



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 Trước đây, những nhà nghiên cứu thị trường cho rằng việc dự đoán giá cổ phiếu là việc không thể, vì nó phụ thuộc nhiều vào sự biến động của thị trường. Nhưng ngày nay lại có những đề xuất chứng minh rằng việc sử dụng các mô hình thống kê, thuật toán học máy hay học sâu nếu được thiết kế thích hợp thì việc dự đoán giá cổ phiếu có thể được thực hiện rất chính xác. Trong nghiên cứu này, chúng tôi dựa trên các mô hình đã có sẵn để thực hiện dự đoán giá cổ phiếu của 3 ngân hàng lớn tại Việt Nam: Ngân hàng Thương mại cổ phần Ngoại thương Việt Nam (VCB), Ngân hàng Thương mại cổ phần Đầu tư và Phát triển Việt Nam (BIDV), Ngân hàng thương mại cổ phần Xuất Nhập khẩu Việt Nam (Eximbank). Để tìm ra các mô hình nào tốt nhất, chúng tôi lần thực hiện các mô hình sau trên bộ dữ liệu của cả 3 ngân hàng để so sánh: Hồi quy tuyến tính (LR), Đường trung bình động tích hợp tự hồi quy (ARIMA), Mạng thần kinh tái phát (RNN), Đơn vị định kỳ có kiểm soát (GRU), Bộ nhớ ngắn hạn dài hạn (LSTM), thuật toán siêu học (Meta-Learning), Mạng thần kinh tái phát thích ứng (AdaRNN), Bộ lọc Kalman (Kalman Filter), trung bình động tự hồi quy vector (VARMA), Nội suy phân cấp thần kinh để dự báo chuỗi thời gian (N-HiTS). Ngoài ra, chúng tôi còn đo hiệu quả của các mô hình bằng cách sử dụng các độ đo sau: Lỗi bình phương trung bình gốc (RMSE), Lỗi phần trăm tuyệt đối trung bình(MAPE), Lỗi tuyệt đối trung bình(MAE).

INDEX TERMS Stock price prediction, machine learning, deep learning.

I. GIỚI THIỆU

Hiện nay, nhóm cổ phiếu ngành ngân hàng có xu hướng tích cực trở lại, và nó được xem là nhóm cổ phiếu “vua” có thể đóng góp vai trò dẫn dắt đà tăng của thị trường chứng khoán. Giá cổ phiếu là những chỉ số quan trọng giúp các nhà đầu tư đưa ra quyết định đúng đắn.

Dựa trên xu thế đó, chúng tôi tập trung nghiên cứu để tìm ra mô hình dự đoán tốt nhất cho giá cổ phiếu của ngân hàng bằng cách sử dụng kết hợp giữa các kiến thức của học máy và học sâu. Chúng tôi sử dụng các mô hình như: Hồi quy tuyến tính (LR), Đường trung bình động tích hợp tự hồi quy (ARIMA), Mạng thần kinh tái phát (RNN), Đơn vị định kỳ có kiểm soát (GRU), Bộ nhớ ngắn hạn dài hạn (LSTM), thuật toán siêu học (Meta-Learning), Mạng thần kinh tái phát thích ứng (AdaRNN), Bộ lọc Kalman (Kalman Filter), trung bình động tự hồi quy vector (VARMA), Nội suy phân cấp thần kinh để dự báo chuỗi thời gian (N-HiTS).

Hy vọng rằng, việc áp dụng các phương pháp này sẽ mở ra một phương pháp đánh giá mới về sự biến động giá của thị trường chứng khoán, giúp các nhà đầu tư có cái nhìn sáng suốt và đưa ra các quyết định đúng đắn.

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

Nội dung.



B. DESCRIPTIVE STATISTICS

Nội dung.

C. CÔNG CỤ

Nội dung.

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

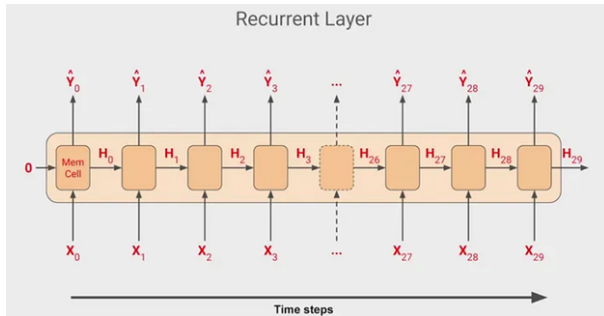
A. LINEAR REGRESSION

Nội dung.

B. ARIMA

Nội dung.

