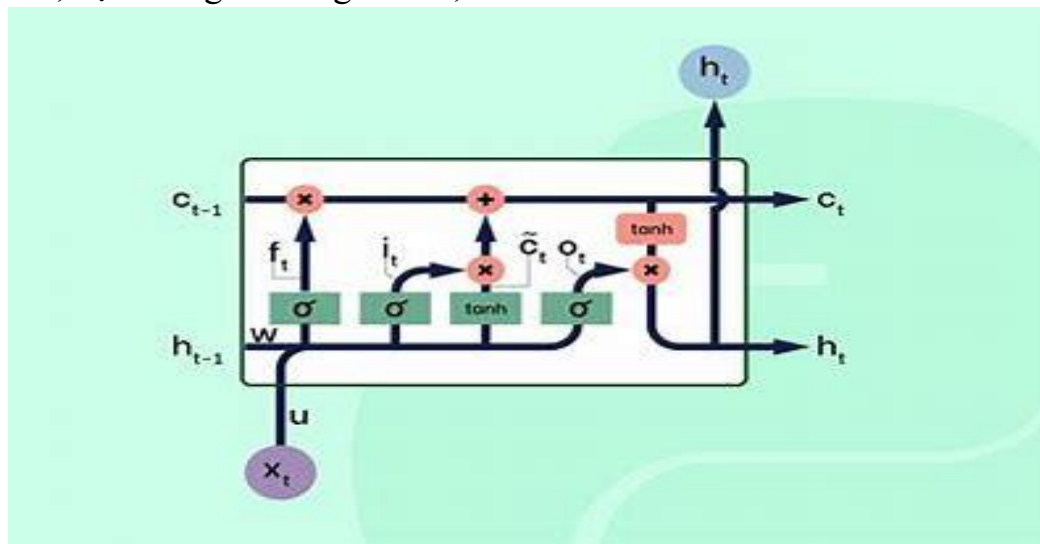
- Deep learning là một hướng đi trong lĩnh vực AI. Dựa trên mạng neural nhân tạo, deep learning có thể biểu diễn đặc trưng từ dữ liệu lớn và phức tạp.
- Deep learning được ứng dụng rộng rãi trong nhiều lĩnh vực như Xử lý ngôn ngữ tự nhiên, Thị giác máy tính, Dịch máy,...
- Deep learning trong chứng khoán: thị trường chứng khoán là nơi mọi người đều mong muốn tìm ra các “câu thần chú” để dự đoán thành công giá cổ phiếu. Mặc dù vậy, đây là một vấn đề gần như không thể dự đoán chính xác 100%.
- Có rất nhiều mô hình giúp dự đoán giá cổ phiếu như là SARIMA, Facebook Prophet, HoltWinters,...trong báo cáo lần này, em sẽ sử dụng LSTM.

2.3.2. GIỚI THIỆU LSTM:

- LSTM là viết tắt của Long Short-Term Memory. LSTM có cấu trúc gồm các module, mỗi module bao gồm 3 cổng: cổng quên, cổng đầu vào, cổng đầu ra. Các cổng này quyết định thông tin nào sẽ được lưu trữ, xóa, chỉnh sửa và truyền đi trong mạng.
- Cổng quên (forget gate): Cổng này quyết định thông tin nào sẽ được bỏ đi khỏi trạng thái ô (cell state) hiện tại. Nó sử dụng một hàm

sigmoid để đánh giá mức độ quan trọng của mỗi phần tử trong trạng thái ô, và chỉ giữ lại những phần tử có giá trị cao.

- Cổng đầu vào (input gate): Cổng này cho phép thông tin tùy chọn cần thiết từ trạng thái ô hiện tại. Nó quyết định thông tin nào có liên quan đến đầu vào hiện tại và cho phép thông tin đó vào. Nó cũng sử dụng một hàm sigmoid để lọc ra những thông tin quan trọng.
- Cổng đầu ra (output gate): Cổng này cập nhật và hoàn thiện trạng thái ẩn (hidden state) tiếp theo. Nó sử dụng một hàm sigmoid để chọn ra những thông tin quan trọng từ trạng thái ô, và một hàm tanh để tạo ra một vector đầu ra. Vector đầu ra này sẽ được nhân với kết quả của hàm sigmoid để tạo ra trạng thái ẩn tiếp theo.
- LSTM có nhiều ứng dụng trong các lĩnh vực liên quan đến dữ liệu chuỗi, như xử lý ngôn ngữ tự nhiên, nhận dạng giọng nói, dự báo thời tiết, dự đoán giá chứng khoán,...



Hình 1: Cấu trúc của LSTM

3. ỨNG DỤNG:

Để áp dụng LSTM vào trong phân tích chứng khoán, em đã xây dựng một mô hình deep learning được tạo ra từ 3 lớp LSTM có cấu trúc như sau:

- Lớp LSTM đầu tiên: số neural là 7 do đầu vào của em có 7 đặc trưng, hàm kích hoạt là ReLU và input_shape là (1, 7).
- Sau lớp đầu tiên em có thêm một lớp Dropout nhằm tránh cho mô hình bị Overfitting. Tỷ lệ là 0.3.
- Lớp LSTM thứ hai: số neural là 64, hàm kích hoạt là ReLU.

- Cũng như sau lớp đầu tiên, em cho thêm một lớp Dropout với tỉ lệ là 0.3.
- Lớp LSTM thứ ba: số neural là 128, hàm kích hoạt là ReLU.
- Lớp Dropout với tỉ lệ là 0.3
- Lớp Dense: đây là lớp cuối cùng, số unit là 1 do bài toán chỉ có 1 kết quả trả về.

Em sử dụng 4 độ đo gồm Mean Squared Error, Root Mean Squared Error, Mean Absolute Error và R-squared để đánh giá mô hình. Kết quả thu được như hình dưới.

	Mean Squared Error	Root Mean Squared Error	Mean Absolute Error	R-squared (R2) Score
0	0.008382	0.091553	0.087812	0.854128

Hình 2: Kết quả đánh giá thu được

Sau khi huấn luyện xong mô hình, em có sử dụng streamlit để xây một demo đơn giản. Anh chị có thể truy cập và sử dụng demo theo đường link dưới đây.

Link: <https://stock-predictionn.streamlit.app>

Hình 3: giao diện cơ bản của demo

4. LỜI KẾT:

Trong quá trình thực hiện bài test này, em có gặp một số khó khăn về mặt chọn lựa các đặc trưng cho việc huấn luyện, lí do là vì kiến thức về chứng khoán của em cũng không phải dạng giáo sư. Cho nên là việc đưa ra quyết định chọn đặc trưng nào cho việc huấn luyện cũng đều dựa trên kinh nghiệm và trải nghiệm của em khi em tham gia vào thị trường chứng khoán. Em rất mong muốn được cải thiện thêm những kinh nghiệm trên và học hỏi thêm những kiến thức khác không chỉ trong IT mà còn là trong thị trường chứng khoán. Em cảm ơn anh chị bên công ty FINPROS đã đọc báo cáo này.