

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

**TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN**

**KHOA HỆ THỐNG THÔNG TIN**

**BÁO CÁO ĐỒ ÁN CUỐI KỲ**

**PHÂN TÍCH DỮ LIỆU KINH DOANH**

**ĐỀ TÀI:**

**Dự báo giá tiền điện tử bằng**

**các mô hình học máy và thống kê**

GV: **Nguyễn Đình Thuân**

GVHDTH: **Nguyễn Thị Viết Hương**

Mã lớp: **IS403.N22.HTCL**

Nhóm thực hiện: Nhóm 6

|  |  |
| --- | --- |
| Dương Nhật Minh | 20520242 |
| Hồ Thị Hằng | 20521285 |
| Nguyễn Hoàng Ngọc Hạnh | 20521287 |

Hồ Chí Minh, tháng 4 2023

**LỜI CẢM ƠN**

Kính gửi giảng viên Nguyễn Đình Thuân, trợ giảng Nguyễn Minh Nhựt và Nguyễn Thị Viết Hương

Chúng em xin gửi lời cảm ơn sâu sắc đến quý thầy cô đã hỗ trợ chúng em trong quá trình thực hiện đồ án. Với tinh thần tận tâm và trách nhiệm cao, quý thầy cô đã giúp đỡ chúng em vượt qua những khó khăn, thách thức và hoàn thành tốt công việc nghiên cứu đồ án của mình.

Đầu tiên, chúng em xin gửi lời cảm ơn đến quý thầy cô đã truyền đạt kiến thức chuyên môn một cách rõ ràng, cặn kẽ và tận tình. Những giải đáp thắc mắc, những lời khuyên hữu ích và những phản hồi chân thành của quý thầy cô đã giúp đỡ chúng em nắm bắt được kiến thức một cách nhanh chóng và hiệu quả hơn.

Ngoài ra, chúng em cũng muốn bày tỏ lòng biết ơn đến quý thầy cô về sự hỗ trợ quan trọng trong việc chỉnh sửa, đánh giá và phê duyệt bản đồ án của chúng em. Những lời nhận xét, sửa chữa và đề xuất cải thiện của quý thầy cô đã giúp chúng em hiểu được những sai sót và thực hiện đồ án của mình tốt hơn.

Chúng em tin rằng, những kiến thức và kinh nghiệm mà quý thầy cô đã truyền đạt sẽ giúp chúng em phát triển nghề nghiệp và trở thành những người có ích cho xã hội. Một lần nữa, chúng em xin chân thành cảm ơn quý thầy cô vì tất cả những điều tốt đẹp mà quý thầy cô đã mang đến cho chúng em trong suốt thời gian qua.

Kính chúc quý thầy cô sức khỏe, thành công và hạnh phúc!

Trân trọng,

Nhóm tác giả

MỤC LỤC

[**Chương I: Giới thiệu** 6](#_Toc138245600)

[**1.** **Tổng quan đề tài:** 6](#_Toc138245601)

[**2.** **Giới thiệu tập dữ liệu:** 6](#_Toc138245602)

[**1.** **Nguồn tải:** 6](#_Toc138245603)

[**2.** **Phân tích tập dữ liệu:** 7](#_Toc138245604)

[**Chương II: Bài báo khoa học liên quan** 9](#_Toc138245605)

[**Chương III: Mô hình** 10](#_Toc138245606)

[**1.** **Linear Regression(Hồi quy tuyến tính)** 10](#_Toc138245607)

[**2.** **Holt-Winters Exponential Smoothing** 11](#_Toc138245608)

[**3.** **Chuỗi thời gian mờ (Fuzzy Time Series)** 13](#_Toc138245609)

[**5.** **Chuỗi thời gian cấu trúc Bayesian(Bayesian structural time series)** 23](#_Toc138245610)

[**6.** **Fully Convolutional Network** 26](#_Toc138245611)

[**7.** **Phân tích hồi quy quá trình Gaussian (Gaussian Process Regression)** 27](#_Toc138245612)

[**8.** **Autoregressive Integrated Moving Average (ARIMA)** 29](#_Toc138245613)

[**9.** **Gated Recurrent Unit (GRU)** 30](#_Toc138245614)

[**Chương IV: Phương pháp** 32](#_Toc138245615)

[**1.** **Giới thiệu** 32](#_Toc138245616)

[**2.** **Các bước thực hiện** 33](#_Toc138245617)

[**Chương V: Kết quả dự đoán** 35](#_Toc138245618)

[**Chương VI: Tổng kết** 41](#_Toc138245619)

[References 42](#_Toc138245620)

MỤC LỤC HÌNH ẢNH

[Hình 1 Tập dữ liệu ETH 6](#_Toc138245576)

[Hình 2 Tập dữ liệu BNB 7](#_Toc138245577)

[Hình 3 Tập dữ liệu XRP 7](#_Toc138245578)

[Hình 4 Biểu đồ Histogram ETH Hình 5 Biểu đồ Boxplot ETH 8](#_Toc138245579)

[Hình 6 Biểu đồ Histogram BNB Hình 7 Biểu đồ boxplot BNB 8](#_Toc138245580)

[Hình 8 Biểu đồ Histogram XRP Hình 9 Biểu đồ Boxplot XRP 8](#_Toc138245581)

[Hình 10 Hàm liên thuộc μA (x) 14](#_Toc138245582)

[Hình 11 Tập dữ liệu ví dụ 17](#_Toc138245583)

[Hình 12 Mờ hóa tập dữ liệu ví dụ 18](#_Toc138245584)

[Hình 13 Xây dựng và gom nhóm các mối quan hệ mờ 18](#_Toc138245585)

[Hình 14 Xác định nhóm quan hệ mờ 19](#_Toc138245586)

[Hình 15 Dự báo và giải mờ tập dữ liệu ví dụ 19](#_Toc138245587)

[Hình 16 Mạng nơ-ron hồi quy đơn giản 20](#_Toc138245588)

[Hình 17 Mạng nơ-ron hồi quy được kết nối đầy đủ 21](#_Toc138245589)

[Hình 18 Mạng nơ-ron hồi quy hoạt động 21](#_Toc138245590)

[Hình 19 Mô tả kích thước của hình ảnh 26](#_Toc138245591)

[Hình 20 Kiến trúc đặc trưng của FCN 27](#_Toc138245592)

[Hình 21 Các bước thực hiện 34](#_Toc138245593)

[Hình 22Kết quả của mô hình FTS với dataset chia theo 7/2/1 38](#_Toc138245594)

[Hình 23. Kết quả của mô hình FCN với dataset chia theo 6/2/2 39](#_Toc138245595)

[Hình 24 kết quả của mô hình FTS với dataset chia theo 6/2/2 39](#_Toc138245596)

[Hình 25 kết quả của mô hình GRU với dataset chia theo 5/3/2 40](#_Toc138245597)

[Hình 26 Kết quả của mô hình FCN với dataset BNB chia theo 7/2/1 40](#_Toc138245598)

