

DỰ ĐOÁN GIÁ CỔ PHIẾU BA NGÂN HÀNG VIỆT NAM - BIDV, VIETCOMBANK, EXIMBANK

TRẦN HOÀNG PHÚC¹, NGUYỄN VIỆT HOÀNG², LÊ BÁ NHẬT LONG³, NGUYỄN HÙNG TUẤN⁴,
và LÊ ANH DUY⁵

¹Khoa Hệ thống thông tin, Trường Đại học Công nghệ Thông tin - UIT, (e-mail: 21522479@gm.uit.edu.vn)

²Khoa Hệ thống thông tin, Trường Đại học Công nghệ Thông tin - UIT, (e-mail: 21522095@gm.uit.edu.vn)

³Khoa Hệ thống thông tin, Trường Đại học Công nghệ Thông tin - UIT, (e-mail: 21522300@gm.uit.edu.vn)

⁴Khoa Hệ thống thông tin, Trường Đại học Công nghệ Thông tin - UIT, (e-mail: 21521633@gm.uit.edu.vn)

⁵Khoa Hệ thống thông tin, Trường Đại học Công nghệ Thông tin - UIT, (e-mail: 21521994@gm.uit.edu.vn)

TÓM TẮT Abstract ở đây

INDEX TERMS Keywords ở đây

I. GIỚI THIỆU

Nội dung chương 1

II. CÁC NGHIÊN CỨU LIÊN QUAN

Linear Regression: nội dung.

GRU: nội dung.

LSTM: nội dung.

ARIMA: nội dung.

Meta-Learning: nội dung.

N-HiTS: nội dung.

III. TÀI NGUYÊN

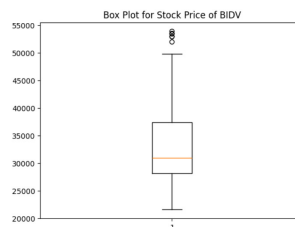
A. NGUỒN DỮ LIỆU

Tập dữ liệu được sử dụng trong nghiên cứu này được lấy từ <https://finance.yahoo.com/>, một nền tảng tài chính uy tín được biết đến với thông tin thị trường toàn diện và cập nhật. Tập dữ liệu bao gồm dữ liệu giá cổ phiếu của ba ngân hàng nổi tiếng tại Việt Nam: Ngân hàng Đầu tư và Phát triển Việt Nam (BIDV), Ngân hàng Thương mại Cổ phần Xuất Nhập Khẩu Việt Nam (Eximbank) và Ngân hàng Thương mại Cổ phần Ngoại Thương Việt Nam (Vietcombank). Bằng cách tận dụng dữ liệu có sẵn trên nền tảng này, chúng tôi đảm bảo một nền tảng đáng tin cậy cho phân tích của mình. Tập dữ liệu bắt đầu từ ngày 1 tháng 3 năm 2019 đến ngày 1 tháng 3 năm 2024, cung cấp một phạm vi thời gian mạnh mẽ cho phân tích của chúng tôi. Mỗi mục trong tập dữ liệu bao gồm các chỉ số tài chính chính, bao gồm:

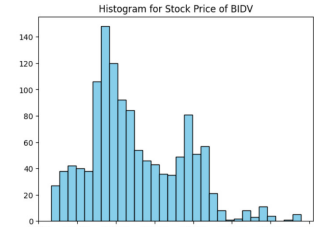
- Date:** Ngày giao dịch.
- Open:** Giá cổ phiếu mở cửa vào đầu ngày giao dịch.

- High:** Giá cổ phiếu cao nhất ghi nhận được trong ngày giao dịch.
- Low:** Giá cổ phiếu thấp nhất ghi nhận được trong ngày giao dịch.
- Close:** Giá cổ phiếu đóng cửa vào một ngày nhất định.
- Adj Close:** Giá đóng cửa được điều chỉnh, bao gồm bất kỳ hành động doanh nghiệp nào như cổ tức hoặc chia cổ phiếu.
- Volume:** Khối lượng giao dịch, chỉ ra tổng số cổ phiếu được giao dịch vào một ngày cụ thể.

