**ĐẠI HỌC QUỐC GIA THÀNH PHỐ HỒ CHÍ MINH**

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**



**BÁO CÁO ĐỒ ÁN**

**MÔN: PHÂN TÍCH DỮ LIỆU KINH DOANH**

**ĐỀ TÀI: DỰ ĐOÁN GIÁ CỔ PHIẾU DỰA TRÊN CÁC MÔ HÌNH THỐNG KÊ VÀ HỌC MÁY**

**LỚP: IS403.N21**

**GIẢNG VIÊN: TS. NGUYỄN ĐÌNH THUÂN**

**GV. NGUYỄN MINH NHỰT**

Nhóm sinh viên thực hiện:

|  |  |
| --- | --- |
| Nguyễn Hữu Thắng | 20520759 |
| Nguyễn Đình Trải | 19522371 |
| Trần Thu Thảo | 20520769 |
| Võ Phạm Thùy Nhung | 20522043 |
| Võ Thị Hà Trang | 20520679 |

Thành phố Hồ Chí Minh, tháng 6 năm 2023

**LỜI CẢM ƠN**

Lời đầu tiên, nhóm chúng em xin gửi lời cảm ơn chân thành nhất đến các thầy đã trực tiếp giảng dạy, tận tình chia sẻ những kiến thức, hỗ trợ và đóng góp những ý kiến quý báu để giúp chúng em trong quá trình học tập cũng như góp ý về bài báo cáo đồ án của nhóm chúng em.

Bộ môn Phân tích dữ liệu kinh doanh là một môn học thú vị, vô cùng bổ ích và có tính thực tế cao. Đảm bảo cung cấp đủ kiến thức, gắn liền với nhu cầu thực tiễn của sinh viên. Tuy nhiên, do vốn kiến thức còn nhiều hạn chế và khả năng tiếp thu thực tế còn nhiều bỡ ngỡ. Mặc dù nhóm đã cố gắng hết sức nhưng chắc chắn bài báo cáo đồ án khó có thể tránh khỏi những thiếu sót và nhiều chỗ còn chưa chính xác, kính mong các thầy xem xét và góp ý để bài báo cáo đồ án của em được hoàn thiện hơn.

Một lần nữa, nhóm chúng em cảm ơn các thầy rất nhiều vì đã đồng hành cùng nhóm qua một học kỳ.

Tp. Hồ Chí Minh, ngày 19 tháng 06 năm 2023

Nhóm sinh viên thực hiện

**Nhóm 16**

**NHẬN XÉT CỦA GIẢNG VIÊN**

# MỤC LỤC

[MỤC LỤC 4](#_Toc138229868)

[DANH MỤC HÌNH ẢNH 7](#_Toc138229869)

[DANH MỤC BẢNG 9](#_Toc138229870)

[CHƯƠNG 1. ĐẶT VẤN ĐỀ 10](#_Toc138229871)

[1.1. Lý do chọn đề tài 10](#_Toc138229872)

[1.2. Mục tiêu 10](#_Toc138229873)

[1.3. Mô tả bài toán 10](#_Toc138229874)

[CHƯƠNG 2. MÔ TẢ DỮ LIỆU 11](#_Toc138229875)

[2.1. Nguồn dữ liệu 11](#_Toc138229876)

[2.2. Các dữ liệu sử dụng 13](#_Toc138229877)

[CHƯƠNG 3. THUẬT TOÁN 15](#_Toc138229878)

[3.1. Linear Regression 15](#_Toc138229879)

[3.2. Holt-Winters 15](#_Toc138229880)

[3.3. ARIMA 17](#_Toc138229881)

[3.4. TDNN 18](#_Toc138229882)

[3.5. NNAR 19](#_Toc138229883)

[3.6. RNN 20](#_Toc138229884)

[3.7. LSTM 21](#_Toc138229885)

[3.8. GRU 23](#_Toc138229886)

[3.9. BVAR 24](#_Toc138229887)

[3.10. AAE 25](#_Toc138229888)

[CHƯƠNG 4. KẾT QUẢ 27](#_Toc138229889)

[4.1. Thống kê mô tả 27](#_Toc138229890)

[4.2. ARE 31](#_Toc138229891)

[**4.2.1 Linear regression 31**](#_Toc138229892)

[**4.2.2 Holt-Winters 32**](#_Toc138229893)

[**4.2.3 ARIMA 33**](#_Toc138229894)

[**4.2.4 TDNN 34**](#_Toc138229895)

[**4.2.5 NNAR 35**](#_Toc138229896)

[**4.2.6 RNN 36**](#_Toc138229897)

[**4.2.7 LSTM 37**](#_Toc138229898)

[**4.2.8 GRU 38**](#_Toc138229899)

[**4.2.9 AAE 39**](#_Toc138229900)

[4.3. ARI 40](#_Toc138229901)

[**4.3.1 Linear regression 40**](#_Toc138229902)

[**4.3.2 Holt-Winters 41**](#_Toc138229903)

[**4.3.3 ARIMA 42**](#_Toc138229904)

[**4.3.4 TDNN 43**](#_Toc138229905)

[**4.3.5 NNAR 44**](#_Toc138229906)

[**4.3.6 RNN 45**](#_Toc138229907)

[**4.3.7 LSTM 46**](#_Toc138229908)

[**4.3.8 GRU 47**](#_Toc138229909)

[**4.3.9 AAE 48**](#_Toc138229910)

[4.4. MAA 49](#_Toc138229911)

[**4.4.1 Linear regression 49**](#_Toc138229912)

[**4.4.2 Holt-Winters 50**](#_Toc138229913)

[**4.4.3 ARIMA 51**](#_Toc138229914)

[**4.4.4 TDNN 52**](#_Toc138229915)

[**4.4.5 NNAR 53**](#_Toc138229916)

[**4.4.6 RNN 54**](#_Toc138229917)

[**4.4.7 LSTM 55**](#_Toc138229918)

[**4.4.8 GRU 56**](#_Toc138229919)

[**4.4.9 AAE 57**](#_Toc138229920)

[CHƯƠNG 5. CÁC ĐỘ ĐO ĐÁNH GIÁ 59](#_Toc138229921)

[5.1. ARE 59](#_Toc138229922)

[**5.1.1 Các độ đo đánh giá theo tỉ lệ chia 7:2:1 59**](#_Toc138229923)

[**5.1.2 Các độ đo đánh giá theo tỉ lệ chia 6:2:2 59**](#_Toc138229924)

[**5.1.3 Thuật toán tốt nhất 60**](#_Toc138229925)

[5.2. ARI 60](#_Toc138229926)

[**5.2.1 Các độ đo đánh giá theo tỉ lệ chia 7:2:1 60**](#_Toc138229927)

[**5.2.2 Các độ đo đánh giá theo tỉ lệ chia 6:2:2 60**](#_Toc138229928)

[**5.2.3 Thuật toán tốt nhất 61**](#_Toc138229929)

[5.3. MAA 61](#_Toc138229930)

[**5.3.1 Các độ đo đánh giá theo tỉ lệ chia 7:2:1 61**](#_Toc138229931)

[**5.3.2 Các độ đo đánh giá theo tỉ lệ chia 6:2:2 62**](#_Toc138229932)

[**5.3.3 Thuật toán tốt nhất 62**](#_Toc138229933)

[CHƯƠNG 6. KẾT LUẬN 63](#_Toc138229934)

[6.1. Kết luận 63](#_Toc138229935)

[6.2. Hướng phát triển 63](#_Toc138229936)

[**6.2.1 Nâng cao hiệu suất thuật toán: 63**](#_Toc138229937)

[**6.2.2 Mở rộng phạm vi ứng dụng của phân tích dữ liệu: 64**](#_Toc138229938)

[BẢNG PHÂN CÔNG, ĐÁNH GIÁ THÀNH VIÊN 65](#_Toc138229939)

[TÀI LIỆU THAM KHẢO 66](#_Toc138229940)

# DANH MỤC HÌNH ẢNH

[Hình 1: Tập dữ liệu ARE 11](#_Toc138228595)

[Hình 2: Tập dữ liệu ARI 12](#_Toc138228596)

[Hình 3: Tập dữ liệu MAA 13](#_Toc138228597)

[Hình 4: Cấu trúc ô LSTM 21](#_Toc138228598)

[Hình 5: Thuật toán GRU 23](#_Toc138228599)

[Hình 6: Thuật toán AAE 25](#_Toc138228600)

[Hình 7: Box Plot của tập dữ liệu ARE 28](#_Toc138228601)

[Hình 8: Histogram của tập dữ liệu ARE 28](#_Toc138228602)

[Hình 9: Box Plot của tập dữ liệu ARI 29](#_Toc138228603)

[Hình 10: Histogram của tập dữ liệu ARI 29](#_Toc138228604)

[Hình 11: Box Plot của tập dữ liệu MAA 30](#_Toc138228605)

[Hình 12: Histogram của tập dữ liệu MAA 30](#_Toc138228606)

[Hình 13: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 6:2:2) 31](#_Toc138228607)

[Hình 14: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 7:2:1) 32](#_Toc138228608)

[Hình 15: Kết quả dự đoán 30 ngày mô hình Holt- Winters (tỉ lệ 6:2:2) 32](#_Toc138228609)

[Hình 16: Kết quả dự đoán 30 ngày mô hình Holt- Winters (tỉ lệ 7:2:1) 33](#_Toc138228610)

[Hình 17: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 6:2:2) 33](#_Toc138228611)

[Hình 18: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 7:2:1) 34](#_Toc138228612)

