

# Phân tích giá xu hướng biến động giá trị của cổ phiếu

Thực hiện: Đỗ Việt Bách

## I. Dữ liệu

	Ticker	Date/Time	Open	High	Low	Close	Volume	Open Interest
0	FPT	12/25/2018 9:15	30.89	30.89	30.89	30.89	35410	0
1	FPT	12/25/2018 9:16	30.81	30.81	30.81	30.81	190	0
2	FPT	12/25/2018 9:17	30.74	30.81	30.74	30.74	1120	0
3	FPT	12/25/2018 9:18	30.74	30.74	30.74	30.74	2120	0
4	FPT	12/25/2018 9:19	30.74	30.74	30.74	30.74	22500	0
...	...	...	...	...	...	...	...	...
97401	FPT	12/22/2020 14:24	58.20	58.20	58.20	58.20	27470	0
97402	FPT	12/22/2020 14:25	58.20	58.20	58.20	58.20	14820	0
97403	FPT	12/22/2020 14:26	58.20	58.20	58.20	58.20	500	0
97404	FPT	12/22/2020 14:29	58.10	58.10	58.10	58.10	2500	0
97405	FPT	12/22/2020 14:46	58.10	58.10	58.10	58.10	11170	0

97406 rows × 8 columns

Hình 1: Dữ liệu cổ phiếu FPT

	Ticker	Date/Time	Open	High	Low	Close	Volume	Open Interest
0	PNJ	2/28/2018 9:15	78.14	78.99	78.14	78.99	270	0
1	PNJ	2/28/2018 9:16	78.94	78.94	78.94	78.94	10	0
2	PNJ	2/28/2018 9:19	78.14	78.14	78.14	78.14	283	0
3	PNJ	2/28/2018 9:20	78.14	78.14	78.14	78.14	480	0
4	PNJ	2/28/2018 9:21	78.14	78.14	78.14	78.14	146	0
...	...	...	...	...	...	...	...	...
125304	PNJ	12/22/2020 14:22	78.20	78.20	78.20	78.20	5000	0
125305	PNJ	12/22/2020 14:23	78.20	78.20	78.20	78.20	1700	0
125306	PNJ	12/22/2020 14:24	78.20	78.20	78.20	78.20	870	0
125307	PNJ	12/22/2020 14:25	78.10	78.10	78.10	78.10	13000	0
125308	PNJ	12/22/2020 14:46	78.30	78.30	78.30	78.30	1500	0

125309 rows × 8 columns

Hình 2: Dữ liệu cổ phiếu PNJ

	Ticker	Date/Time	Open	High	Low	Close	Volume	Open Interest
0	MSN	12/25/2017 9:15	73.1	73.1	73.1	73.1	4210	0
1	MSN	12/25/2017 9:16	73.0	73.0	73.0	73.0	5000	0
2	MSN	12/25/2017 9:18	73.5	73.5	73.5	73.5	210	0
3	MSN	12/25/2017 9:20	73.2	73.5	73.1	73.1	2050	0
4	MSN	12/25/2017 9:21	73.0	73.0	73.0	73.0	1380	0
...	...	...	...	...	...	...	...	...
135349	MSN	12/22/2020 14:23	82.8	82.8	82.8	82.8	700	0
135350	MSN	12/22/2020 14:24	82.7	82.8	82.7	82.8	4660	0
135351	MSN	12/22/2020 14:25	82.8	82.8	82.8	82.8	50	0
135352	MSN	12/22/2020 14:27	82.8	82.8	82.8	82.8	300	0
135353	MSN	12/22/2020 14:46	82.8	82.8	82.8	82.8	200	0

135354 rows × 8 columns

Hình 3: Dữ liệu cổ phiếu MSN

	Ticker	Date/Time	Open	High	Low	Close	Volume	Open Interest
0	VIC	10/17/2018 9:39	98.5	98.5	98.5	98.5	390	0
1	VIC	10/17/2018 9:40	98.5	98.5	98.5	98.5	760	0
2	VIC	10/17/2018 9:41	98.6	98.6	98.6	98.6	2060	0
3	VIC	10/17/2018 9:42	98.6	98.6	98.6	98.6	5540	0
4	VIC	10/17/2018 9:43	98.6	98.7	98.6	98.7	760	0
...	...	...	...	...	...	...	...	...
101261	VIC	12/22/2020 14:22	105.8	105.8	105.7	105.7	1870	0
101262	VIC	12/22/2020 14:23	105.7	105.7	105.7	105.7	2510	0
101263	VIC	12/22/2020 14:24	105.7	105.7	105.7	105.7	3620	0
101264	VIC	12/22/2020 14:25	105.7	105.7	105.7	105.7	100	0
101265	VIC	12/22/2020 14:46	105.7	105.7	105.7	105.7	90	0