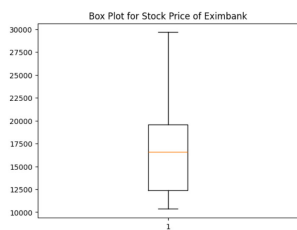
B. DESCRIPTIVE STATISTICS



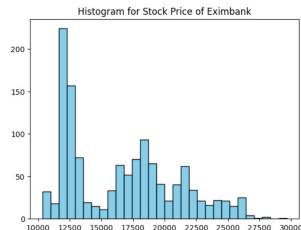
Hình 1. BIDV stock price's boxplot



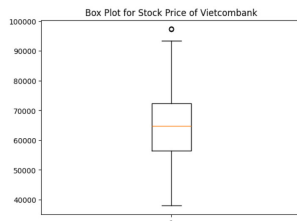
Hình 2. BIDV stock price's histogram



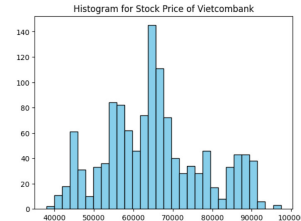
Hình 3. Eximbank stock price's boxplot



Hình 4. Eximbank stock price's histogram



Hình 5. VCB stock price's boxplot



Hình 6. VCB stock price's histogram

C. CÔNG CỤ

Trong nghiên cứu của chúng tôi, chúng tôi đã sử dụng các công cụ phân tích thống kê khác nhau trong Python để hiểu rõ hơn dữ liệu và đưa ra những kết luận ý nghĩa. Các công cụ này, bao gồm numpy, pandas, sklearn và matplotlib.pyplot, đã giúp chúng tôi khám phá ra những phát hiện đáng chú ý. Để biết kết quả chi tiết, vui lòng xem bảng mô tả và biểu đồ được cung cấp.

D. TỶ LỆ PHÂN CHIA TẬP DỮ LIỆU

Nội dung.

E. CÁC CHỈ SỐ ĐÁNH GIÁ MÔ HÌNH

Nội dung.

IV. PHƯƠNG PHÁP NGHIÊN CỨU

A. LINEAR REGRESSION

Nội dung.

B. ARIMA

Nội dung.

C. RNN

Nội dung.

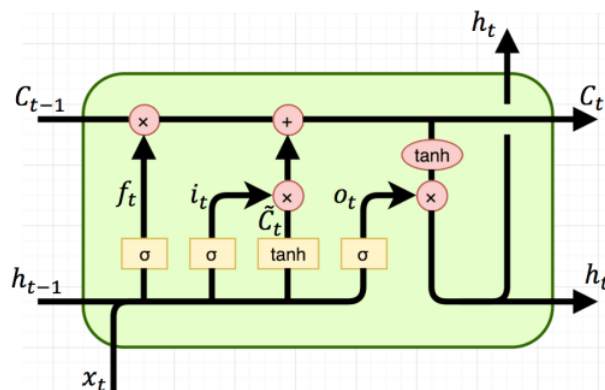
D. LSTM

Mạng bộ nhớ dài-ngắn (Long Short Term Memory networks), thường được gọi là LSTM - là một dạng đặc biệt của RNN, nó có khả năng học được các phụ thuộc xa. LSTM được giới thiệu bởi Hochreiter & Schmidhuber (1997), và sau đó đã được cải tiến và phổ biến bởi rất nhiều người trong ngành. Chúng hoạt động cực kì hiệu quả trên nhiều bài toán khác nhau nên dần đã trở nên phổ biến như hiện nay.

LSTM là một mạng nơ-ron tuần tự sâu trong học sâu, cho phép thông tin tồn tại lâu dài.

Đây là một loại đặc biệt của Mạng Nơ-ron Tái Phát, có khả năng xử lý vấn đề vanishing gradient gặp phải bởi RNN.

- Output: c , $htct$, ht . Ở đây, c biểu diễn trạng thái của ô (cell state) và h biểu diễn trạng thái ẩn (hidden state).
- Input: $ct-1$, $ht-1$, $ht-1$. Trong đó, $xtxt$ là đầu vào tại trạng thái thứ t của mô hình. $ct-1$, $ht-1$, $ht-1$ là đầu ra từ lớp trước. h chơi vai trò tương tự như s trong RNN, trong khi c là điểm mới của LSTM.



Kí hiệu σ , \tanh có nghĩa là dùng sigma, tanh và activation function

f_t , i_t , o_t tương ứng với forget gate, input gate và output gate.

- Forget gate: $f_t = \sigma(U_f * x_t + W_f * h_{t-1} + b_f)$

- Input gate: $i_t = \sigma(U_i * x_t + W_i * h_{t-1} + b_i)$

- Output gate: $o_t = \sigma(U_o * x_t + W_o * h_{t-1} + b_o)$

Nhận xét: $0 < f_t, i_t, o_t < 1$; b_f, b_i, b_o là các hệ số bias.

$c_t = \tanh(U_c * x_t + W_c * h_{t-1} + b_c)$.

$c_t = f_t * c_{t-1} + i_t * c_t$, forget gate quyết định xem cần lấy bao nhiêu từ cell state trước và input gate sẽ quyết định lấy bao nhiêu từ input của state và hidden layer của layer trước.