C. RNN



$$h_t = \sigma(W_h x_t + U_h h_{t-1} + b_t)$$

$$y_t = \sigma(W_y h_t + b_y)$$

- h_t : Vector lớp ẩn tại thời điểm t
- X_t : vector đầu vào tại thời điểm t
- \hat{Y}_t : vector đầu ra tại thời điểm t
- W_h, U_h, W_y : Là ma trận trọng số ngẫu nhiên
- b_h, b_y : là các bias
- σ_h, σ_y : là các hàm kích hoạt

Một số hàm kích hoạt phổ biến là:

Sigmoid

$$f(x) = \frac{1}{1 + e^{-x}}$$

Tanh

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$

RELU

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ x & \text{for } x \geq 0 \end{cases}$$

Thuật toán Đầu vào: chuỗi thời gian $x_1, x_2, x_3, \dots, x_t$, nhãn thực tế tương ứng $y_1, y_2, y_3, \dots, y_t$, Đầu ra: Tính toán hàm

mất mát (Loss Function) và cập nhật các trọng số W_h, U_h, W_y, b_h, b_y

$$\text{Loss Function} = \frac{1}{N} \sum_{i=1}^N (-y_i + \hat{y}_i)^2$$

Lặp lại quá trình trên cho đến khi hàm mất mát giảm đủ (không thay đổi trong một số bước lặp nhất định, không thể giảm được nữa)

D. LSTM

Nội dung.

E. GRU

Nội dung.

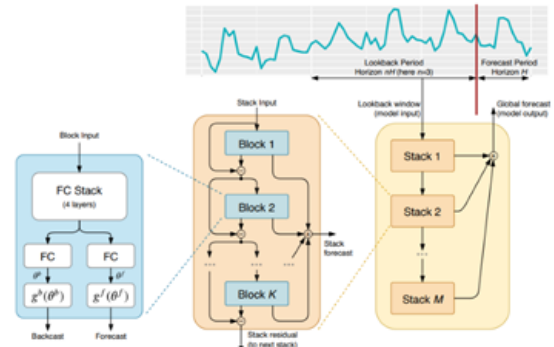
F. KALMAN FILTER

Nội dung.

G. META-LEARNING

Nội dung.

H. NBEATS



$$h_{\ell,1} = FC_{\ell,1}(x_{\ell}), \quad h_{\ell,2} = FC_{\ell,2}(h_{\ell,1}), \quad h_{\ell,3} = FC_{\ell,3}(h_{\ell,2}), \quad h_{\ell,4} = FC_{\ell,4}(h_{\ell,3}).$$

$$\theta_{\ell}^b = \text{LINEAR}_{\ell}^b(h_{\ell,4}), \quad \theta_{\ell}^f = \text{LINEAR}_{\ell}^f(h_{\ell,4}).$$

$$\hat{y}_{\ell} = \sum_{i=1}^{\dim(\theta_{\ell}^f)} \theta_{\ell,i}^f v_i^f, \quad \hat{x}_{\ell} = \sum_{i=1}^{\dim(\theta_{\ell}^b)} \theta_{\ell,i}^b v_i^b.$$

Mô tả cấu trúc mô hình:

Khối Đầu vào (Time series input): của mô hình là một chuỗi thời gian $Y_{t-n+1:t}$ chứa các đặc điểm của quá khứ.

Ngăn xếp khối đầu vào (Stack Input): cấu trúc này bao gồm các khối được xếp chồng lên nhau. Mỗi khối thực hiện 2 nhiệm vụ chính là dự đoán lại quá khứ (backcast) và dự báo tương lai (forecast)

Broadcast: mỗi khối cố gắng tái tạo lại đầu vào (phần quá khứ) để đảm bảo phần dư trước khi chuyển trạng khối tiếp theo

Forecast: Mỗi khi đưa ra dự báo cho tương lai và các dự báo này được tổng hợp lại để tạo ra dự báo cuối cùng

Fully Connected Layers: Mỗi khối báo gồm nhiều lớp fully-connected. Các lớp này sử dụng các hàm kích hoạt để học các biểu diễn phi tuyến của dữ liệu

Residual Connections(kết nối dư): sau mỗi khối, phần dư giữa đầu vào và backcast được tính toán và sử dụng làm đầu vào cho khối tiếp theo

Global Forcecast: Dự báo cuối cùng được tạo ra bằng cách tổng hợp các dự báo từ tất cả các khối. Các dự báo này được kết hợp lại để tạo ra dự báo cuối cùng cho chuỗi thời gian tương lai.

V. KẾT QUẢ

A. THIẾT LẬP MÔ HÌNH

1) Linear Regression

Nội dung.

2) RNN

Nội dung.

3) ARIMA

Nội dung.

4) VARMA

Nội dung.