[Hình 27 Kết quả của mô hình FTS với dataset BNB chia theo 7/2/1 41](#_Toc138245599)

MỤC LỤC BẢNG

[Bảng 1 Bảng độ chính xác dự đoán tập dữ liệu ETH 36](#_Toc138235155)

[Bảng 2 Bảng độ chính xác dự đoán tập dữ liệu BNB 36](#_Toc138235156)

[Bảng 3 Bảng độ chính xác dự đoán tập dữ liệu XRP 37](#_Toc138235157)

NHẬN XÉT CỦA GIẢNG VIÊN

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

…………………………………………………………………………………………

**Chương I: Giới thiệu**

1. **Tổng quan đề tài:**

Tiền điện tử đã trở thành một phần không thể thiếu trong hệ thống tài chính toàn cầu, thu hút sự quan tâm và tham gia của hàng triệu người trên thế giới. Trong số các loại tiền điện tử phổ biến, Ethereum (ETH), Binance Coin (BNB) và Ripple (XRP) đã nổi lên như những ngôi sao sáng trong ngành công nghệ chuỗi khối. Sự tăng trưởng và tiềm năng của các loại tiền điện tử này đã khiến các nhà đầu tư và người dùng kỳ vọng vào sự tăng trưởng và lợi nhuận từ việc nắm giữ chúng.

Bài viết này sử dụng các mô hình thống kê như LN, ARIMA, Fuzzy Time Series, Holt-Winters, axcnd machine learning như GRU, RNN, AR-Net, BSTS, GPR để dự đoán giá đóng cửa. đóng cửa) của tiền điện tử (ETH, BNB và XRP). Dựa trên thông tin và phân tích chi tiết, bài viết này hy vọng sẽ cung cấp cái nhìn tổng quan về triển vọng giá của ETH, BNB và XRP, đồng thời giúp người đọc hiểu rõ hơn về các rủi ro và cơ hội. đầu tư vào lĩnh vực tiền điện tử tiềm năng này.

1. **Giới thiệu tập dữ liệu:**
2. **Nguồn tải:**

Các tập dữ liệu được lấy từ trang [finance.yahoo.com](https://uithcm.sharepoint.com/sites/Phntchdliukinhdoanh84/Shared%20Documents/General/BÁO%20CÁO%20PHÂN%20TÍCH%20DỮ%20LIỆU%20KINH%20DOANH.docx) từ ngày 20.6.2018 đến ngày 20.6.2023 bao gồm:

- Tập dữ liệu của ETH:

Ảnh có chứa văn bản, ảnh chụp màn hình, Phông chữ, số

Mô tả được tạo tự động

Hình 1 Tập dữ liệu ETH

- Tập dữ liệu của BNB:

Ảnh có chứa văn bản, Phông chữ, số, hàng

Mô tả được tạo tự động

Hình 2 Tập dữ liệu BNB

- Tập dữ liệu của XRP:

Ảnh có chứa văn bản, Phông chữ, số, ảnh chụp màn hình

Mô tả được tạo tự động

Hình 3 Tập dữ liệu XRP

1. **Phân tích tập dữ liệu:**

**-** Số liệu:

|  |  |  |  |
| --- | --- | --- | --- |
| Measure | ETH | BNB | XRP |
| Min | 84.31 | 4.5287 | 0.1396 |
| Max | 4812.09 | 675.6841 | 1.8392 |
| Mean | 1238.63 | 178.1599 | 0.475 |
| Count | 1827 | 1827 | 1827 |
| Median | 626.41 | 35.2872 | 0.381 |
| Range | 4727.78 | 671.1554 | 1.67 |
| Variance | 1444095 | 33500.7816 | 0.081 |
| Standard Deviation | 1201.71 | 183.0322 | 0.285 |
| Coefficent of Variation | 0.97 | 1.027348 | 0.601 |
| Skewness | 0.94 | 0.62957 | 1.687 |
| Kurtosis | -0.12 | -0.86894 | 2.723 |

**-** Biểu đồ:

**+** ETH:

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

Mô tả được tạo tự độngẢnh có chứa biểu đồ, ảnh chụp màn hình, hàng, Hình chữ nhật

Mô tả được tạo tự động

Hình 4 Biểu đồ Histogram ETH Hình 5 Biểu đồ Boxplot ETH

+ BNB:

A picture containing text, screenshot, diagram, plot

Description automatically generatedA diagram of a box plot

Description automatically generated with low confidence

Hình 6 Biểu đồ Histogram BNB Hình 7 Biểu đồ boxplot BNB

+XRP:

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

Mô tả được tạo tự độngẢnh có chứa văn bản, ảnh chụp màn hình, biểu đồ, thiết kế

Mô tả được tạo tự động

Hình 8 Biểu đồ Histogram XRP Hình 9 Biểu đồ Boxplot XRP

**Chương II: Bài báo khoa học liên quan**

Poongodi M. và cộng sự [1] đã tiến hành Dự đoán giá của tiền điện tử Ethereum trên blockchain trong một hệ thống tài chính công nghiệp, được chấp nhận vào ngày 25 tháng 11 năm 2019 tại Tạp chí Quốc tế về Công nghệ Máy tính và Kỹ thuật Điện của công ty xuất bản học thuật Elsevier BV. Trong nghiên cứu này, dự đoán giá được thực hiện bằng hai phương pháp học máy, gồm hồi quy tuyến tính (LN) và máy vector hỗ trợ (SVM), bằng cách sử dụng một chuỗi thời gian gồm các giá đóng cửa của tiền điện tử ether hàng ngày. Khi sử dụng mô hình đề xuất, phương pháp SVM có độ chính xác cao hơn (96,06%) so với phương pháp LN (85,46%).

Saadah và cộng sự [2] đã áp dụng một số phương pháp học máy và học sâu để dự đoán giá của Bitcoin, Ethereum và Ripple. Các phương pháp bao gồm k-nearest neighbors, support vector machine và LSTM. Kết quả thực nghiệm cho thấy LSTM đạt được chỉ số RMSE tối ưu cho cả ba loại tiền điện tử, với chỉ số RMSE là 928,62 cho Bitcoin, 11,69 cho Ethereum và 0,16 cho Ripple.

V. Derbentsev và cộng sự [3] đã tiến hành dự báo Bitcoin (BTC), Ethereum (ETH) và Ripple (XRP) từ ngày 19 tháng 8 năm 2013 đến ngày 19 tháng 7 năm 2016 thông qua mạng RNN và LSTM. Nghiên cứu này đã điều chỉnh các đơn vị và số lượng epoch và sử dụng chỉ số MAPE. Kết quả cho thấy độ chính xác ngoài mẫu của việc dự báo giá hàng ngày trong ngắn hạn bằng RNN và LSTM, dựa trên chỉ số MAPE, đối với ba loại tiền điện tử có vốn hóa cao nhất (BTC, ETH và XRP) nằm trong khoảng từ 0,92% đến 2,61%.