101266 rows × 8 columns

Hình 4: Dữ liệu cổ phiếu VIC

- Dựa trên 4 dữ liệu trên em lựa chọn giá trị cổ phiếu lúc đóng cửa để dự đoán (Close). Bởi vì cột Open, High, Low có sự tương đồng. Ở đây, Volume không có sự phân bố giá trị đồng đều nên em loại bỏ để giảm thời gian thực thi của mô hình, mặc dù Volume có ảnh hưởng tương đối đến kết quả dự đoán

## II. Phương pháp

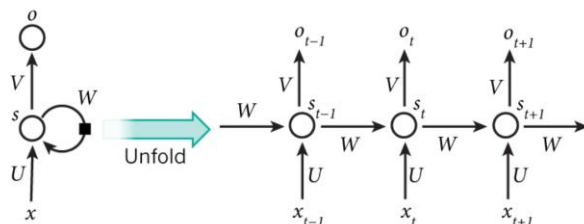
Để dự đoán sự biến thiên của các giá trị có 3 phương pháp chính là:

- Phương pháp Time Series Analysis (Phân tích chuỗi thời gian): sử dụng các mô hình Time Series như ARIMA, SARIMA ... để phân tích và dự đoán xu hướng của chuỗi thời gian.
- Phương pháp Machine Learning: sử dụng các mô hình học máy như Linear Regression, Decision Trees, Random Forests, SVM ... để dự đoán giá dựa trên các đặc trưng (features) khác nhau của dữ liệu.
- Phương pháp Deep Learning: sử dụng các mạng neural sâu (deep neural networks) như LSTM, GRU để mô hình hóa dữ liệu chuỗi thời gian và dự đoán giá trị.

Sau khi phân tích bằng các mô hình ARIMA và Linear Regression không đưa ra kết quả chính xác thì em sử dụng 2 mô hình chính là RNN và GRU để thực hiện phân tích đưa ra kết quả tốt hơn.

### 2.1. Mô hình RNN

Mô hình RNN (Recurrent Neural Network) sử dụng các chuỗi thông tin, tức là RNN có khả năng nhớ các thông tin được tính toán từ trước.



Hình 5: Mô hình RNN

- $x_t$  là đầu vào tại bước  $t$ , với  $x_t$  là một vector.
- $s_t$  là trạng thái ẩn của bước  $t$ . Nó chính là bộ nhớ của mạng.  $s_t$  được tính toán dựa trên các trạng thái ẩn phía trước và đầu vào tại bước đó.  $s_t = f(Ux_t + Ws_{t-1})$
- $o_t$  là đầu ra tại bước  $t$ . Ví dụ như là pha tang của cổ phiếu hoặc ngược lại.

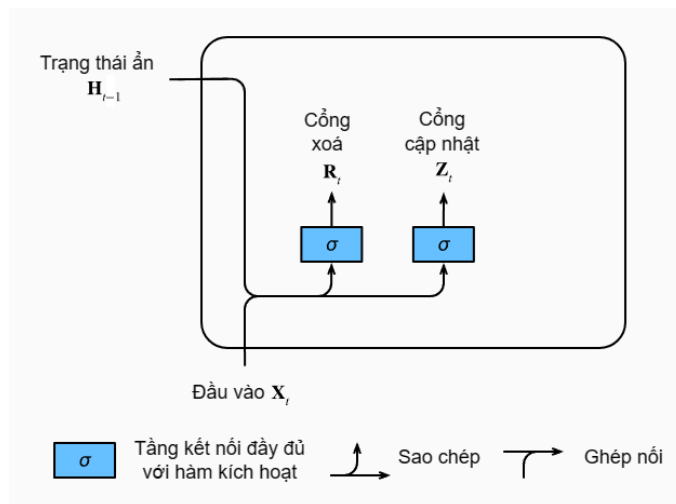
Mô hình RNN không phải mô hình phổ biến để tính toán giá trị cổ phiếu như LSTM hay GRU trong Deep Learning. Nó thường được sử dụng để phân tích giọng nói hay ảnh hơn. Mô hình này là cơ sở để phát triển các mô hình dự báo sự biến thiên của giá trị sau này như LSTM hay GRU.

Việc xây dựng mô hình thuần này khá phức tạp nên em sẽ xây dựng dựa trên thư viện keras.

## 2.2. Mô hình GRU

Mô hình GRU (Gated Recurrent Unit) – Nút hồi tiếp có cổng được xây dựng vào năm 2014, là biến thể gọn hơn của LSTM. Mô hình này hoạt động tốt với tập dữ liệu lớn nhưng yêu cầu thực thi nhanh hơn do có ít tham số hơn so với LSTM.

Sự khác biệt chính giữa RNN thông thường và GRU là GRU hỗ trợ việc kiểm soát trạng thái ẩn. Điều này có nghĩa là ta có các cơ chế được học để quyết định khi nào nên cập nhật và khi nào nên xóa trạng thái ẩn.