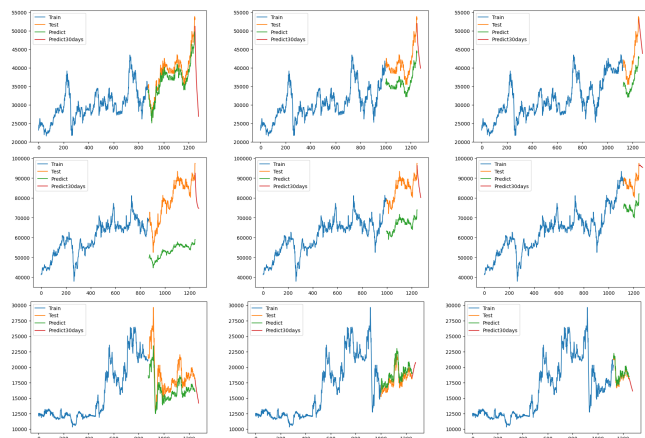
5) Kalman Filter

Nội dung.

6) Meta-Learning

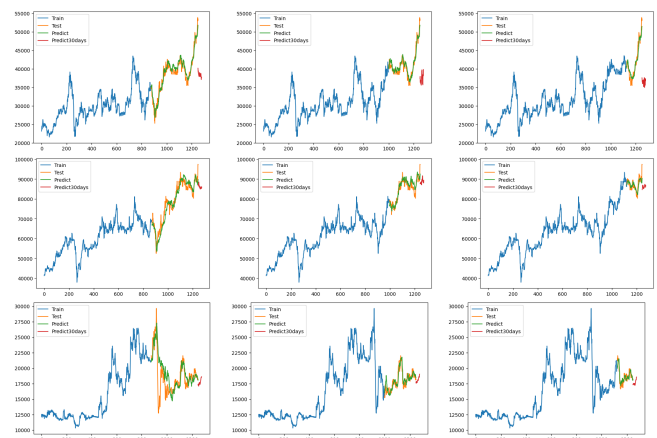
Nội dung.

7) RNN



Hình 1. Kết quả chạy của mô hình VARMA

8) NBeats



Hình 2. Kết quả chạy của mô hình VARMA

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) VARMA

Nội dung.

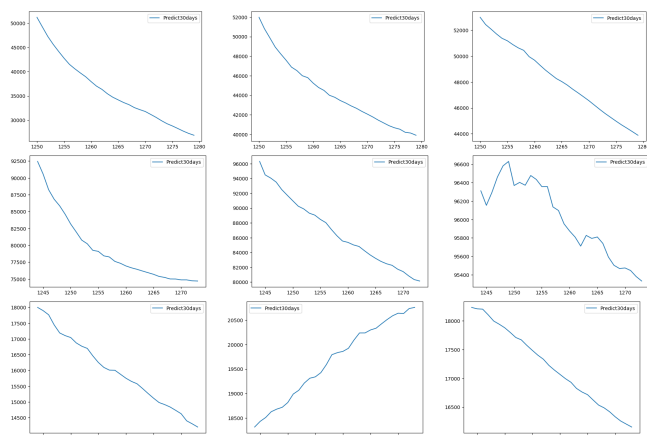
4) Kalman Filter

Nội dung.

5) Meta-Learning

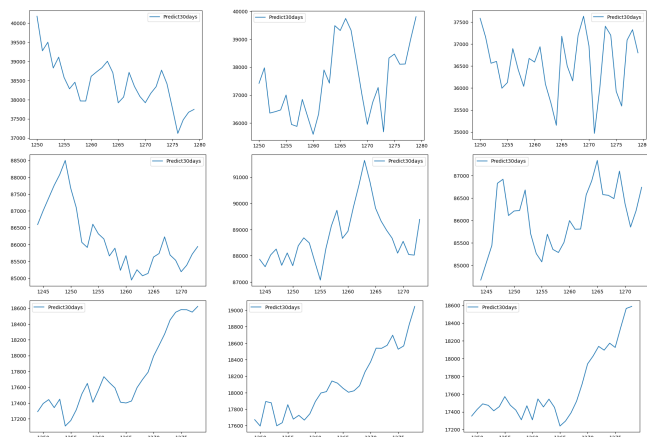
Nội dung.

6) RNN



Hình 3. Dự đoán 30 ngày tới của mô hình RNN

7) NBeats



Hình 4. Dự đoán 30 ngày tới của mô hình RNN

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] Firmananda. "Comparison Random Forest Regression and Linear Regression For Forecasting BBKA Stock Price". *Jurnal Teknik Industri Terintegrasi* 6(3):718-732, Jul, 2023.