[Hình 19: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 6:2:2) 34](#_Toc138228613)

[Hình 20: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 7:2:1) 35](#_Toc138228614)

[Hình 21: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 6:2:2) 35](#_Toc138228615)

[Hình 22: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 7:2:1) 36](#_Toc138228616)

[Hình 23: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 6:2:2) 36](#_Toc138228617)

[Hình 24: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 7:2:1) 37](#_Toc138228618)

[Hình 25: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 6:2:2) 37](#_Toc138228619)

[Hình 26: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 7:2:1) 38](#_Toc138228620)

[Hình 27: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 6:2:2) 38](#_Toc138228621)

[Hình 28: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 7:2:1) 39](#_Toc138228622)

[Hình 29: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2) 39](#_Toc138228623)

[Hình 30: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 7:2:1) 40](#_Toc138228624)

[Hình 31: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 6:2:2) 40](#_Toc138228625)

[Hình 32: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 7:2:1) 41](#_Toc138228626)

[Hình 33: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 6:2:2) 41](#_Toc138228627)

[Hình 34: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 7:2:1) 42](#_Toc138228628)

[Hình 35: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 6:2:2) 42](#_Toc138228629)

[Hình 36: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 7:2:1) 43](#_Toc138228630)

[Hình 37: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 6:2:2) 43](#_Toc138228631)

[Hình 38: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 7:2:1) 44](#_Toc138228632)

[Hình 39: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 6:2:2) 44](#_Toc138228633)

[Hình 40: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 7:2:1) 45](#_Toc138228634)

[Hình 41: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 6:2:2) 45](#_Toc138228635)

[Hình 42: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 7:2:1) 46](#_Toc138228636)

[Hình 43: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 6:2:2) 46](#_Toc138228637)

[Hình 44: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 7:2:1) 47](#_Toc138228638)

[Hình 45: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 6:2:2) 47](#_Toc138228639)

[Hình 46: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 7:2:1) 48](#_Toc138228640)

[Hình 47: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2) 49](#_Toc138228641)

[Hình 48: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 7:2:1) 49](#_Toc138228642)

[Hình 49: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 6:2:2) 49](#_Toc138228643)

[Hình 50: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 7:2:1) 50](#_Toc138228644)

[Hình 51: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 6:2:2) 50](#_Toc138228645)

[Hình 52: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 7:2:1) 51](#_Toc138228646)

[Hình 53: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 6:2:2) 51](#_Toc138228647)

[Hình 54: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 7:2:1) 52](#_Toc138228648)

[Hình 55: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 6:2:2) 52](#_Toc138228649)

[Hình 56: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 7:2:1) 53](#_Toc138228650)

[Hình 57: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 6:2:2) 53](#_Toc138228651)

[Hình 58: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 7:2:1) 54](#_Toc138228652)

[Hình 59: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 6:2:2) 55](#_Toc138228653)

[Hình 60: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 7:2:1) 55](#_Toc138228654)

[Hình 61: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 6:2:2) 56](#_Toc138228655)

[Hình 62: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 7:2:1) 56](#_Toc138228656)

[Hình 63: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 6:2:2) 57](#_Toc138228657)

[Hình 64: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 7:2:1) 57](#_Toc138228658)

[Hình 65: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2) 58](#_Toc138228659)

[Hình 66: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2) 58](#_Toc138228660)

# DANH MỤC BẢNG

[Bảng 1: Mô tả dữ liệu 13](#_Toc138196689)

[Bảng 2 Thống kê mô tả 27](#_Toc138196690)

[Bảng 3: Các độ đo đánh giá theo tỉ lệ chia 7:2:1 trên tập ARE 59](#_Toc138196691)

[Bảng 4: Các độ đo đánh giá theo tỉ lệ chia 6:2:2 trên tập ARE 60](#_Toc138196692)

[Bảng 5: Các độ đo đánh giá theo tỉ lệ chia 7:2:1 trên tập ARI 60](#_Toc138196693)

[Bảng 6: Các độ đo đánh giá theo tỉ lệ chia 6:2:2 trên tập ARI 61](#_Toc138196694)

[Bảng 7: Các độ đo đánh giá theo tỉ lệ chia 7:2:1 trên tập MAA 62](#_Toc138196695)

[Bảng 8: Các độ đo đánh giá theo tỉ lệ chia 6:2:2 trên tập MAA 62](#_Toc138196696)

1. ĐẶT VẤN ĐỀ
   1. Lý do chọn đề tài

Cổ phiếu là một lĩnh vực kinh doanh quan trọng và luôn thu hút sự quan tâm của các nhà đầu tư và chuyên gia trong lĩnh vực tài chính. Trong lĩnh vực cổ phiếu, hiểu và dự báo giá cổ phiếu là một kỹ năng quan trọng và cần thiết để giúp các chuyên gia đưa ra quyết định hợp lý về đầu tư và giao dịch cổ phiếu. Với sự phát triển nhanh chóng của công nghệ và các kỹ thuật phân tích dữ liệu, việc sử dụng các phương pháp phân tích dữ liệu để dự báo giá cổ phiếu là một xu hướng đang được các chuyên gia trong ngành tài chính quan tâm.

* 1. Mục tiêu

Mục tiêu của đề tài là áp dụng các kỹ thuật phân tích dữ liệu để dự báo giá cổ phiếu trong tương lai. Việc này sẽ giúp cho các chuyên gia trong lĩnh vực tài chính đưa ra các quyết định đầu tư và giao dịch cổ phiếu một cách hợp lý và có cơ sở dựa trên phân tích dữ liệu. Đồng thời, việc phân tích dữ liệu cũng giúp cải thiện hiệu quả của việc dự báo giá cổ phiếu và giảm thiểu rủi ro đầu tư trong lĩnh vực tài chính. Ngoài ra, mục tiêu của đề tài này còn là tăng cường kiến thức và kỹ năng phân tích dữ liệu của sinh viên trong môn Phân tích dữ liệu kinh doanh.

* 1. Mô tả bài toán

Từ những dữ liệu cổ phiếu trên các trang mạng trực tuyến hiện nay, chúng ta có thể xây dựng một chương trình có khả năng tự động thu thập dữ liệu theo thời gian thực, lưu trữ dữ liệu và áp dụng các thuật toán phân tích để dự đoán và đánh giá giá cổ phiếu. Chương trình này sẽ giúp các nhà đầu tư và chuyên gia trong lĩnh vực tài chính có thể theo dõi và đưa ra quyết định đầu tư thông qua việc trực quan hóa các giá trị dự đoán của cổ phiếu.

Bằng việc thu thập dữ liệu cổ phiếu từ các nguồn trực tuyến, chương trình sẽ tự động cập nhật thông tin về giá cổ phiếu, chỉ số thị trường, tin tức tài chính và các dữ liệu khác. Sau đó, thông qua việc áp dụng các thuật toán phân tích dữ liệu như dự báo chuỗi thời gian, mạng thần kinh nhân tạo, hay mô hình học máy, chương trình sẽ tiến hành dự đoán giá cổ phiếu trong tương lai.