Kiki Ramadani, Dodi Devianto [4] trình bày một mô hình dự báo sử dụng logic mờ để ước tính giá của Bitcoin tại HỘI NGHỊ QUỐC TẾ VỀ KHOA HỌC VÀ KHOA HỌC ỨNG DỤNG (ICSAS2020) với Phân tích chuỗi Markov FTS. Trong bài báo nghiên cứu này, kết quả của Phân tích chuỗi Markov FTS có chỉ số MAPE tốt nhất là 8,8%, nhưng Phương pháp Logic Chen phân đoạn FTS lại là tốt nhất vì có các chỉ số MAE và RMSE tốt hơn (355,51 và 678,01).

Mohammad J. Hamayel và Amani Yousef Owda [5] đã trình bày một cách sáng tạo để dự đoán giá tiền điện tử bằng cách sử dụng các thuật toán học máy như GRU, LSTM và bi-LSTM, cho thấy kết quả đầy hứa hẹn trong lĩnh vực dự báo giá. Chỉ số RMSE, MAPE của GRU cho BTC: 174,129 và 0,2454%; ETH: 26,59 và 0,8267%, LTC: 0,825 và 0,2116%, vì vậy GRU có thể dự đoán giá tiền điện tử tốt hơn LSTM và bi-LSTM nhưng nhìn chung tất cả các thuật toán đều cho kết quả dự đoán xuất sắc.

Dian Utami Sutiksno và cộng sự [6] đã có một nghiên cứu đáng chú ý trong lĩnh vực dự báo dữ liệu lịch sử của Bitcoin. Bài viết này sử dụng phương pháp ARIMA và chỉ báo α-Sutte để dự đoán xu hướng và biến động của giá Bitcoin trong tương lai. Trong nghiên cứu này, ARIMA không phải là phương pháp có độ chính xác tốt nhất để dự đoán dữ liệu lịch sử của bitcoin với MSE, MAE là 295797,315 và 497,657.

**Chương III: Mô hình**

1. **Linear Regression(Hồi quy tuyến tính)**

**Hồi quy** là một phương pháp thống kê để thiết lập mối quan hệ giữa một biến phụ thuộc và các biến độc lập.

*Ví dụ: Tuổi = 5 + Chiều cao \* 10 + Cân nặng \* 13*

Ở đây chúng ta đang thiết lập mối quan hệ giữa Chiều cao & Cân nặng với Tuổi của một người.

**Hồi quy tuyến tính** là một phương pháp để dự đoán biến phụ thuộc (Y) dựa trên giá trị của một hay nhiều biến độc lập (X)

Hồi quy tuyến tính có 2 loại là: Hồi quy tuyến tính đơn giản và hồi quy tuyến tính đa biến

**Hồi quy tuyến tính đơn giản** là phương pháp thiết lập mối quan hệ giữa biến phụ thuộc Y và một biến độc lập X. Được biểu diễn bởi phương trình:

Y = β0 + β1X + e

**Hồi quy tuyến tính đa biến** là phương pháp thiết lập mối quan hệ giữa biến phụ thuộc Y và nhiều biến độc lập X. Được biểu diễn bởi phương trình:

Y = β0 + β1X1 + β2X2 +…+ βnXn+ e

Hồi quy tuyến tính có thể được sử dụng để dự đoán giá trị của biến phụ thuộc dựa trên các biến độc lập. Với mối quan hệ tuyến tính được xác định, mô hình hồi quy có thể được sử dụng để tạo dự đoán và ước tính các giá trị trong tương lai.

1. **Holt-Winters Exponential Smoothing**

***Exponential Smoothing (ES)*** là một phương pháp phân tích chuỗi thời gian và dự báo.

Có 3 loại chính của ES là: Simple Exponential Smoothing, Double Exponential Smoothing và Triple Exponential Smoothing.

***Simple Exponential Smoothing (SES)*** là dạng cơ bản nhất, thích hợp để áp dụng cho dữ liệu không có tính xu hướng và không có tính mùa.

***Double Exponential Smoothing*** là dạng mở rộng của SES bằng cách thêm thành phần xu hướng, thích hợp để áp dụng cho dữ liệu có tính xu hướng và không có tính mùa.

***Triple Exponential Smoothing*** hay còn gọi là Holt-Winters Exponential Smoothing. Đây là dạng phức tạp nhất và là dạng mở rộng của 2 dạng trên, thích hợp để áp dụng cho dữ liệu có tính xu hướng và có tính mùa.

Trong bài này, sẽ tìm hiểu cụ thể về dạng thứ 3 là Holt-Winters

Holt-Winters lần đầu tiên được đề xuất bởi Charles C. Holt, Peter Winters, vào năm 1960 sau khi đọc một cuốn sách xử lý tín hiệu từ những năm 1940 về làm mịn hàm mũ.

Tùy thuộc vào bản chất của thành phần theo mùa, Holt Winter cũng có 2 dạng đó là: Holt Winter Additive và Holt Winter Multiplicative

Ở đây chúng ta sẽ trình bày dạng Holt Winter Multiplicative

Phương pháp này có 3 hệ số làm mịn là α, β và γ. Với các bước như sau:

Đầu tiên, chúng ta sẽ ước lượng giá trị của đối tượng nghiên cứu tại thời điểm t với công thức:

Xt = αYt / St – m + (1 – α)(Xt-1 + Tt-1)

Trong đó:

+ Xt là giá trị ước lượng tại thời điểm t

+ Xt-1 là giá trị ước lượng tại thời điểm t-1

+ Yt là giá trị thực tế ở thời điểm t

+ St – m là giá trị ước lượng cho chỉ số thời vụ

+ Tt-1 là thành phần xu hướng tại thời điểm t-1

+ m là chu kỳ trong thời vụ

+ α là trọng số.

Tiếp theo, ta sẽ ước lượng thành phần xu hướng:

Tt = β(Xt – Xt-1) + (1 – β)Tt-1

Trong đó:

+ Tt: là thành phần xu hướng tại thời điểm t

+ β : là trọng số được gán cho thành phần xu hướng

Sau cùng, ta ước lượng thành phần thời vụ S tại thời điểm t

St = γYt / Xt + (1 – γ)St – m

Trong đó:

+ St là thành phần thời vụ tại thời điểm t

+ γ là trọng số được gán cho thành phần thời vụ

Phương pháp của Holt-Winters cho thấy cách phân tích chủ động, nghĩa là khi có một giá trị quan sát mới, nó sẽ được dùng để ước lượng thành phần xu hướng và thời vụ để cập nhật, và sau đó dùng thông tin này để dự báo cho tương lai.

Điều quan trọng là xác định hệ số α, β, γ sao cho sai số dự báo nhỏ nhất.

**Các cách chọn trọng số:**

+ Dựa trên kinh nghiệm phân tích, sự am hiểu về đối tượng nghiên cứu cũng như các yếu tố tác động đến nó đã biết từ trước đó. Ví dụ các nhà phân tích biết được các giá trị ở quá khứ gần hơn thường giúp ích nhiều cho việc dự báo nên sẽ đánh trọng số cao và ngược lại

+ Dựa trên sai số dự báo, trọng số nào có sai số dự báo thấp nhất sẽ được chọn. Cách đánh giá dựa trên so sánh các chỉ tiêu MAE, MSE, RMSE, MAPE.