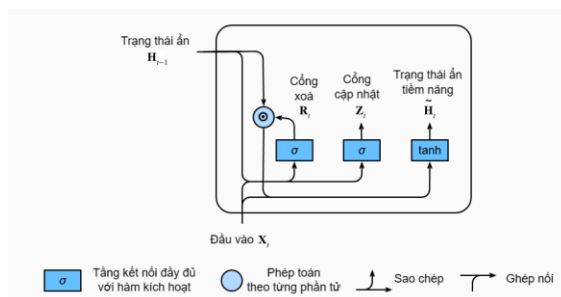


Hình 6: Mô hình GRU

- Bước đầu tiên, ta cần biến đổi tập dữ liệu thành các vector trong khoảng (0, 1).
- Bắt đầu xây dựng mô hình GRU bằng các cổng xóa và cổng cập nhật trong GRU. Với đầu vào là batch  $X_t \in \mathbb{R}(n \times d)$  ( $n$  là số lượng mẫu,  $d$  là số lượng đầu vào) và trạng thái ẩn  $H_{t-1} \in \mathbb{R}(n \times h)$  ( $h$  là số lượng trạng thái ẩn). Cổng xóa  $R_t$  và cổng cập nhật  $Z_t$  được tính như sau:

$$R_t = \sigma(X_t W_{xr} + H_{t-1} W_{hr} + b_r),$$

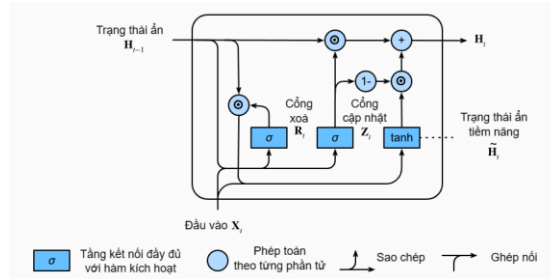
$$Z_t = \sigma(X_t W_{xz} + H_{t-1} W_{hz} + b_z).$$



Hình 7: Cổng xóa

- Đây là minh họa luồng tính toán của cổng xóa. Với  $\tilde{H}_t$  là trạng thái tiềm năng được tính theo công thức:

$$\tilde{H}_t = \tanh(\mathbf{X}_t \mathbf{W}_{xh} + (\mathbf{R}_t \odot \mathbf{H}_{t-1}) \mathbf{W}_{hh} + \mathbf{b}_h).$$



Hình 8: Cổng cập nhật

- Cổng cập nhật xác định mức độ giống nhau giữa trạng thái mới  $H_t$  và trạng thái cũ  $H_{t-1}$ . Với  $H_t$  là trạng thái mới được tính theo công thức:

$$\mathbf{H}_t = \mathbf{Z}_t \odot \mathbf{H}_{t-1} + (1 - \mathbf{Z}_t) \odot \tilde{\mathbf{H}}_t.$$

- Nếu  $Z_t = 1$ , trạng thái mới được giữ như trạng thái cũ. Bỏ qua  $X_t$  trong chuỗi thời gian  $t$  (cổng xóa).
- Nếu  $Z_t \sim 0$ , trạng thái ẩn  $H_t \sim$  trạng thái tiềm năng  $H_t \sim$  (cổng cập nhật).

Việc xây dựng mô hình này phức tạp và chưa có công thức tính rõ ràng nên em xây dựng mô hình này dựa trên thư viện keras.

### III. Kết quả

#### 3.1. Các độ đo

- *RMSE (Root Mean Square Error)*: Sai số dự đoán

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Trong đó:

$n$  là số lượng mẫu trong tập dữ liệu.

$y_i$  là giá trị thực tế.

$\hat{y}_i$  là giá trị dự đoán tương ứng với  $y_i$ .

RMSE càng gần 0, thì độ chính xác càng cao.

- *MAPE (Mean Absolute Percent Error)*: Sai lệch trung bình giữa tỷ lệ phần trăm của dự báo và thực tế.

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100$$

Trong đó:

$n$  là số lượng mẫu trong tập dữ liệu.

$y_i$  là giá trị thực tế.

$\hat{y}_i$  là giá trị dự đoán tương ứng với  $y_i$ .

MAPE càng gần 0, thì độ chính xác càng cao.

- *Accuracy*: Trong mô hình dự báo không thường sử dụng độ đo %, và đây là công thức em thu nhặt được để tính Accuracy.