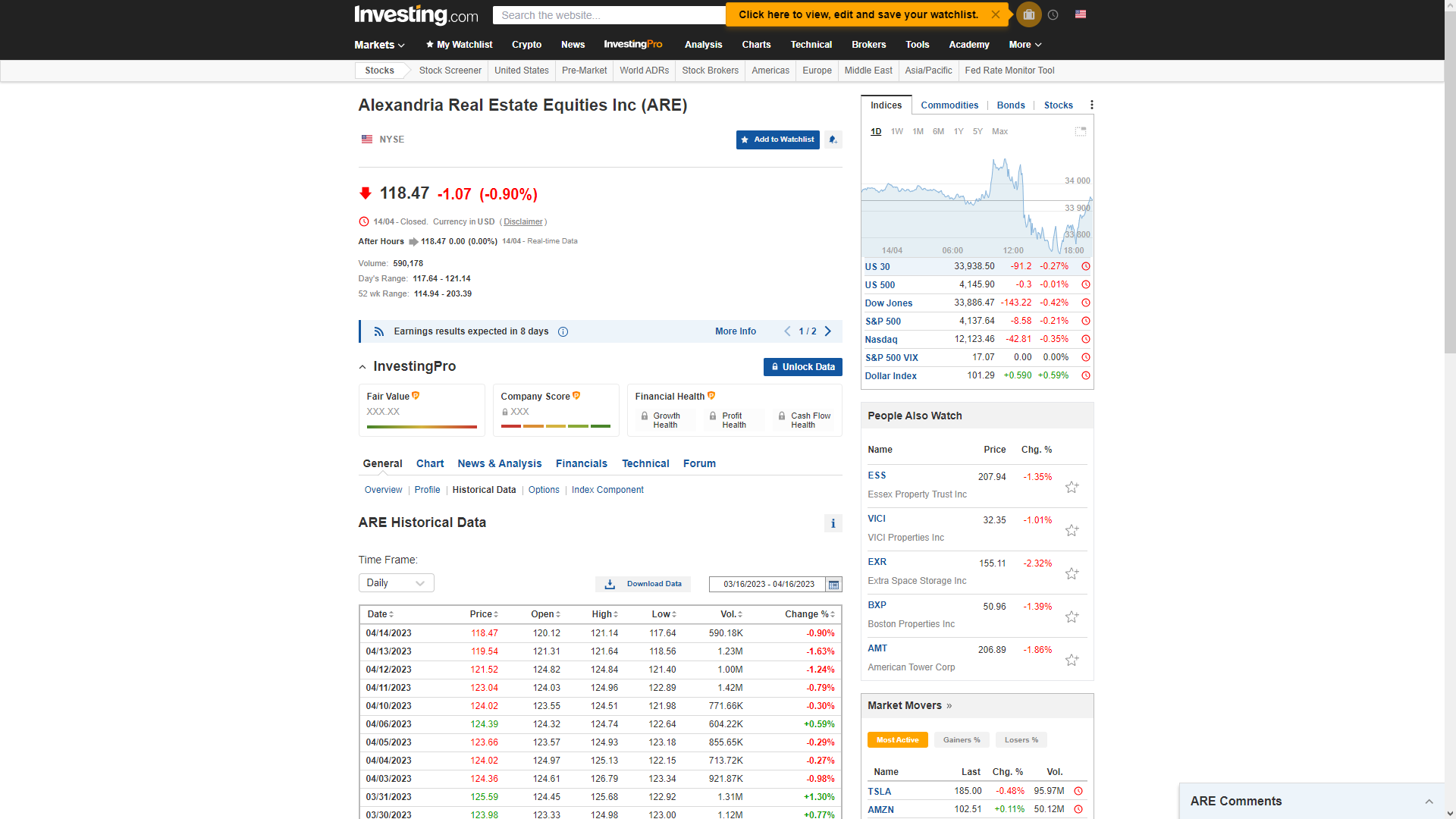
1. MÔ TẢ DỮ LIỆU
   1. Nguồn dữ liệu

Dữ liệu về các chỉ số bất động sản được lấy từ Investing.com, Link: <https://www.investing.com> với 3 bộ dữ liệu:

* Alexandria Real Estate Equities Inc (ARE): bộ dữ liệu của Công ty Cổ phần Bất động sản Alexandria (ARE).
* Apollo Commercial Real Estate Finance Inc (ARI): bộ dữ liệu của Tập đoàn Tài chính Bất động sản Thương mại Apollo (ARI)
* Mid-America Apartment Communities Inc (MAA): là một công ty bất động sản có trụ sở tại Hoa Kỳ.

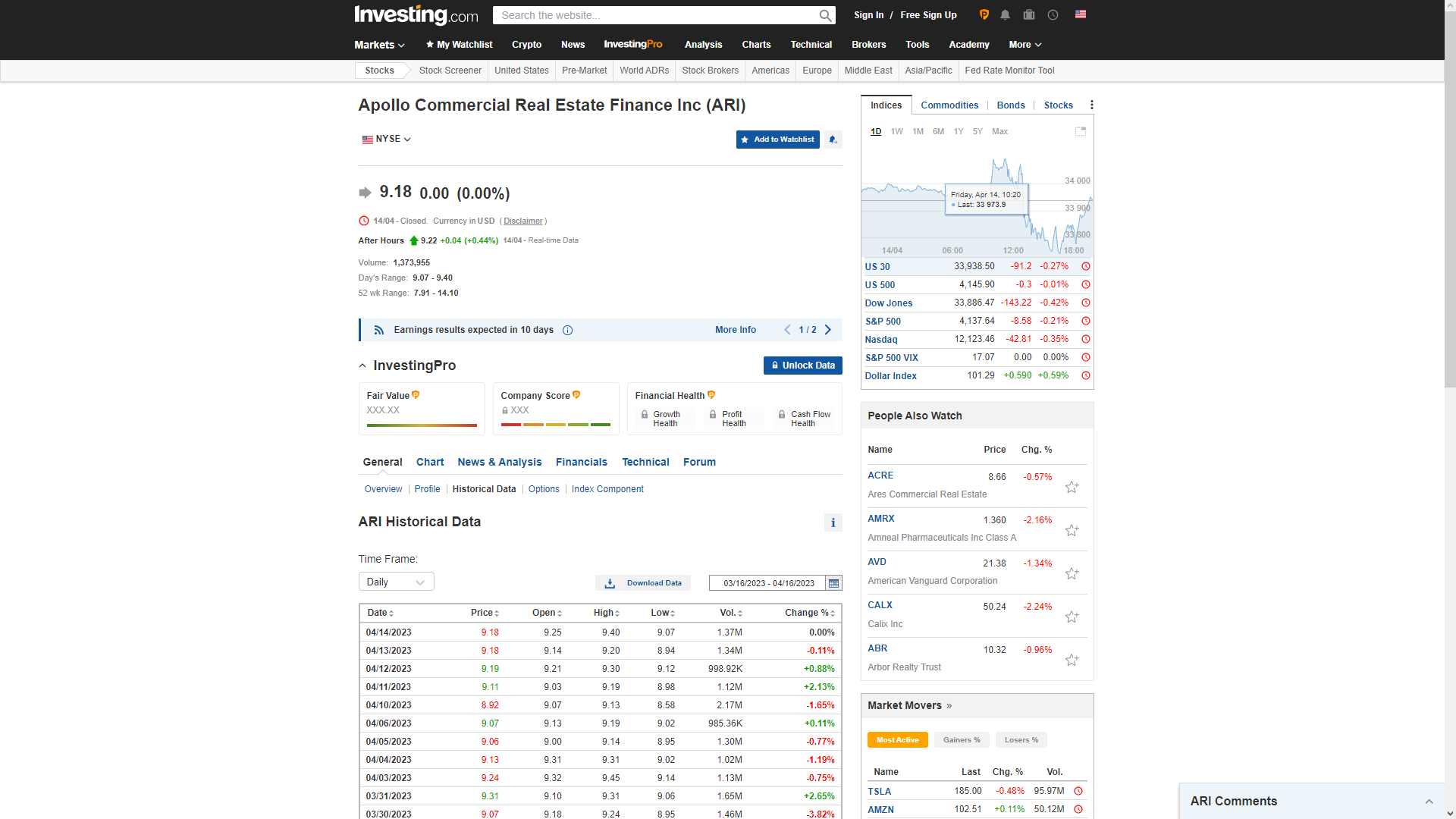
Ba bộ dữ liệu đề cập đến dữ liệu lịch sử về các công ty bất động sản đã nêu trên. Các bộ dữ liệu này chứa thông tin về giá cổ phiếu của các công ty trong một khoảng thời gian nhất định, cho phép nhà đầu tư và các chuyên gia tài chính phân tích và dự đoán xu hướng thị trường bất động sản.

Alexandria Real Estate Equities là một trong những công ty bất động sản hàng đầu tại Hoa Kỳ, chuyên tập trung vào các khu vực nghiên cứu y khoa và khoa học. Bộ dữ liệu lịch sử của công ty này bao gồm thông tin về giá cổ phiếu của công ty từ năm 2010 đến nay.



Hình 1: Tập dữ liệu ARE

Apollo Commercial Real Estate Finance là một công ty bất động sản tài chính, cung cấp các dịch vụ cho các khách hàng trong ngành bất động sản thương mại. Bộ dữ liệu lịch sử của công ty này bao gồm thông tin về giá cổ phiếu của công ty từ năm 2009 đến nay.



Hình 2: Tập dữ liệu ARI

Mid-America Apartment Communities (MAA) là một công ty bất động sản tài chính, hoạt động trong lĩnh vực cho thuê căn hộ và quản lý bất động sản thương mại. Bộ dữ liệu lịch sử của MAA bao gồm thông tin về giá cổ phiếu của công ty từ năm 2016 đến hiện tại. Dữ liệu này được thu thập từ trang web Investing.com, một nguồn thông tin phổ biến về thị trường tài chính và dữ liệu cổ phiếu.

Ảnh có chứa văn bản, phần mềm, Trang web, Website

Mô tả được tạo tự động

Hình 3: Tập dữ liệu MAA

Tất cả ba bộ dữ liệu này đều cung cấp thông tin quan trọng về các công ty bất động sản và giá cổ phiếu của chúng, là tài nguyên quan trọng cho các nhà đầu tư và chuyên gia tài chính để phân tích và dự đoán xu hướng thị trường bất động sản.

* 1. Các dữ liệu sử dụng
* Mô tả dữ liệu:

|  |  |
| --- | --- |
| Tên thuộc tính | Ý nghĩa |
| Date | Ngày giao dịch của cổ phiếu. |
| Price | Giá đóng cửa của cổ phiếu. |
| Open | Giá mở cửa của cổ phiếu. |
| High | Giá cao nhất của cổ phiếu. |
| Low | Giá thấp nhất của cổ phiếu. |
| Vol | Khối lượng giao dịch của cổ phiếu trong ngày giao dịch tương ứng. |
| Change% | Phần trăm thay đổi giá của cổ phiếu so với ngày giao dịch trước đó. |

Bảng 1: Mô tả dữ liệu

Trong bộ dữ liệu, chúng ta tập trung vào phân tích chỉ số "**Price**" của các công ty Mid-America Apartment Communities (MAA), Alexandria Real Estate Equities (ARE), và Apollo Commercial Real Estate Finance (ARI). Chỉ số "**Price**" thể hiện giá cổ phiếu của các công ty này trong một khoảng thời gian nhất định. Bằng việc phân tích chỉ số này, chúng ta có thể hiểu được xu hướng giá cổ phiếu, biến động và hiệu suất của các công ty trong lĩnh vực bất động sản tài chính.

Phân tích chỉ số "**Price**" giúp nhà đầu tư và chuyên gia trong lĩnh vực tài chính đánh giá hiệu suất và sự phát triển của các công ty trong thị trường bất động sản tài chính. Bằng việc theo dõi biến động và xu hướng giá cổ phiếu, chúng ta có thể đưa ra những quyết định đầu tư thông minh và có căn cứ.