$h_t = o_t * \tanh(c_t)$, output gate quyết định xem cần lấy bao nhiêu từ cell state để trở thành output của hidden state. Ngoài ra h_t cũng được dùng để tính ra output y_t cho state t .

E. GRU

Nội dung.

F. VARMA

Nội dung.

G. KALMAN FILTER

Nội dung.

H. META-LEARNING

Meta-learning là một phương pháp trong học máy nhằm huấn luyện các mô hình để học một cách hiệu quả các nhiệm vụ mới với lượng dữ liệu hạn chế. Một trong những thuật toán nổi tiếng nhất trong meta-learning là Model-Agnostic Meta-Learning (MAML).

Model-Agnostic Meta-Learning (MAML)

Mục tiêu: MAML nhằm tìm một khởi tạo tốt cho các tham số của mô hình, giúp mô hình có thể được tinh chỉnh nhanh chóng trên một nhiệm vụ mới chỉ với một vài bước cập nhật gradient.

Khái niệm chính

- 1. Phân phối nhiệm vụ ($p(T)p(T)$): Một phân phối trên các nhiệm vụ mà mô hình cần thích nghi.
- 2. Mô hình cơ bản (f_θ): Mô hình được tham số hóa bởi θ .
- 3. Bộ dữ liệu hỗ trợ (Support Set): Bộ dữ liệu nhỏ từ đó mô hình học nhiệm vụ.
- 4. Bộ dữ liệu truy vấn (Query Set): Bộ dữ liệu được dùng để đánh giá mô hình sau khi nó đã thích nghi với nhiệm vụ.

Các bước của thuật toán MAML

- 1. Khởi tạo tham số mô hình: Bắt đầu với các tham số khởi tạo θ .
- 2. Lấy mẫu một loạt các nhiệm vụ: Lấy mẫu một loạt các nhiệm vụ T_i từ phân phối nhiệm vụ $p(T)p(T)$
- 3. Đối với mỗi nhiệm vụ T_i
Lấy mẫu bộ dữ liệu hỗ trợ $D_{T_i}^{\text{train}}$ và bộ dữ liệu truy vấn $D_{T_i}^{\text{test}}$.

Tính gradient của hàm mất mát trên bộ dữ liệu hỗ trợ với tham số θ :

$$\nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta})$$

Cập nhật các tham số sử dụng gradient:

$$\theta'_i = \theta - \alpha \nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta})$$

- 4. Cập nhật Meta

Tính hàm mất mát trên bộ dữ liệu truy vấn sử dụng tham số đã cập nhật θ'_i

$$\mathcal{L}_{T_i}(f_{\theta'_i})$$

Tính gradient của hàm mất mát trên bộ dữ liệu truy vấn đối với các tham số khởi tạo θ

$$\nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta'_i})$$

Cập nhật các tham số khởi tạo θ bằng cách trung bình các gradient qua loạt nhiệm vụ:

$$\theta \leftarrow \theta - \beta \sum_i \nabla_{\theta} \mathcal{L}_{T_i}(f_{\theta'_i})$$

, với β là tốc độ học của meta.

I. NBEATS

Nội dung.

J. N-HITS

Nội dung.

V. KẾT QUẢ

A. THIẾT LẬP MÔ HÌNH

1) Linear Regression

Nội dung.

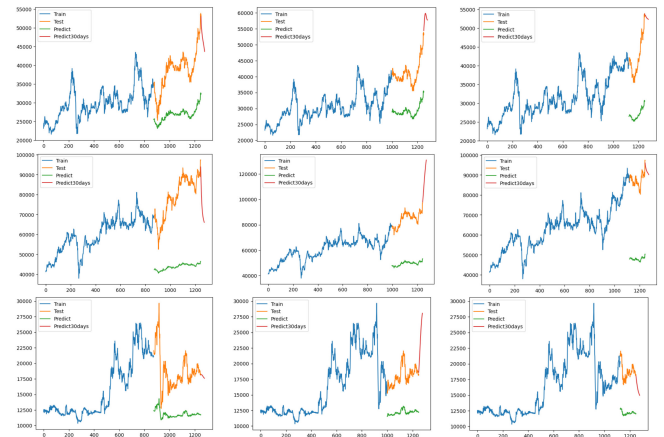
2) ARIMA

Nội dung.

3) RNN

Nội dung.

4) LSTM



Hình 7. Kết quả chạy của mô hình Meta-Learning

5) GRU

Nội dung.