+ Dựa vào tính biến động, mức độ tác động của yếu tố ngẫu nhiên không thể kiểm soát. Ví dụ sau khi loại bỏ yếu tố thời vụ, yếu tố xu hướng mà dữ liệu Time series chưa thực sự bằng phẳng (nhìn trên đồ thị) thì không thể đặt trọng số quá cao cho các giá trị quá khứ ở mốc thời gian gần với mốc dự báo, và ngược lại nếu dữ liệu có vẻ bằng phẳng hơn.

Để dự báo giá trị trong tương lai ta sử dụng công thức:

Ft+ h = (Xt + h\*Tt)\*St+h-m

Trong đó:

+ Ft+ h là giá trị dự đoán trong tương lai tại thời điểm t+h

+ h là số bước thời gian muốn dự đoán.

1. **Chuỗi thời gian mờ (Fuzzy Time Series)**

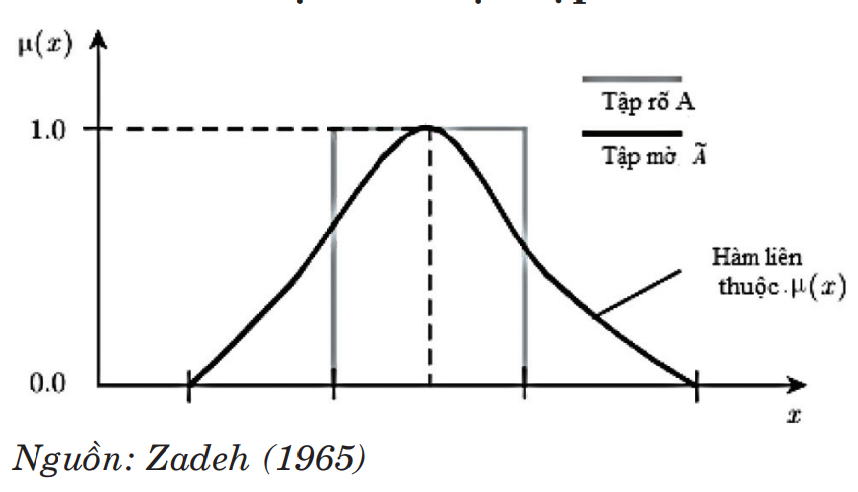
***Ra đời:*** Năm 1965, Zadeh đề xuất khái niệm tập mờ như một công cụ để kiểm tra mức độ thành viên chưa biết. Chuỗi thời gian mờ là một phương pháp phân tích bắt nguồn từ khái niệm tập mờ.

Năm 1993, Song và Chissom đã kết hợp thành công khái niệm tập mờ với mô hình chuỗi thời gian và bắt đầu nghiên cứu về chuỗi thời gian mờ.

* Tập mờ: Giả sử H là một không gian nền, nếu V là một tập con rõ của H thì ta có thể xây dựng hàm chỉ báo (Indicator finction) chính xác như sau:

Nếu A là một tập con mờ của H, thì quan hệ x thuộc A hay không thuộc A không được đánh giá rõ ràng bằng hàm chỉ báo như trên mà được thể hiện qua một hàm liên thuộc (membership function) cho biết khả năng để x thuộc vào tập A.

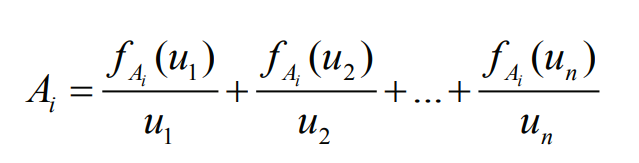
Nếu hàm cho kết quả 0 thì x không thuộc A, nếu hàm cho kết quả 1 thì x là thành viên toàn phần của tập hợp. Các phần tử x có giá trị nằm trong khoảng (0,1) được gọi là các thành viên mờ của tập A.



Hình 10 Hàm liên thuộc μA (x)

Hàm liên thuộc có các tính chất sau:

* Chuỗi thời gian mờ: Cho U là một không gian nền với .

Một tập mờ (không hạn chế số lượng) của U được xác định bởi: 

Trong đó: là hàm liên thuộc của tập mờ với thuộc tập mờ và là hàm thể hiện khả năng thuộc vào của .

Ta có:

Ảnh có chứa văn bản

Mô tả được tạo tự động

Nếu:

- uk hoàn toàn không thuộc Ai, = 0.

- uk hoàn toàn thuộc Ai, = 1.

- uk  có thuộc Ai, = (0,1).

**Định nghĩa 1:** Song và Chissom(1993) định nghĩa chuỗi thời gian mờ như sau:

Cho với *t= …,* 0, 1, 2… Nếu là một tập mờ trên tập nền thì tập hợp được gọi là chuỗi thời gian mờ trên nền .

F(t) là một hàm của thời gian t, nghĩa là các giá trị của F(t) có thể khác nhau ở các thời điểm khác nhau do không gian nền có thể khác nhau ở các thời điểm nhau.

**Định nghĩa 2:** Giả sử Giả sử F(t) chỉ bị F(t-1) tác động (nghĩa là F(t) 🡪 F(t-1)) thì mối quan hệ này có thể được biểu diễn dưới dạng , trong đó, R(t,t-1) là một quan hệ mờ và được gọi là mô hình bậc nhất của F(t).

**Định nghĩa 3:** Cho F(t) là một chuỗi thời gian mờ. Nếu tại mọi thời điểm t, F(t) = F(t - 1) và F(t) có hữu hạn phần tử thì F(t) được gọi là chuỗi thời gian mờ bất biến theo thời gian. Nếu ngược lại, F(t) được gọi là chuỗi thời gian mờ biến thể theo thời gian.

**Ví dụ:**

- Quan sát thời tiết tại một địa điểm trong một năm có rất nhiều trạng thái: nóng, rất nóng, lạnh, rất lạnh, hơi lạnh, hơi nóng, ….

- Tâm trạng của một người bình thường trong một khoảng thời gian có rất nhiều trạng thái: tốt, rất tốt, buồn, rất buồn, tệ, ….

***Nhược điểm*** của phương pháp dự báo bằng chuỗi thời gian mờ:

+ Chuỗi thời gian mờ chỉ áp dụng dự báo tốt trong ngắn hạn. Tuy nhiên đã có nhiều nghiên cứu mở rộng ứng dụng phương pháp chuỗi thời gian mờ để dự báo trong dài hạn.

+ Phương pháp chuỗi thời gian mờ còn khá mới mẻ trong các nghiên cứu lớn và chưa có những phương pháp riêng để đánh giá mức độ chính xác của dự báo đối với chuỗi thời gian mờ.

***Ưu điểm*** của phương pháp dự báo bằng chuỗi thời gian mờ:

+ Không quan tâm đến tính dừng của chuỗi thời gian. Chuỗi thời gian dừng hay không dừng đều có thể áp dụng phương pháp chuỗi thời gian mờ.