Tổng quan, phân tích chỉ số "Price" của các công ty MAA, ARE và ARI trong bộ dữ liệu sẽ cung cấp thông tin quan trọng về giá cổ phiếu và hiệu suất của các công ty trong lĩnh vực bất động sản tài chính.

1. THUẬT TOÁN
   1. Linear Regression

Trong trường hợp đơn giản nhất, mô hình hồi quy cho phép tồn tại mối quan hệ tuyến tính giữa biến phụ thuộc y và một biến độc lập duy nhất x.

Yt=β0+β1xt+εt

Trong đó:

Yt: là biến phụ thuộc

β0: hệ số chặn

β1: hệ số hồi quy cho biến độc lập

εt: sai số

Ảnh có chứa hàng, biểu đồ, ảnh chụp màn hình, Sơ đồ

Mô tả được tạo tự động

Hình 1: Mô hình Linear Regression

Theo hình trên ta thấy các hệ số β0 và β1 lần lượt biểu thị giao điểm và hệ số góc của đường thẳng. Hệ số chặn β¬0 đại diện cho giá trị dự đoán của y khi x=0, β1 đại diện cho sự thay đổi của y khi x thay đổi.

* 1. Holt-Winters

Thuật toán Holt-winters là một phương pháp dự báo chuỗi thời gian, kết hợp Exponential Smoothing với các yếu tố mùa vụ (seasonality) và xu hướng (trend). Ba thành phần làm mịn (smoothing) cần thiết để dự báo trong mô hình holt-winters là level, trend, seasonal, những thành phần này nằm trong khoảng từ 0 đến 1.

Dựa trên yếu tố mùa vụ, các mô hình Holt-winter được phân thành hai loại là mô hình cộng (additive models) và mô hình nhân (multiplicative models).

* ADDITIVE MODEL

Mô hình này được sử dụng khi kích thước dữ liệu không phụ thuộc vào yếu tố mùa vụ. Dự báo cho dữ liệu trước 1 ngày (Ft) và dự báo trước m ngày (Ft+m) có thể được tính toán bằng các phương trình sau:

Trong đó: Các thành phần Level , Trend và Seasonal với giá trị ‘t’ nhất định.

: Giá trị dữ liệu tại một thời điểm nhất định, t

: Hệ số làm mịn

*m*: Giai đoạn dự đoán trước

*s*: Thời lượng theo mùa

Để dự báo, trước tiên khởi tạo và ước tính các thành phần seasonal , có thể được thực hiện bằng các phương trình sau:

Trong đó, IDt: giá trị dữ liệu tại thời điểm tức thời, t.

* MULTIPLICATIVE MODEL

Mô hình được sử dụng khi kích thước dữ liệu phụ thuộc vào yếu tố mùa vụ. Dự báo cho dữ liệu *m* ngày tới ( Ft+m ) có thể được biểu diễn bằng các phương trình sau:

Trong mô hình nhân, các phương trình cho giá trị khởi tạo của level và trend giống với phương trình level và trend trong mô hình cộng. Các thành phần theo mùa khởi tạo được đưa ra là:

* 1. ARIMA

Mô hình ARIMA là viết tắt của quá trình tự hồi quy (AutoRegression -AR), quá trình trung bình trượt (Moving Average – MA) và tích hợp sai phân Integrated – I.

* *Các loại mô hình ARIMA*

Mô hình ARIMA không có tính mùa vụ.

Mô hình ARIMA có tính mùa vụ (Seasonal ARIMA – SARIMA).

* *Chuỗi dừng*

Một chuỗi thời gian có tính dừng là một chuỗi các giá trị mean, variance, autocorrelation không thay đổi theo thời gian và nó không bao hàm các yếu tố xu thế. Với hầu hết các phương pháp thống kê dự báo, đều phải đảm bảo tính dừng của chuỗi dữ liệu vì thế việc kiểm tra tính dừng là rất quan trọng.

Để kiểm định tính dừng của dữ liệu ta có hai phương pháp kiểm định phổ biến: Kiểm định Dickey Fuller3(DF) và Dickey Fuller cải tiến(ADF4).

* *Mô hình ARIMA không có tính mùa vụ*

Với p,d,q lần lượt là các số không âm

I(d): Integrated – So sánh sự khác nhau giữa d quan sát (Hiệu giữa giá trị hiện tại và d giá trị trước đó).

AR(p): Autoregression – là quá trình tìm mối quan hệ giữa dữ liệu hiện tại p dữ liệu quá khứ trước đó. (Gọi là lag)

yt = a0 + a1 yt-1 + a­2 yt-2 + … + ap yt-p + ԑt

Điều kiện dừng của việc chọn p:

MA(q): Moving Average – là quá trình tìm mối quan hệ giữa dữ liệu hiện tại và q phần lỗi quá khứ trước đó

yt =β0 + β1 ԑt-1 + β2 ԑt-2 + … + βq ԑt-q + µt

Điều kiện dừng của việc chọn q:

* 1. TDNN

TDNN (Time Delay Neural Network) là một mô hình mạng nơ-ron thời gian được sử dụng để mô hình hóa và dự đoán chuỗi thời gian. Nó khác biệt với mạng nơ-ron thông thường bằng cách kết nối các nút ẩn với các giá trị vào của các bước thời gian trước đó. Cấu trúc này cho phép TDNN nhìn vào quá khứ và sử dụng thông tin lịch sử để dự đoán giá trị tiếp theo trong chuỗi thời gian. Một kiến trúc TDNN thông thường bao gồm lớp đầu vào, một hoặc nhiều lớp ẩn và lớp đầu ra.

* Lớp đầu vào: Lớp này có số nút bằng số quan sát trễ (lags) của dữ liệu thời gian mà ta muốn sử dụng. Mỗi nút đầu vào đại diện cho một giá trị quan sát trễ.
* Lớp ẩn: Lớp ẩn của TDNN sẽ có một số nút tùy chọn. Mỗi nút ẩn sẽ có một hàm kích hoạt phi tuyến như sigmoid, tanh hoặc ReLU.
* Lớp đầu ra: Lớp này có một nút đầu ra duy nhất để dự đoán giá trị tiếp theo của chuỗi thời gian.
* Biểu thức chung cho mạng TDNN nhiều lớp khi dự đoán giá trị yt+1 là:

[1]

Trong đó:

*f* và *g*là hàm kích hoạt ở lớp ẩn và lớp đầu ra.

*p* là số lượng nút đầu vào.

*q* là số lượng các nút ẩn.

*Bij* là trọng số được gắn vào kết nối giữa nút đầu vào thứ i và nút thứ j của lớp ẩn.

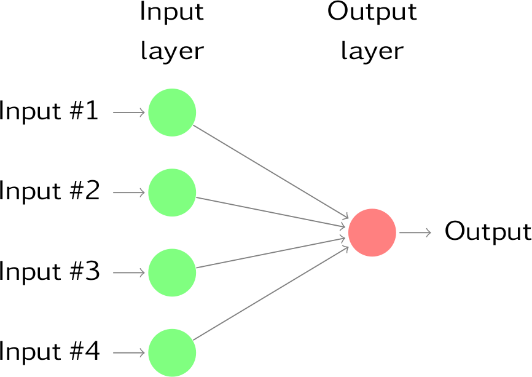
*αj* là trọng số được gắn vào kết nối giữa nút ẩn vào nút đầu ra.

*Yt-i* là đầu vào thứ i (lag) của mô hình.

Trong quá trình huấn luyện, chúng ta sẽ điều chỉnh các trọng số Wi và W’j thông qua các thuật toán huẩn luyện như lan truyền ngược (backpropagation) để tối thiểu hóa sai số dự đoán giữa đầu ra thực tế và đầu ra dự đoán của mạng. Thông qua việc lặp lại quá trình này trên nhiều epoch và sử dụng các thuật toán tối ưu như gradient descent, mạng TDNN sẽ học cách mô hình hóa và dự đoán dữ liệu chuỗi thời gian.

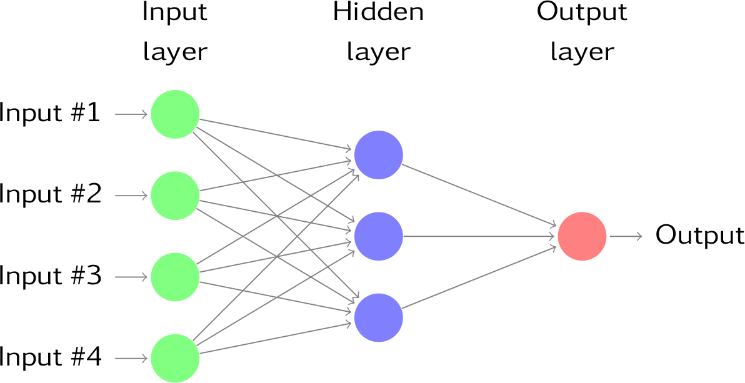
* 1. NNAR
* *Kiến trúc mạng nơ-ron (Neural Network)*

Mạng nơ-ron bao gồm nhiều nơ-ron được tổ chức thành các lớp. Các lớp bao gồm lớp đầu vào, lớp đầu ra và cũng có thể có lớp trung gian chứa các nơ-ron ẩn. Với một mạng nơ-ron đơn giản nhất sẽ không chứa các lớp ẩn và lúc này nó tương đương với hồi quy tuyến tính. Dưới đây là hình minh họa một mạng nơ-ron đơn giản:



Hình 2: Cấu trúc mạng nơ ron đơn giản

Khi chúng ta thêm một lớp trung gian với các nơ-ron ẩn, mạng nơ-ron sẽ trở nên phi tuyến tính được minh họa như hình bên dưới:



Hình 3: Cấu trúc mạng nơ ron 1 lớp ẩn

* *Neural Network Autoregression* (*NNAR)*

Tương tự như khi sử dụng các giá trị lag trong mô hình hồi quy tuyến tính, khi dùng các quan sát trong quá khứ (các giá trị lag) làm đầu vào cho mạng nơ-ron để dự đoán các giá trị đầu ra ta gọi đây là mô hình mạng nơ-ron tự hồi quy - Neural Network Autoregression (NNAR). Xét các mạng chuyển tiếp với một lớp ẩn được ký hiệu:

NNAR(P, k), trong đó P là số lag đầu vào, k là các nút trong lớp ẩn.