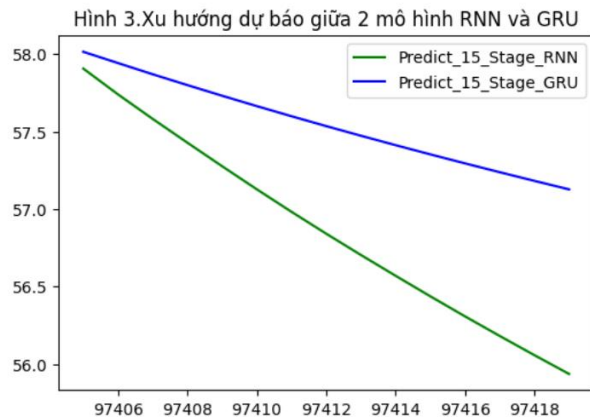
$$\text{Accuracy (\%)} = \left( 1 - \frac{\text{RMSE}}{\text{Data Range}} \right) \times 100$$

Trong đó:

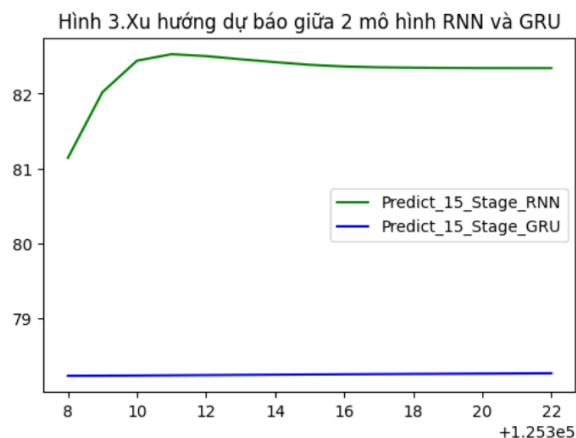
RMSE (Root Mean Squared Error) là độ lỗi bình phương trung bình giữa các giá trị dự đoán và giá trị thực tế trên tập test hoặc tập validation.

Data Range là phạm vi giá trị của biến mục tiêu, ở đây là giá trị cổ phiếu trong cột 'Close' của dataframe df.

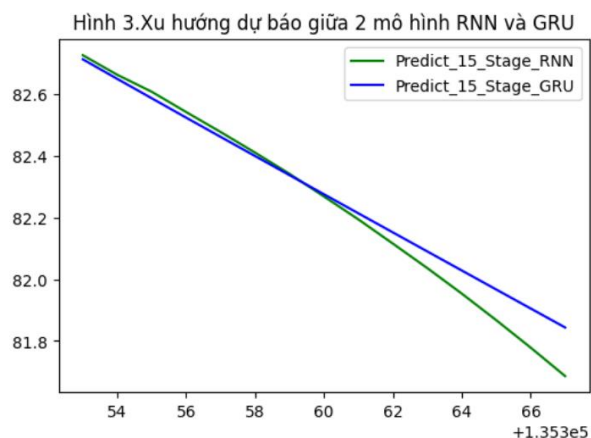
### 3.2.Nhận xét kết quả



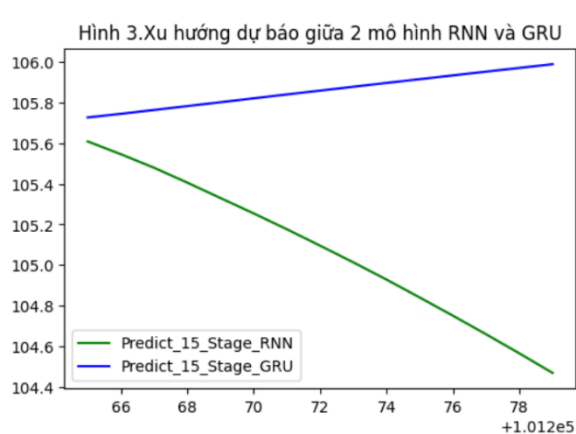
Hình 9: Dự đoán giá cổ phiếu FPT



Hình 10: Dự đoán giá cổ phiếu PNJ



Hình 11: Dự đoán giá cổ phiếu MSN



Hình 12: Dự đoán giá cổ phiếu VIC

- Các mô hình đều có Accuracy, RMSE, MAPE cao. Và dự đoán tốt trên tập test và validate.
- Tuy nhiên, mô hình GRU dự báo tốt trên tập dữ liệu FPT và MSN. Mô hình RNN dự báo tốt trên tập dữ liệu VIC. Trên tập dữ liệu PNJ, GRU cho độ chính xác chưa đáng tin cậy.
- Chi tiết kết quả được lưu trữ trên:

<https://github.com/Bachos0605/Test Phan Tich Gia Co Phieu Findpros>

#### *IV. Tài liệu tham khảo*

- [1] <https://dominhhai.github.io/vi/2017/10/what-is-rnn/>
- [2] [https://d2l.ai/vn.com/chapter\\_recurrent-modern/gru\\_vn.html](https://d2l.ai/vn.com/chapter_recurrent-modern/gru_vn.html)