+ Dự báo được trong trường hợp số liệu thu thập được ở dạng mờ, không biết con số chính xác

+ Dự báo được cho những trường hợp dữ liệu thu thập được không liên tục.

***Trình tự thực hiện dự báo:***

+ Xác định không gian nền U = [Dmin - D1,Dmax + D2]

+ Chia U thành các khoảng tương ứng

+ Xác định các tập mờ

+ Mờ hóa chuỗi thời gian đã có

+ Xây dựng và gom nhóm các mối quan hệ mờ

+ Xác định nhóm quan hệ mờ dùng dự báo

+ Dự báo và giải mờ:

* Vế phải có 1 phần từ: trung bình cộng của khoảng tương ứng.
* Vế phải có nhiều phần từ: trung bình cộng từng khoảng, rồi lấy trung bình cộng chung của tất cả các giả trị trung bình cộng vừa tính.
* Vế phải không có phần từ: tự gán vế phải bằng vế trái.

- Ví dụ: Cho tập dữ liệu:

A picture containing text, screenshot, pattern

Description automatically generated

Hình 11 Tập dữ liệu ví dụ

+ Xác định không gian nền U: U = [Dmin - D1,Dmax + D2] = [13000, 20000] (D1, D2 tùy ý).

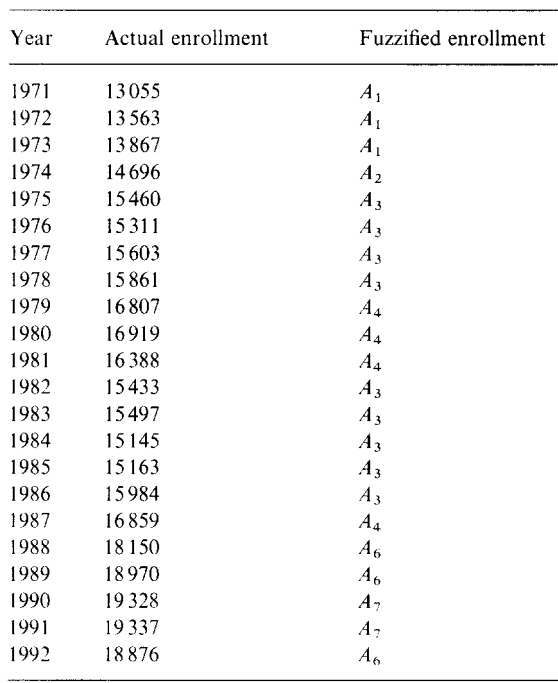
+ Chia U thành các khoảng bằng nhau: ul =

[13000, 14000], u2 = [14000, 15 000], u3 = [15000, 16000], u4 = [16000, 17 000], u5 = [17000, 18000], u6 = [18000, 19000], and u7 = [19000,20000].

+ Xác định các tập mờ : A1, A2, A3, A4, A5, A6, A7 (không giới hạn) tương đương 7 khoảng không gian nền được chia.

+ Mờ hóa chuỗi thời gian đã có:

Ví dụ: năm 1971 có enrollment=13055 thuộc u1=[13000, 14000] 🡪 Fuzzified enrollment = A1.



Hình 12 Mờ hóa tập dữ liệu ví dụ

+ Xây dựng và gom nhóm các mối quan hệ mờ

A picture containing font, text, handwriting, line

Description automatically generated

Hình 13 Xây dựng và gom nhóm các mối quan hệ mờ

+ Xác định nhóm quan hệ mờ:

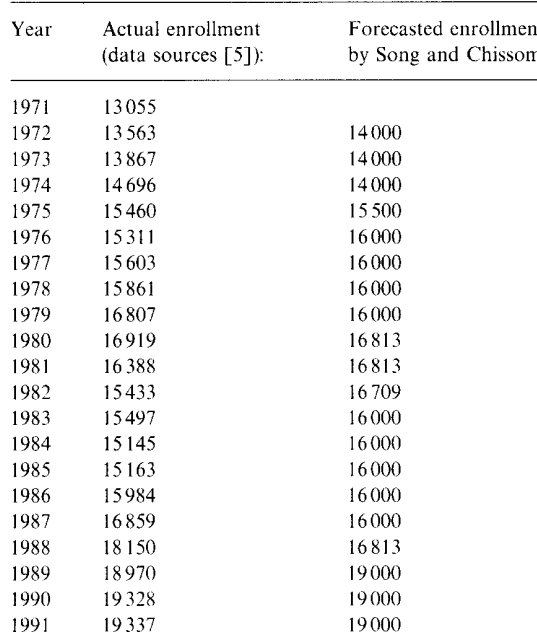
A picture containing text, font, white, line

Description automatically generated

Hình 14 Xác định nhóm quan hệ mờ

+ Dự báo và giải mờ:

Ví dụ: dự đoán 1972, ta có enrollment 1971 = 13055 thuộc A1, mà ta có group 1 A1🡪A1 và A1🡪A2, vậy kết quả dự đoán 1972 = ((13000+14000)/2 + (14000+15000)/2)/2 = 14000

****

Hình 15 Dự báo và giải mờ tập dữ liệu ví dụ

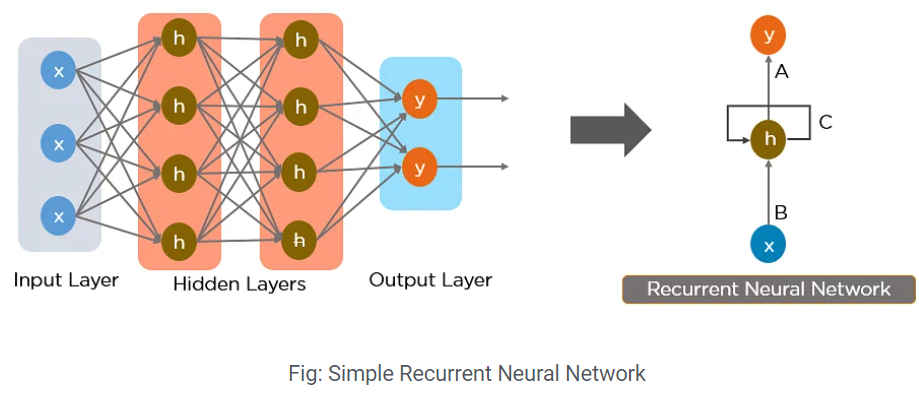
1. **Mạng nơ-ron hồi quy (Recurrent Neural Network)**

***Ra đời:*** Wilhelm Lenz và Ernst Ising đã tạo và phân tích mô hình Ising (1925), về cơ bản là một mạng thần kinh tái phát nhân tạo không học tập (RNN) bao gồm các phần tử ngưỡng giống như nơ-ron. Vào năm 1972, Shun'ichi Amari đã làm cho kiến ​​trúc này trở nên thích ứng. RNN học tập của ông đã được phổ biến rộng rãi bởi John Hopfield vào năm 1982.