* *Hàm tương quan một phần (PACF)*

PACF hay còn gọi là hàm tự tương quan một phần là hàm dùng để xác định bậc tự tương quan trên laq đến p. PACF này được viết bằng phương trình:

* 1. RNN

RNN (Recurrent Neural Network) là một loại neural network được sử dụng chủ yếu cho dữ liệu có cấu trúc chuỗi như chuỗi thời gian, ngôn ngữ tự nhiên và dữ liệu văn bản. Thuật toán RNN sử dụng khái niệm "trạng thái ẩn" (hidden state) để duy trì thông tin về quá khứ và sử dụng thông tin đó khi xử lý dữ liệu trong tương lai.

Chúng ta ký hiệu trạng thái ẩn và đầu vào tại thời điểm t lần lượt là Ht ∈ Rn x h và Xt∈Rn x d, trong đó n là số lượng mẫu, d là số lượng đầu vào của mỗi mẫu và h là số lượng đơn vị ẩn. Hơn nữa, chúng ta sử dụng ma trận trọng số Wxh ∈ Rd×h, ma trận Whh ∈ Rh×h để truyền thông tin từ trạng thái ẩn đến trạng thái ẩn, và tham số bias bh ∈ R1×h.

Cuối cùng, tất cả thông tin này được truyền qua một hàm kích hoạt φ, thường là hàm sigmoid logistic hoặc hàm tanh, để chuẩn bị các gradient để sử dụng trong quá trình lan truyền ngược. Kết hợp tất cả các ký hiệu này lại với nhau, ta thu được Phương trình 1 là biến ẩn ẩn và Phương trình 2 là biến đầu ra.

Ht = φh (XtWxh + Ht−1Whh + bh) (1)

Ot = φo (HtWho + bo) (2)

Vì Ht được tính toán đệ quy dựa trên Ht−1 và quá trình này diễn ra cho mỗi bước thời gian, RNN bao gồm các dấu vết của tất cả các trạng thái ẩn trước đó của Ht−1 cũng như chính Ht−1.

Nếu chúng ta so sánh notation cho RNN với notation tương tự cho Feedforward Neural Networks thẳng, chúng ta có thể thấy rõ sự khác biệt đã được mô tả trước đó. Trong Phương trình 3, chúng ta có thể thấy quá trình tính toán cho biến ẩn ẩn, trong khi Phương trình 4 thể hiện biến đầu ra.

H = φh (XWxh + bh) (3)

O = φo (HWho + bo) (4)

RNN được sử dụng rộng rãi trong nhiều lĩnh vực như xử lý ngôn ngữ tự nhiên, dịch máy, nhận dạng giọng nói, dự đoán chuỗi thời gian và nhiều ứng dụng khác có liên quan đến dữ liệu chuỗi.

* 1. LSTM

LSTM là một loại RNN đặc biệt với các tính năng bổ sung để ghi nhớ chuỗi dữ liệu. Việc ghi nhớ xu hướng trước đó của dữ liệu có thể thực hiện được thông qua một số cổng cùng với một dòng bộ nhớ được tích hợp trong một LSTM điển hình.

Mỗi LSTM là một tập hợp các ô hoặc mô-đun hệ thống, nơi các luồng dữ liệu được thu thập và lưu trữ. Các ô giống như một đường vận chuyển (đường phía trên trong mỗi ô) kết nối từ mô-đun này sang mô-đun khác truyền dữ liệu từ quá khứ và thu thập chúng cho mô-đun hiện tại. Do việc sử dụng một số cổng trong mỗi ô, dữ liệu trong mỗi ô có thể được xử lý, lọc hoặc thêm vào cho các ô tiếp theo.

Cấu trúc bên trong của một ô LSTM được thể hiện trong hình:

A picture containing diagram, line, plan, technical drawing

Description automatically generated

Hình 4: Cấu trúc ô LSTM

Một mô hình LSTM bao gồm ba cổng: *cổng quên* (forget gate), *cổng đầu vào* (input gate) và *cổng đầu ra* (output gate).

* *Forget gate:* Hàm sigmoid thường được sử dụng cho cổng này để đưa ra quyết định về thông tin nào cần được xóa khỏi bộ nhớ LSTM. Đầu ra của cổng này là ft, có giá trị nằm trong khoảng từ 0 đến 1, trong đó 0 biểu thị loại bỏ hoàn toàn giá trị đã học và 1 ngụ ý bảo toàn toàn bộ giá trị. Đầu ra này được tính như sau:
* *Input gate*: Cổng này đưa ra quyết định liệu thông tin mới có được thêm vào bộ nhớ LSTM hay không. Cổng này bao gồm hai lớp: 1) lớp *sigmoid* và 2) lớp “*tanh*”. Lớp *sigmoid* quyết định những giá trị nào cần được cập nhật và lớp *tanh* tạo một vectơ chứa các giá trị ứng cử viên mới sẽ được thêm vào bộ nhớ LSTM. Đầu ra của hai lớp này được tính toán như sau:

Trong đó *it* biểu thị liệu giá trị có cần được cập nhật hay không và biểu thị một vectơ các giá trị ứng cử viên mới sẽ được thêm vào bộ nhớ LSTM. Sự kết hợp của hai lớp này cung cấp một bản cập nhật cho bộ nhớ LSTM trong đó giá trị hiện tại bị quên bằng cách sử dụng lớp cổng quên thông qua phép nhân giá trị cũ ( tức là *ct-1*), sau đó thêm giá trị ứng cử viên mới vào . Phương trình sau biểu diễn phương trình toán học của nó:

trong đó là kết quả của cổng quên, là giá trị từ 0 đến 1 trong đó 0 biểu thị loại bỏ hoàn toàn giá trị; trong khi đó, 1 ngụ ý bảo toàn hoàn toàn giá trị.

* *Output gate:* Đầu tiên, cổng này sử dụng một lớp *sigmoid* để đưa ra quyết định phần nào của bộ nhớ LSTM đóng góp vào đầu ra. Sau đó, nó thực hiện một hàm *tanh* phi tuyến tính để ánh xạ các giá trị giữa −1 và 1. Cuối cùng, kết quả được nhân với đầu ra của một lớp sigmoid. Phương trình sau đây biểu thị các công thức để tính toán đầu ra:

Trong đó *ot* là giá trị đầu ra và *ht* là biểu diễn của nó dưới dạng giá trị giữa −1 và 1.

* 1. GRU

Gated Recurrent Unit (GRU) là một loại mô hình mạng nơ-ron hồi quy (RNN) được giới thiệu nhằm giải quyết vấn đề mất mát thông tin xa trong các mạng RNN truyền thống. GRU sử dụng cơ chế cổng để điều chỉnh quá trình truyền thông tin trong mạng và giúp mô hình nhớ được thông tin quan trọng trong quá khứ.

Ảnh có chứa ảnh chụp màn hình, biểu đồ, hình vuông

Mô tả được tạo tự động

Hình 5: Thuật toán GRU

GRU có cấu trúc đơn giản hơn so với mô hình Long Short-Term Memory (LSTM) nhưng vẫn đạt được hiệu suất tương tự trong nhiều nhiệm vụ. Nó bao gồm hai cổng chính:

* Update gate: Cổng cập nhật xác định tỷ lệ thông tin nào sẽ được cập nhật từ trạng thái trước đó và thông tin mới nào sẽ được truyền vào trạng thái hiện tại. Nó quyết định liệu mô hình có nên quên hoặc ghi nhớ thông tin từ quá khứ.

[2]

Trong đó:

zt là giá trị của Update gate tại thời điểm t.

ht-1 là trạng thái ẩn trước đó.

xt là trạng thái đầu vào hiện tại.

Wz là trọng số của xt.

Uz là trọng số của ht-1.

* Reset gate: Cổng khôi phục xác định tỷ lệ thông tin nào từ trạng thái trước đó sẽ bị đặt lại (reset) để tạo không gian cho thông tin mới. Nó giúp loại bỏ những thông tin không cần thiết hoặc không quan trọng.

[2]

Trong đó:

rt là giá trị của Reset gate tại thời điểm t.

ht-1 là trạng thái ẩn trước đó.

xt là trạng thái đầu vào hiện tại.

Wr là trọng số của xt.

Ur là trọng số của ht-1.

Các cổng cập nhật và khôi phục được kết hợp để tính toán trạng thái ẩn mới (ht) trong GRU. Các giá trị của hai cổng này sẽ quyết định phần nào của thông tin từ trạng thái trước đó sẽ được cập nhật và phần nào sẽ bị đặt lại, từ đó giúp mô hình hiểu và ghi nhớ thông tin quan trọng trong chuỗi dữ liệu. GRU sử dụng một lớp ẩn duy nhất để tính toán trạng thái ẩn mới. Điều này giúp mô hình tiết kiệm tài nguyên tính toán và thời gian huấn luyện.