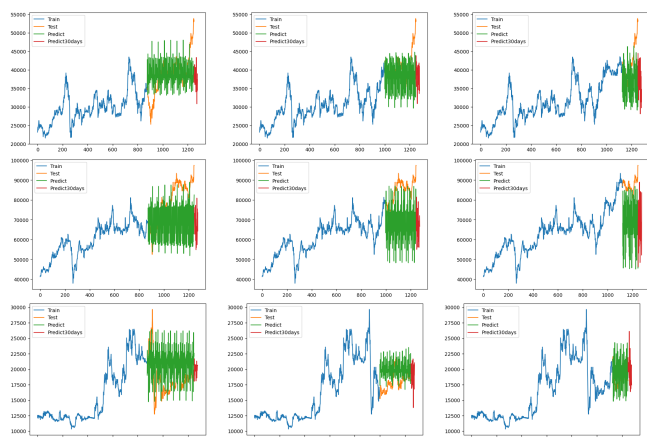
6) VARMA

Nội dung.

7) Kalman Filter

Nội dung.

8) Meta-Learning



Hình 8. Kết quả chạy của mô hình Meta-Learning

9) NBeats

Nội dung.

10) N-HiTS

Nội dung.

B. ĐÁNH GIÁ MÔ HÌNH

1) Evaluation models with BIDV dataset

Nội dung.

2) Evaluation models with VCB dataset

Nội dung.

3) Evaluation models with EIB dataset

Nội dung.

C. DỰ ĐOÁN GIÁ CHO 30 NGÀY TỚI

1) Linear Regression

Nội dung.

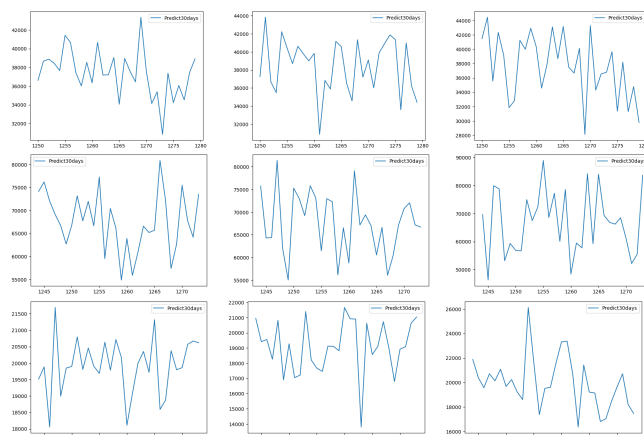
2) ARIMA

Nội dung.

3) RNN

Nội dung.

4) LSTM



Hình 9. Dự đoán 30 ngày tới của mô hình Meta-Learning

5) GRU

Nội dung.

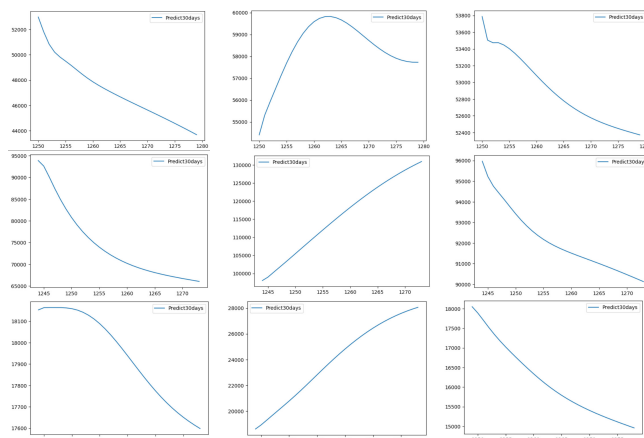
6) VARMA

Nội dung.

7) Kalman Filter

Nội dung.

8) Meta-Learning



Hình 10. Dự đoán 30 ngày tới của mô hình Meta-Learning

9) NBeats

Nội dung.

10) N-HiTS

Nội dung.

TÀI LIỆU

- [1] Arif Mudi Priyatno, Lailatul Syifa Tanjung, Wahyu Febri Ramadhan, Putri Cholidhazia, Putri Zulia Jati and Fahmi Iqbal Firmananda. "Comparison Random Forest Regression and Linear Regression For Forecasting BBKA Stock Price". *Jurnal Teknik Industri Terintegrasi* 6(3):718-732, Jul, 2023.



- [2] Arif Mudi Priyatno, Lailatul Syifa Tanjung, Wahyu Febri Ramadhan, Putri Cholidhazia, Putri Zulia Jati and Fahmi Iqbal Firmananda. "Comparison Random Forest Regression and Linear Regression For Forecasting BBKA Stock Price". *Jurnal Teknik Industri Terintegrasi* 6(3):718-732, Jul, 2023.
- [3] Arif Mudi Priyatno, Lailatul Syifa Tanjung, Wahyu Febri Ramadhan, Putri Cholidhazia, Putri Zulia Jati and Fahmi Iqbal Firmananda. "Comparison Random Forest Regression and Linear Regression For Forecasting BBKA Stock Price". *Jurnal Teknik Industri Terintegrasi* 6(3):718-732, Jul, 2023.