- **Mạng nơ-ron hồi quy** là một loại mạng nơ-ron trong đó đầu ra từ bước trước được cung cấp làm đầu vào cho bước hiện tại. Trong các mạng nơ-ron truyền thống, tất cả các đầu vào và đầu ra đều độc lập với nhau, nhưng trong trường hợp cần dự đoán từ tiếp theo của câu, thì các từ trước đó là bắt buộc và do đó cần phải ghi nhớ các từ trước đó. Do đó, RNN ra đời, giải quyết vấn đề này với sự trợ giúp của lớp ẩn (Hidden Layer).

- Tính năng chính và quan trọng nhất của RNN là trạng thái ẩn (Hidden State) của nó, ghi nhớ một số thông tin về chuỗi. Trạng thái đó còn được gọi là trạng thái bộ nhớ (Memory State), vì nó ghi nhớ đầu vào trước đó cho mạng. Nó sử dụng các tham số giống nhau cho dữ liệu mỗi đầu vào vì nó thực hiện cùng một nhiệm vụ trên tất cả các dữ liệu đầu vào hoặc các lớp ẩn để tạo ra dữ liệu đầu ra. Điều này làm giảm độ phức tạp của các tham số, không giống như các mạng thần kinh khác.

- Cách chuyển đổi từ mạng nơ ron chuyển tiếp (Feed-Foward Neural Network) liệu thành mạng nơ ron hồi quy:



Hình 16 Mạng nơ-ron hồi quy đơn giản

Trong đó:

+ x: Các nút trong các lớp đầu vào

+ h: Các nút trong lớp ẩn

+ y: Các nút trong lớp đầu ra

Các nút trên được nén để tạo thành một lớp trong mạng nơ-ron hồi quy.

+ A, B, C là các tham số của mạng.

- Tại bất kỳ thời điểm t nào, đầu vào hiện tại là sự kết hợp của đầu vào tại x(t) và x(t-1). Đầu ra tại bất kỳ thời điểm nào được tìm lưu vào mạng để cải thiện đầu ra.

A picture containing text, screenshot, diagram, circle

Description automatically generated

Hình 17 Mạng nơ-ron hồi quy được kết nối đầy đủ

A picture containing circle, diagram

Description automatically generated

Hình 18 Mạng nơ-ron hồi quy hoạt động

+ Trong các mạng nơ-ron hồi quy, thông tin quay vòng qua một vòng lặp đến lớp ẩn ở giữa.

+ Lớp đầu vào 'x' nhận đầu vào của mạng nơ-ron và xử lý nó rồi chuyển nó lên lớp giữa.

+ Lớp giữa 'h' có thể bao gồm nhiều lớp ẩn, mỗi lớp có các chức năng kích hoạt và trọng số cũng như độ lệch riêng. Nếu có một mạng nơ-ron trong đó các tham số khác nhau của các lớp ẩn khác nhau không bị ảnh hưởng bởi lớp trước đó, tức là: mạng nơ-ron không có bộ nhớ, thì có thể sử dụng mạng nơ-ron hồi quy.

h(t) được tính bằng công thức: A black text on a white background

Description automatically generated with medium confidence

Một số activation function thường được sử dụng: linear function: f(x) = x; sigmoid function: A picture containing font, white, text, calligraphy

Description automatically generated; tanh function: .

y(t) được tính bằng công thức: A black letter on a white background

Description automatically generated with low confidence

trong đó: Whh, Wxh, Why lần lượt là trọng số của hidden state trước đó, của input đến hidden state hiện tại và trọng số của hidden state đến output.

- Mạng nơ-ron hồi quy sẽ chuẩn hóa các chức năng kích hoạt khác nhau, trọng số và độ lệch để mỗi lớp ẩn có cùng tham số. Sau đó, thay vì tạo nhiều lớp ẩn, nó sẽ tạo một lớp và lặp lại nhiều lần theo yêu cầu.

***- Ưu điểm của RNN:***

+ Khả năng xử lý các chuỗi có độ dài thay đổi: RNN được thiết kế để xử lý các chuỗi đầu vào có độ dài thay đổi, khiến chúng rất phù hợp cho các tác vụ như nhận dạng giọng nói, xử lý ngôn ngữ tự nhiên và phân tích chuỗi thời gian.

+ Bộ nhớ của các đầu vào trong quá khứ: RNN có bộ nhớ về các đầu vào trong quá khứ, cho phép chúng nắm bắt thông tin về ngữ cảnh của chuỗi đầu vào. Điều này làm cho chúng hữu ích cho các nhiệm vụ như mô hình hóa ngôn ngữ, trong đó ý nghĩa của một từ phụ thuộc vào ngữ cảnh mà từ đó xuất hiện.

+ Chia sẻ thông số: RNN chia sẻ cùng một bộ tham số trong tất cả các bước thời gian, giúp giảm số lượng tham số cần học và có thể dẫn đến khả năng khái quát hóa tốt hơn.

+ Ánh xạ phi tuyến tính: RNN sử dụng các hàm kích hoạt phi tuyến tính, cho phép chúng tìm hiểu các ánh xạ phức tạp, phi tuyến tính giữa đầu vào và đầu ra.

+ Xử lý tuần tự: RNN xử lý các chuỗi đầu vào một cách tuần tự, giúp chúng hoạt động hiệu quả về mặt tính toán và dễ dàng song song hóa.

+ Linh hoạt: RNN có thể được điều chỉnh cho nhiều loại tác vụ và loại đầu vào, bao gồm chuỗi văn bản, lời nói và hình ảnh.

+ Cải thiện độ chính xác: RNN đã được chứng minh là đạt được hiệu suất cao nhất trong nhiều nhiệm vụ mô hình hóa trình tự, bao gồm mô hình hóa ngôn ngữ, nhận dạng giọng nói và dịch máy.

*-* ***Nhược điểm của RNN:***

+ Biến mất hoặc bùng nổ (Vaninshing or Exploding): RNN có thể gặp vấn đề biến mất hoặc bùng nổ độ dốc (gradient degree), điều này có thể gây khó khăn cho việc huấn luyện mạng một cách hiệu quả. Điều này xảy ra khi độ dốc của hàm mất mát đối với các tham số trở nên rất nhỏ hoặc rất lớn khi chúng lan truyền theo thời gian.

+ Độ phức tạp tính toán: RNN có thể tốn kém về mặt tính toán để đào tạo, đặc biệt là khi xử lý các chuỗi dài. Điều này là do mạng phải xử lý từng đầu vào theo trình tự, điều này có thể chậm.

+ Khó khăn trong việc nắm bắt sự phụ thuộc dài hạn: Mặc dù RNN được thiết kế để nắm bắt thông tin về các đầu vào trong quá khứ, nhưng chúng có thể gặp khó khăn trong việc nắm bắt các phụ thuộc dài hạn trong chuỗi đầu vào. Điều này là do độ dốc có thể trở nên rất nhỏ khi chúng lan truyền theo thời gian, điều này có thể khiến mạng quên thông tin quan trọng.

+ Thiếu song song: RNN vốn đã tuần tự, điều này gây khó khăn cho việc tính toán song song. Điều này có thể hạn chế tốc độ và khả năng mở rộng của mạng.