Tóm lại, GRU là một kiến trúc mạng nơ-ron hồi quy mạnh mẽ và phổ biến trong việc xử lý dữ liệu chuỗi, bao gồm xử lý ngôn ngữ tự nhiên, dịch máy, nhận dạng giọng nói và nhiều ứng dụng khác. Nó cải thiện khả năng mô hình ghi nhớ thông tin xa và giúp giải quyết vấn đề mất mát thông tin trong các mạng RNN truyền thống.

* 1. BVAR

Bayesian Vector Autoregression (BVAR) sử dụng các phương pháp Bayes để ước lượng mô hình VAR. BVAR khác với các mô hình VAR tiêu chuẩn ở chỗ các tham số mô hình được coi là biến ngẫu nhiên, có xác suất biết trước, thay vì các giá trị cố định.

Thuật toán BVAR (Bayesian Vector Autoregression) là một phương pháp dự báo trong lĩnh vực thống kê bayesian được sử dụng để mô hình hóa và dự đoán hệ thống các biến ngẫu nhiên liên kết với nhau trong thời gian. BVAR xem xét mối quan hệ đồng thời giữa các biến phụ thuộc và các biến độc lập trong một hệ thống dữ liệu chuỗi thời gian.

Công thức chính của thuật toán BVAR được biểu diễn dưới dạng mô hình Vector Autoregression (VAR) với một tiếp cận Bayesian. Mô hình VAR có dạng:

Yt = c + A1 \* Yt-1 + A2 \* Yt-2 + ... + Ap \* Yt-p + et

Trong đó:

Yt là vector các biến phụ thuộc tại thời điểm t

c là vector hệ số chặn

A1, A2, ..., Ap là các ma trận hệ số hồi quy

Yt-1, Yt-2, ..., Yt-3 là vector các biến phụ thuộc tại các thời điểm trước đó

et là vector các thành phần lỗi ngẫu nhiên

Thuật toán BVAR áp dụng phương pháp Bayesian để ước lượng các tham số trong mô hình VAR. Nó sử dụng các tiền kiến thức thông qua việc đặt các prior xác định trước cho các tham số và ước lượng posterior của chúng thông qua Markov Chain Monte Carlo (MCMC) hoặc các phương pháp khác.

Việc sử dụng tiếp cận Bayesian trong BVAR cho phép tính toán phân phối posterior của các tham số, giúp ta đánh giá được sự không chắc chắn trong dự đoán và cung cấp phạm vi tin cậy cho kết quả dự báo.

* 1. AAE

AAE (Adversarial Autoencoder) là một kiến trúc mô hình học sâu kết hợp giữa **Autoencoder** và mạng **Adversarial** nhằm chủ yếu để tạo ra dữ liệu mới có cùng phân phối với dữ liệu huấn luyện ban đầu và thực hiện các nhiệm vụ như phân loại, phân đoạn và tái tạo dữ liệu.

A picture containing text, diagram, font, screenshot

Description automatically generated

Hình 6: Thuật toán AAE

* **Autoencoder** là một kiến trúc neural network được sử dụng để học biểu diễn tự động (latent representation) của dữ liệu đầu vào. Nó bao gồm hai phần chính: encoder (bộ mã hóa) và decoder (bộ giải mã). Encoder chuyển đổi dữ liệu đầu vào thành biểu diễn ẩn (latent representation) có chiều thấp hơn, trong khi decoder chuyển đổi biểu diễn ẩn trở lại dữ liệu ban đầu.
* Mạng **Adversarial** là một mạng được sử dụng để tạo ra dữ liệu giả mô phỏng dữ liệu thực tế. Nó bao gồm hai thành phần chính: generator (bộ sinh) và discriminator (bộ phân biệt). Generator nhận đầu vào là một biểu diễn ẩn ngẫu nhiên và cố gắng tạo ra dữ liệu giả mô phỏng dữ liệu thực tế. Trong khi đó, discriminator nhận đầu vào là dữ liệu thực tế và dữ liệu giả và cố gắng phân biệt chúng.

1. KẾT QUẢ
   1. Thống kê mô tả

|  |  |  |  |
| --- | --- | --- | --- |
| **Giá trị** | **ARE** | **ARI** | **MAA** |
| **Count** | 2129 | 2129 | 2129 |
| **Mean** | 136.61879 | 15.45914 | 122.69520 |
| **Min** | 71.65 | 4.24 | 71.15 |
| **Max** | 223.57 | 19.68 | 229.44 |
| **Mode** | 92.64 | 18.66 | 100.0 |
| **Median** | 128.24 | 16.62 | 107.98 |
| **Variance** | 1118.77045 | 10.98388 | 1385.89197 |
| **ST dev** | 33.44 | 3.31419 | 37.22757 |
| **CV** | 0.26082 | 0.19941 | 0.34476 |
| **Skew** | 0.36263 | -0.85507 | 0.93720 |
| **Kurtosis** | -0.62161 | -0.48045 | -0.06742 |
| **25%** | 112.29 | 13.0 | 96.33 |
| **50%** | 128.24 | 16.62 | 107.98 |
| **75%** | 160.71 | 18.26 | 147.31 |
| **100%** | 223.57 | 19.68 | 229.44 |

Bảng 2 Thống kê mô tả

Ảnh có chứa biểu đồ, ảnh chụp màn hình, hàng, Hình chữ nhật

Mô tả được tạo tự động

Hình 7: Box Plot của tập dữ liệu ARE

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, Sơ đồ

Mô tả được tạo tự động

Hình 8: Histogram của tập dữ liệu ARE

Ảnh có chứa ảnh chụp màn hình, biểu đồ, Hình chữ nhật, hàng

Mô tả được tạo tự động

Hình 9: Box Plot của tập dữ liệu ARI

Ảnh có chứa văn bản, biểu đồ, ảnh chụp màn hình, Sơ đồ

Mô tả được tạo tự động

Hình 10: Histogram của tập dữ liệu ARI

Ảnh có chứa biểu đồ, ảnh chụp màn hình, văn bản, hàng

Mô tả được tạo tự động

Hình 11: Box Plot của tập dữ liệu MAA

Ảnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, Sơ đồ

Mô tả được tạo tự động

Hình 12: Histogram của tập dữ liệu MAA

Theo bảng 1 và các hình bên trên, trong 3 loại cổ phiếu thì ARE và MAA có giá trị trung bình gần tương đương nhau và chênh lệch khá lớn so với ARI. Phương sai và độ lệch chuẩn của ARE và MAA khá lớn, cho thấy sự biến động lớn của dữ liệu. Skew của 3 bộ dữ liệu có sự khác nhau dẫn đến độ phân phối của chúng cũng khác nhau, MAA (skewness gần bằng 0) xu hướng gần đối xứng, ARE (skewness âm) phân phối lệch về phía trái, ARI (skewness dương) xu hướng lệch về phía bên phải. Kurtosis, Cả 3 đều âm, phân phối của cả 3 đều nhọn hơn so với phân phối chuẩn, trong đó MAA có phân phối gần với phân phối chuẩn (kurtosis gần bằng 0).

* 1. ARE
     1. Linear regression

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 13: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 14: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 7:2:1)

* + 1. Holt-Winters

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 15: Kết quả dự đoán 30 ngày mô hình Holt- Winters (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 16: Kết quả dự đoán 30 ngày mô hình Holt- Winters (tỉ lệ 7:2:1)

* + 1. ARIMA

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 17: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 18: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 7:2:1)

* + 1. TDNN

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 19: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 20: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 7:2:1)

* + 1. NNAR

A picture containing text, plot, line, font

Description automatically generated

Hình 21: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 6:2:2)

A picture containing text, plot, line, diagram

Description automatically generated

Hình 22: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 7:2:1)

* + 1. RNN

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 23: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 24: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 7:2:1)

* + 1. LSTM

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 25: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 26: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 7:2:1)

* + 1. GRU

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 27: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 28: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 7:2:1)

* + 1. AAE

**A picture containing text, screenshot, font, plot

Description automatically generated**

Hình 29: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 30: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 7:2:1)

* 1. ARI
     1. Linear regression

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 31: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

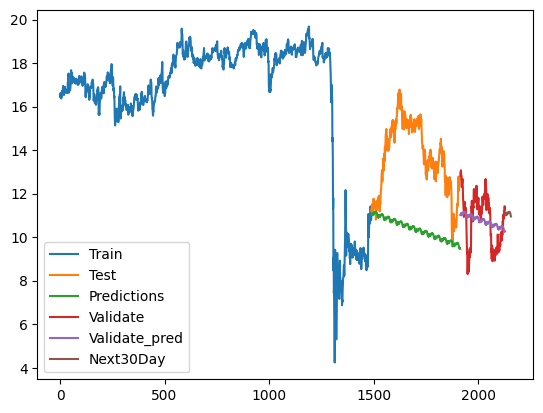
Hình 32: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 7:2:1)

* + 1. Holt-Winters

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 33: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 6:2:2)



Hình 34: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 7:2:1)

* + 1. ARIMA

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 35: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 36: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 7:2:1)

* + 1. TDNN

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 37: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 38: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 7:2:1)

* + 1. NNAR

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 39: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 40: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 7:2:1)

* + 1. RNN

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 41: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 42: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 7:2:1)

* + 1. LSTM

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 43: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 44: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 7:2:1)

* + 1. GRU

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 45: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 46: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 7:2:1)

* + 1. AAE

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 47: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2)A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 48: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 7:2:1)

* 1. MAA
     1. Linear regression

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 49: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 50: Kết quả dự đoán 30 ngày mô hình Linear Regression (tỉ lệ 7:2:1)

* + 1. Holt-Winters

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 51: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 52: Kết quả dự đoán 30 ngày mô hình Holt-Winters (tỉ lệ 7:2:1)

* + 1. ARIMA

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 53: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 54: Kết quả dự đoán 30 ngày mô hình ARIMA (tỉ lệ 7:2:1)

* + 1. TDNN

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 55: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 56: Kết quả dự đoán 7 ngày mô hình TDNN (tỉ lệ 7:2:1)

* + 1. NNAR

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 57: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, line

Description automatically generated

Hình 58: Kết quả dự đoán 30 ngày mô hình NNAR (tỉ lệ 7:2:1)

* + 1. RNN

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 59: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 60: Kết quả dự đoán 30 ngày mô hình RNN (tỉ lệ 7:2:1)

* + 1. LSTM

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 61: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 62: Kết quả dự đoán 30 ngày mô hình LSTM (tỉ lệ 7:2:1)

* + 1. GRU

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 63: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 64: Kết quả dự đoán 30 ngày mô hình GRU (tỉ lệ 7:2:1)

* + 1. AAE

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 65: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2)

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 66: Kết quả dự đoán 30 ngày mô hình AAE (tỉ lệ 6:2:2)

1. **CÁC ĐỘ ĐO ĐÁNH GIÁ**
   1. ARE
      1. Các độ đo đánh giá theo tỉ lệ chia 7:2:1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Thuật toán | RMSE | | MAPE | | MAE | |
|  | **Test** | **Val** | **Test** | **Val** | **Test** | **Val** |
| Linear Regression | 22.52 | 55.35 | 10.16 | 39.09 | 18.51 | 52.12 |
| Holt Winters | 21.37 | 37.52 | 9.35 | 25.13 | 16.77 | 32.76 |
| ARIMA | 25.70 | 30.91 | 11.06 | 20.43 | 21.14 | 26.52 |
| TDNN | 5.45 | 2.89 | 2.45 | 1.59 | 4.32 | 2.22 |
| NNAR | 5.46 | 3.96 | 2.49 | 2.20 | 4.58 | 3.09 |
| RNN | 3.307 | 2.709 | 1.304 | 1.47 | 2.402 | 2.061 |
| LSTM | 3.16 | 2.74 | 1.35 | 1.50 | 2.50 | 2.09 |
| GRU | 2.722 | 2.726 | 1.09 | 1.48 | 1.97 | 2.08 |
| AAE | 0.58 | 0.49 | 3.91 | 3.70 | 0.52 | 0.41 |

Bảng 3: Các độ đo đánh giá theo tỉ lệ chia 7:2:1 trên tập ARE

* + 1. Các độ đo đánh giá theo tỉ lệ chia 6:2:2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Thuật toán | RMSE | | MAPE | | MAE | |
|  | **Test** | **Val** | **Test** | **Val** | **Test** | **Val** |
| Linear Regression | 14.45 | 42.62 | 6.27 | 25.82 | 11.04 | 14.45 |
| Holt Winters | 13.90 | 39.66 | 7.03 | 23.83 | 11.36 | 35.59 |
| ARIMA | 16.51 | 39.16 | 6.74 | 23.46 | 10.93 | 35.14 |
| TDNN | 5.35 | 6.35 | 2.57 | 2.32 | 4.04 | 4.64 |
| NNAR | 5.14 | 5.36 | 2.42 | 2.65 | 4.31 | 4.03 |
| RNN | 3.41 | 3.22 | 1.43 | 1.54 | 2.34 | 2.52 |
| LSTM | 3.37 | 3.14 | 1.40 | 1.50 | 2.28 | 2.45 |
| GRU | 3.32 | 3.03 | 1.34 | 1.43 | 2.15 | 2.28 |
| AAE | 29.76 | 28.81 | 16.43 | 13.64 | 28.50 | 24.34 |

Bảng 4: Các độ đo đánh giá theo tỉ lệ chia 6:2:2 trên tập ARE

* + 1. Thuật toán tốt nhất

Đối với tập dữ liệu ARE, mô hình đạt hiệu suất tốt nhất cho kết quả RMSE tối ưu 0.58 (7-2-1) là AAE, 3.32 (6-2-2) là GRU. Nhưng dựa cách chia và độ đo trên độ đo RMSE chúng tôi kết luận mô hình GRU, RNN và LSTM là các mô hình dự đoán 30 ngày với hiệu suất tốt trên tập dữ liệu này.

* 1. ARI
     1. Các độ đo đánh giá theo tỉ lệ chia 7:2:1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Thuật toán | RMSE | | MAPE | | MAE | |
|  | **Test** | **Val** | **Test** | **Val** | **Test** | **Val** |
| Linear Regression | 1.65 | 2.88 | 10.41 | 26.14 | 1.31 | 2.65 |
| Holt Winters | 3.67 | 1.11 | 23.29 | 9.14 | 3.32 | 0.95 |
| ARIMA | 3.08 | 1.20 | 18.62 | 9.71 | 2.69 | 0.98 |
| TDNN | 0.257 | 0.29 | 1.50 | 1.97 | 0.201 | 0.208 |
| NNAR | 0.36 | 0.40 | 2.19 | 2.83 | 0.29 | 0.30 |
| RNN | 0.256 | 0.294 | 1.46 | 2.01 | 0.195 | 0.256 |
| LSTM | 0.282 | 0.291 | 1.68 | 2.00 | 0.226 | 0.213 |
| GRU | 0.252 | 0.285 | 1.47 | 1.94 | 0.197 | 0.205 |
| AAE | 0.583 | 0.490 | 3.91 | 3.70 | 0.526 | 0.412 |

Bảng 5: Các độ đo đánh giá theo tỉ lệ chia 7:2:1 trên tập ARI

* + 1. Các độ đo đánh giá theo tỉ lệ chia 6:2:2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Thuật toán | RMSE | | MAPE | | MAE | |
|  | **Test** | **Val** | **Test** | **Val** | **Test** | **Val** |
| Linear Regression | 7.93 | 8.62 | 72.41 | 74.23 | 7.29 | 8.40 |
| Holt Winters | 7.21 | 6.97 | 65.42 | 59.71 | 6.49 | 6.71 |
| ARIMA | 7.01 | 6.66 | 63.25 | 57.17 | 6.22 | 6.43 |
| TDNN | 0.52 | 0.36 | 3.62 | 2.47 | 0.35 | 0.27 |
| NNAR | 1.05 | 0.80 | 8.42 | 6.24 | 0.78 | 0.68 |
| RNN | 1.03 | 1.21 | 11.34 | 10.21 | 1.10 | 1.14 |
| LSTM | 1.29 | 0.97 | 10.71 | 7.75 | 0.97 | 0.83 |
| GRU | 1.22 | 1.08 | 10.50 | 9.01 | 1.00 | 1.00 |
| AAE | 0.62 | 0.38 | 3.87 | 2.65 | 0.42 | 0.31 |

Bảng 6: Các độ đo đánh giá theo tỉ lệ chia 6:2:2 trên tập ARI

* + 1. Thuật toán tốt nhất

Đối với tập dữ liệu ARI, mô hình đạt hiệu suất tốt nhất cho kết quả RMSE tối ưu 0.252 (7-2-1) là GRU, 0.62 (6-2-2) là AAE. Dựa trên độ đo RMSE chúng tôi kết luận mô hình GRU, RNN, LSTM là các mô hình dự đoán 30 ngày với hiệu suất tốt trên tập dữ liệu này.

* 1. MAA
     1. Các độ đo đánh giá theo tỉ lệ chia 7:2:1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Thuật toán | RMSE | | MAPE | | MAE | |
|  | **Test** | **Val** | **Test** | **Val** | **Test** | **Val** |
| Linear Regression | 22.52 | 55.35 | 10.16 | 39.09 | 18.51 | 52.12 |
| Holt Winters | 53.01 | 31.29 | 24.20 | 18.23 | 46.30 | 29.27 |
| ARIMA | 60.26 | 34.12 | 27.82 | 20.44 | 53.09 | 32.68 |
| TDNN | 13.17 | 5.55 | 5.70 | 2.82 | 11.03 | 4.08 |
| NNAR | 9.45 | 4.32 | 4.11 | 2.28 | 7.89 | 3.62 |
| RNN | 2.52 | 2.62 | 1.04 | 1.30 | 1.88 | 2.04 |
| LSTM | 3.60 | 2.70 | 1.43 | 1.36 | 2.69 | 2.13 |
| GRU | 2.62 | 2.64 | 1.11 | 1.28 | 2.00 | 2.02 |
| AAE | 22.02 | 11.83 | 10.92 | 6.87 | 18.95 | 10.52 |

Bảng 7: Các độ đo đánh giá theo tỉ lệ chia 7:2:1 trên tập MAA

* + 1. Các độ đo đánh giá theo tỉ lệ chia 6:2:2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Thuật toán | RMSE | | MAPE | | MAE | |
|  | **Test** | **Val** | **Test** | **Val** | **Test** | **Val** |
| Linear Regression | 26.36 | 45.84 | 12.81 | 18.74 | 19.56 | 36.36 |
| Holt Winters | 23.98 | 42.99 | 16.36 | 16.85 | 21.39 | 32.96 |
| ARIMA | 24.05 | 42.82 | 16.43 | 16.78 | 21.46 | 32.81 |
| TDNN | 5.94 | 11.21 | 2.76 | 4.98 | 4.19 | 9.43 |
| NNAR | 4.23 | 4.44 | 2.21 | 2.01 | 2.78 | 3.12 |
| RNN | 2.95 | 3.11 | 1.55 | 1.33 | 2.03 | 2.40 |
| LSTM | 3.00 | 3.76 | 1.568 | 1.566 | 2.05 | 2.89 |
| GRU | 2.89 | 2.90 | 1.49 | 1.26 | 1.92 | 2.24 |
| AAE | 5.08 | 10.46 | 2.50 | 4.11 | 3.64 | 8.00 |

Bảng 8: Các độ đo đánh giá theo tỉ lệ chia 6:2:2 trên tập MAA

* + 1. Thuật toán tốt nhất

Đối với tập dữ liệu MAA, mô hình GRU và RNN đạt hiệu suất tốt nhất cho kết quả RMSE lần lượt là 2.52 (7-2-1), 2.95 (6-2-2) là RNN. Dựa trên độ đo RMSE chúng tôi kết luận GRU, RNN, LSTM là các mô hình dự đoán 30 ngày với hiệu suất tốt trên tập dữ liệu này.