+ Khó khăn trong việc lựa chọn kiến ​​trúc phù hợp: Có nhiều biến thể khác nhau của RNN, mỗi biến thể đều có những ưu điểm và nhược điểm riêng. Việc chọn kiến ​​trúc phù hợp cho một nhiệm vụ nhất định có thể là một thách thức và có thể yêu cầu thử nghiệm và điều chỉnh rộng rãi.

+ Khó khăn trong việc giải thích đầu ra: Đầu ra của RNN có thể khó diễn giải, đặc biệt là khi xử lý các đầu vào phức tạp như ngôn ngữ tự nhiên hoặc âm thanh. Điều này có thể gây khó khăn cho việc hiểu cách mạng đưa ra dự đoán.

1. **Chuỗi thời gian cấu trúc Bayesian(Bayesian structural time series)**

***Ra đời:*** Công trình của West và Harrison (1997) đã sử dụng phương pháp mô hình tuyến tính dựa trên lý thuyết xác suất Bayesian cho chuỗi thời gian, và đây có thể coi là một bước tiến đầu tiên trong phát triển BSTS.

Một trong những công trình quan trọng nhất sau này là bài báo của Scott và Varian có tiêu đề "Predicting the Present with Bayesian Structural Time Series" (2014). Trong bài báo này, họ giới thiệu BSTS và mô tả cách tiếp cận và ứng dụng BSTS trong mô hình hóa chuỗi thời gian.

Mô hình chuỗi thời gian cấu trúc là một khung thống kê được thiết kế độc đáo để mô hình hóa dữ liệu chuỗi thời gian. Mô hình này phân tách một chuỗi thời gian thành nhiều thành phần khác nhau, bao gồm tính thời vụ, xu hướng và các thành phần chu kỳ bằng cách sử dụng phương pháp tiếp cận không gian trạng thái. Với việc phân tách chuỗi thời gian, mô hình này cũng có thể thực hiện phép nội suy cấu trúc trong đó chuỗi hàng quý có thể được nội suy thành chuỗi hàng tháng bằng cách áp dụng phương pháp tổng của khoảng thời gian, trung bình của khoảng thời gian hoặc cuối khoảng thời gian. Ý nghĩa thống kê của mô hình này là vô số và có vai trò chính trong lĩnh vực khoa học dữ liệu.

**Bayesian Structural Time Series** là một phương pháp dự báo chuỗi thời gian sử dụng các mô hình thời gian có cấu trúc dựa trên phân tích Bayesian. BSTS cho phép mô hình hóa các yếu tố thay đổi trong chuỗi thời gian, bao gồm các yếu tố mùa vụ, xu hướng, thành phần ngẫu nhiên và các sự kiện đặc biệt.

BSTS sử dụng sự kết hợp của các mô hình tuyến tính và phi tuyến để phân tích và dự báo chuỗi thời gian. Các mô hình này cho phép mô hình hóa các yếu tố tác động lẫn nhau và tìm hiểu cách chúng ảnh hưởng đến chuỗi thời gian.

Đối với các tập dữ liệu giá trị của tiền ảo, nhìn chung các tập dữ liệu không có một xu hướng tăng hay giảm nhất định, cũng không có tính mùa vụ, do đó nhóm thực hiện chọn thành phần cho mô hình gồm:

+ Autoregressive component:

* + - Đóng vai trò dự đoán giá trị dựa trên lịch sử của dữ liệu đầu vào, giúp nắm bắt các xu hướng mẫu lặp lại trong dữ liệu và sử dụng để dự đoán giá trị tiếp theo.
    - Xây dựng một mô hình hồi quy tuyến tính hoặc phi tuyến để dự đoán giá trị dựa trên giá trị trước đó; thường được biểu diễn dưới dạng một chuỗi Autoregressive hoặc một mô hình hồi quy tuyến tính.

A picture containing font, text, typography

Description automatically generated

trong đó:

+ c là hằng số

+ φ1 , …, φp là các tham số

+ εt là nhiễu trắng

+ p : số lượng giá trị trong quá khứ có trong mô hình AR

+ Dynamic:

* + - Đóng vai trò trong việc mô hình hóa sự biến đổi không xác định và không thể dự đoán trước của dữ liệu. Thành phần này cho phép mô hình phản ứng linh hoạt và thích ứng với sự thay đổi trong dữ liệu đầu vào.
    - Xây dựng một mô hình động (dynamic model) để mô tả sự biến đổi và dao động của dữ liệu. Thành phần này thường được biểu diễn bằng một mô hình ngẫu nhiên, chẳng hạn như quy trình Gauss-Markov (Gaussian-Markov process) hoặc quy trình ARIMA (Autoregressive Integrated Moving Average), để mô phỏng các yếu tố không xác định trong dữ liệu.
    - Cho phép mô hình BSTS cập nhật và thích ứng với sự thay đổi trong dữ liệu theo thời gian, giúp mô hình có khả năng dự đoán và ước lượng chính xác hơn. Nó giúp mô hình đánh giá và điều chỉnh mức độ biến động và sự không chắc chắn trong dữ liệu, từ đó cung cấp dự đoán chính xác hơn về giá trị tiếp theo.

Từ đó ta có công thức chung:

**Y(t) = dynamic\_component(t) + autoreg\_component(t)**

1. **Fully Convolutional Network**

**FCN (Fully Convolutional Network**) là một kiến trúc dựa trên mạng CNN nhằm mục đích segment một hình ảnh đầu vào bằng cách dự báo nhãn cho từng pixel trên ảnh input.

FCN chủ yếu được sử dụng trong mục đích xử lý hình ảnh, tuy nhiên FCN cũng được dùng trong bài toán time series bằng cách sử dụng tích chập 1D.

Và ở bài viết này, chúng ta sẽ tìm hiểu FCN cho bài toán Time series.

**Fully Convolutional Network (FCN)** trong bài toán Time series là một thuật toán học sâu tận dụng mạng nơ-ron tích chập CNN cho kết nối end-to-end phân loại chuỗi thời gian đơn biến.