1. KẾT LUẬN
   1. Kết luận

Trong nghiên cứu này, chúng tôi đã tiến hành phân tích và so sánh hiệu suất của một loạt thuật toán dự đoán giá cổ phiếu trong môi trường dữ liệu thời gian. Các thuật toán được xem xét bao gồm Linear Regression, ARIMA, Holt-Winters, TDNN, NNAR, AAE, BVAR, RNN, LSTM và GRU. Kết quả cho thấy LSTM, GRU và RNN là ba thuật toán cho ra kết quả tốt nhất trong việc dự đoán giá cổ phiếu. Cả ba thuật toán này đã chứng minh khả năng dự đoán giá cổ phiếu một cách hiệu quả. Tuy nhiên, chúng tôi nhận thấy rằng việc áp dụng thuật toán BVAR gặp hạn chế và chưa đạt được kết quả như mong đợi, đây là điểm thiếu sót lớn của chúng tôi.

Cần lưu ý rằng các thuật toán dự đoán giá cổ phiếu chỉ dựa trên dữ liệu lịch sử và có rất nhiều yếu tố khác có thể ảnh hưởng đến giá cổ phiếu. Do đó, việc dự đoán giá cổ phiếu vẫn còn mức độ không chắc chắn. Khi áp dụng dự đoán giá cổ phiếu trong thực tế, cần phân tích kỹ lưỡng và cân nhắc các yếu tố khác để có kết quả chính xác.

Tóm lại, nghiên cứu này đã đánh giá hiệu suất của một loạt thuật toán dự đoán giá cổ phiếu. LSTM, GRU và RNN đã cho ra kết quả tốt nhất trong số các thuật toán được nghiên cứu. Tuy nhiên, chúng tôi cũng ghi nhận hạn chế trong việc áp dụng thuật toán BVAR. Những kết quả này có thể đóng góp quan trọng cho phát triển và ứng dụng các mô hình dự đoán giá cổ phiếu trong thực tế.

* 1. Hướng phát triển

Sau quá trình học tập và nghiên cứu cùng thực nghiệm trên dữ liệu thực, hướng phát triển của nhóm bao gồm: nâng cao hiệu xuất của thuật toán và mở rộng phạm vi ứng dụng của phân tích dữ liệu kinh doanh.

* + 1. Nâng cao hiệu suất thuật toán:
* Tối ưu hóa thuật toán hiện tại: Một cách để nâng cao hiệu suất của các thuật toán đã sử dụng trong đồ án là tìm hiểu về các phương pháp tối ưu hóa thuật toán hiện có. Tìm hiểu và nghiên cứu thêm các phương pháp như tăng tốc GPU (Graphics Processing Unit), tối ưu hóa đa luồng (multithreading), hoặc tối ưu hóa mã hóa (code optimization) để giảm thời gian chạy và tăng tốc độ xử lý của thuật toán.
* Sử dụng các thuật toán tiên tiến hơn: Ngoài việc nghiên cứu và sử dụng các thuật toán đã đề cập trong đồ án, có thể khám phá và áp dụng các thuật toán mới nhất và tiên tiến hơn. Điều này có thể đòi hỏi tìm hiểu về các phương pháp và công nghệ mới trong lĩnh vực phân tích dữ liệu kinh doanh.
  + 1. Mở rộng phạm vi ứng dụng của phân tích dữ liệu:
* Phân tích dữ liệu thời gian thực: có thể nghiên cứu và triển khai phương pháp phân tích dữ liệu thời gian thực, cho phép theo dõi và phân tích dữ liệu trong thời gian gần như thực tế. Điều này có thể rất hữu ích trong việc đưa ra quyết định nhanh chóng và phát hiện các xu hướng mới ngay khi chúng xuất hiện.
* Mở rộng ứng dụng trong các lĩnh vực mới: Đồ án có thể tập trung vào một lĩnh vực cụ thể, nhưng có thể xem xét mở rộng phạm vi ứng dụng của phân tích dữ liệu đối với các lĩnh vực khác như marketing, tài chính, nguồn nhân lực, quản lý chuỗi cung ứng và hơn thế nữa. Việc áp dụng phân tích dữ liệu vào các lĩnh vực này có thể giúp tìm ra thông tin quan trọng, nhận diện xu hướng và mô hình dự báo để đưa ra quyết định thông minh và hiệu quả.

# BẢNG PHÂN CÔNG, ĐÁNH GIÁ THÀNH VIÊN

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Công việc** | **Nguyễn Hữu Thắng** | **Nguyễn Đình Trải** | **Trần Thu Thảo** | **Võ Thị Hà Trang** | **Võ Phạm Thùy Nhung** |
| Phân tích đề tài | x | x | x | x | x |
| Thu thập dữ liệu | x | x | x | x | x |
| Trực quan hóa dữ liệu | x | x | x | x | x |
| Thống kê mô tả |  |  | x | x | x |
| Nghiên cứu thuật toán | x | x | x | x | x |
| Thuật toán Linear Regression, NNAR |  |  |  |  | x |
| Thuật toán Holt\_Winters, LSTM |  |  |  | x |  |
| Thuật toán TDNN, GRU |  |  | x |  |  |
| Thuật toán ARIMA, BVAR | x |  |  |  |  |
| Thuật toán RNN, AAE |  | x |  |  |  |
| Đánh giá các độ đo | x | x | x | x | x |
| Phân tích kết quả | x | x | x | x | x |
| Đánh giá thuật toán | x | x | x | x | x |
| Viết báo cáo, bài báo | x | x | x | x | x |
| Đánh giá | 100% | 100% | 100% | 100% | 100% |
| Nhận xét | Tốt | Tốt | Tốt | Tốt | Tốt |

# TÀI LIỆU THAM KHẢO

|  |  |
| --- | --- |
| [1] | Jha, Girish K., and Kanchan Sinha, "Time-delay neural networks for time series prediction: an application to the monthly wholesale price of oilseeds in India," *Neural Computing and Applications 24,* 2014. |
| [2] | Cho, Kyunghyun, et al., "Learning phrase representations using RNN encoder-decoder for statistical machine translation.". |
| [3] | Siami-Namini, Sima, Neda Tavakoli, and Akbar Siami Namin., "A comparative analysis of forecasting financial time series using arima, lstm, and bilstm.," 2019. |
| [4] | Venkateswarlu, G., and A. D. Sarma., "Performance of Holt-Winter and exponential smoothing methods for forecasting ionospheric TEC using IRNSS data," *2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT),* 2017. |
| [5] | Hyndman, Rob J., and George Athanasopoulos, Forecasting: principles and practice, OTexts, 2018. |
| [6] | TensorFlow Time series forecasting, "TensorFlow," [Online]. Available: https://www.tensorflow.org/tutorials/structured\_data/time\_series?hl=en. |
| [7] | As' ad, Mohamad., "Finding the Best ARIMA Model to Forecast Daily Peak Electricity Demand," *2012.* |
| [8] | Schmidt, Robin M., "Recurrent neural networks (rnns): A gentle introduction and overview.," 2019. |
| [9] | Siami-Namini, Sima, and Akbar Siami Namin., "Forecasting economics and financial time series: ARIMA vs. LSTM.," 2018. |
| [10] | Makhzani, Alireza, et al, "Adversarial autoencoders.," 2015. |
| [11] | Benmahdjoub, Khalida, Zohra Ameur, and Mina Boulifa., "Forecasting of rainfall using time delay neural network in Tizi-Ouzou (Algeria)," *Energy Procedia 36 ,* 2013. |
| [12] | As’ad, Mohamad, and Sigit Setyowibowo Sujito, "Prediction of Daily Gold Prices Using an Autoregressive Neural Network.," *Inform 5.2,* 2020. |
| [13] | Mou, Hanlin, and Junsheng Yu., "CNN-LSTM prediction method for blood pressure based on pulse wave.," *Electronics 10.14,* 2021. |
| [14] | Abrego, Lisandro, and Pär Österholm., "External linkages and economic growth in Colombia: insights from a bayesian Var model.," 2008. |
| [15] | Mondal, Prapanna, Labani Shit, and Saptarsi Goswami., "Study of effectiveness of time series modeling (ARIMA) in forecasting stock prices.," *International Journal of Computer Science, Engineering and Applications 4.2,* 2014. |