Ảnh có chứa văn bản, ảnh chụp màn hình, thiết kế

Mô tả được tạo tự động

Hình 19 Mô tả kích thước của hình ảnh

Kích thước của đầu vào hình ảnh trong mạng tích chập hoàn toàn (FCN) là H × W. Sau mỗi lần gộp, kích thước của bản đồ đặc trưng sẽ trở nên nhỏ hơn tương ứng. Việc chuyển đổi các lớp được kết nối đầy đủ thành các lớp tích chập cho phép mạng phân loại tạo ra bản đồ nhiệt H × W. Việc thêm các lớp giải mã sẽ tạo ra một mạng hiệu quả cho việc học dày đặc từ đầu đến cuối

Thành phần cốt lõi của FCN là lớp tích chập, chịu trách nhiệm chính cho việc học tính năng. Nó chứa một số bản đồ tính năng được xử lý bởi các hạt tích chập. Mỗi hạt tích chập chỉ xử lý dữ liệu cho trường tiếp nhận của nó với cùng trọng số được chia sẻ, do đó giảm số lượng tham số miễn phí và cho phép FCN sâu hơn với ít tham số hơn. Công thức tính các lớp tích chập như sau:

Ảnh có chứa văn bản, Phông chữ, biểu đồ, màu trắng

Mô tả được tạo tự động

Trong đó:

+ 𝑐𝑜𝑛𝑣(𝑖,𝑗) là kết quả tích chập hay còn gọi là feature map;

+ M cho biết kích thước của convolution kernel (M × M);

+ 𝑤𝑢,𝑣 là trọng số của convolution kernel tại dòng 𝑢 và cột 𝑣

+ 𝑥𝑖+𝑢,𝑗+𝑣 là đầu vào

+ 𝑏 là độ lệch

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

Mô tả được tạo tự động

Hình 20 Kiến trúc đặc trưng của FCN

Kiến trúc đặc trưng của FCN. Một khối tích chập chứa M lớp tích chập liên tục và p lớp tổng hợp. Trong một mạng chập, N khối chập liên tục có thể được xếp chồng lên nhau, theo sau là q lớp chập hoàn toàn và s lớp giải chập.

1. **Phân tích hồi quy quá trình Gaussian (Gaussian Process Regression)**

**Phân tích hồi quy quá trình Gaussian (GPR)** là một phương pháp máy học phi tham số (non-parametric), không cần tham số để tạo mô hình dự đoán. Dùng để mô hình hóa và dự đoán quan hệ giữa các biến đầu vào và đầu ra dựa trên quá trình Gaussian(GP).

Quá trình Gaussian(GP) là một tập hợp các biến ngẫu nhiên tuân theo phân phối Gaussian, được đặc trưng bởi vectơ trung bình và ma trận hiệp phương sai.

Công thức:

Diagram

Description automatically generated with medium confidence

Trong đó:

x và x' là hai vector đầu vào

||x - x'||² là khoảng cách giữa hai vector

σ là độ rộng của kernel

Hàm hiệp phương sai này có hai siêu tham số là θ=(,l). Để đánh giá các siêu tham số ta suy diễn sử dụngcông thức xác suất Bayes như sau:



trong đó:

p(f|Y,X) gọi là xác suất hậu nghiệm (posterior)

p(Y|X,f) gọi là xác suất khả năng (likelihood)

p(f|X) gọilà xác suất tiên nghiệm

p(Y|X) gọi là xác suất biên(marginal likelihood). Các siêu tham số hàm hiệp phương sai tìm được sao cho hàm logarit của xác suất biên sau đây đạt giá trị lớn nhất.

A picture containing font, text, line, white

Description automatically generated

Phương pháp được sử dụng để tìm điểm tối ưu cho hàm logarit xác suất biên ở trên là phương pháp Gradient liên hợp. Sau khi tìm được các siêu tham số cho hàm hiệp phương sai, ma trận hiệp phương sai Khoàn toàn xác định. Xác suất có điều kiện p(f∗ | f) mang ý nghĩa là, đối với bộ dữ liệu huấn luyện tại các điểm f, việc dự đoán tại các điểm dữ liệu kiểm thử f∗sẽ cho độ chính xác với xác suất bao nhiêu. Phân phối của xác suất có điều kiện p(f∗ | f)cũng là phân phối quá trình Gauss có dạng:

A group of letters on a white background

Description automatically generated with low confidence

Trong đó:

A picture containing text, font, line, white

Description automatically generated

Đối với biến đổi lấysai phân bậc một ta có:



Trong đó:

: là đầu vào cho phương pháp

: là kết quả dự đoán chuỗi  tại một thời điểm kế tiếp

: là kết quả dự đoán tại một điểm kế tiếp của chuỗi đầu vào 

Công thức trên cho phép truy ngược kết quả dự đoán chuỗi xu thế  từ kết quả dự đoán biến đổi xu thế .

1. **Autoregressive Integrated Moving Average (ARIMA)**

Mô hình ARIMA (AutoRegressive Integrated Moving Average) là một trong những mô hình dự báo thời gian cơ bản, sử dụng các giá trị trong quá khứ và các lỗi dự báo của một chuỗi thời gian để dự đoán giá trị trong tương lai.

*\*Chuỗi thời gian: một chuỗi các điểm dữ liệu được thu thập theo thứ tự thời gian.*

ARIMA tích hợp của 2 quá trình: quá trình tự hồi quy bậc p – AR(p) và quá trình trung bình trượt bậc q – MA(q). Mặt khác, cần phải dùng tích hợp sai phân I(d) (hay còn gọi là toán tử trễ) để làm cho chuỗi thời gian trở thành chuỗi dừng.

*\*Chuỗi dừng: một chuỗi thời gian có tính dừng là một chuỗi có các giá trị như trung bình, phương sai và các giá trị tương quan (autocorrelation) của quá trình không thay đổi theo thời gian và không bao hàm yếu tố xu thế.*

*\* Để kiểm tra tính dừng của chuỗi thời gian, ta có hai phương pháp*

*kiểm định phổ biến: Kiểm định Dickey Fuller3 (DF) và Dickey Fuller cải tiến (ADF4).*

+ Quá trình tự hồi quy bậc p – AR(p): quá trình tìm mối quan hệ giữa dữ liệu hiện tại và p dữ liệu trước đó (lag).

+ Quá trình trung bình vượt bậc q – MA(q): quá trình tìm mối quan hệ giữa dữ liệu hiện tại và d phần lỗi trước đó.

+ Tích hợp sai phân I(d): so sánh sự khác nhau giữa d quan sát (Hiệu giữa

giá trị hiện tại và d giá trị trước đó) để biến một chuỗi thành chuỗi dừng.

• p : số lượng giá trị trong quá khứ có trong mô hình AR

A picture containing font, text, typography

Description automatically generated

trong đó:

+ c là hằng số

+ φ1 , …, φp là các tham số

+ εt là nhiễu trắng

• d : số lần chuỗi thời gian bị sai khác.

• q : số lỗi dự báo trong quá khứ có trong mô hình MA hoặc kích thước của cửa sổ trung bình động. Nó được đặt tên là mô hình MA vì chỉ có có thể được coi là trung bình động có trọng số của các lỗi dự báo trong quá khứ

Phương trình mô hình đầy đủ của ARIMA(p, d, q) là:



trong đó ∇yt là chuỗi thời gian khác biệt, có thể chênh lệch nhiều lần

- Mô hình ARIMA không phải là mô hình dự báo hoàn hảo ứng với bất kỳ dữ liệu chuỗi thời gian nào và chỉ hoạt động tốt nhất nếu dữ liệu phụ thuộc nhiều vào thời gian và dự báo dạng điểm thời gian ~ những dữ liệu ngẫu nhiên thường ít hoạt động với mô hình ARIMA.

1. **Gated Recurrent Unit (GRU)**

Gated recurrent units (GRUs) giống như một bộ nhớ dài hạn ngắn hạn (LSTM) với một cổng quên, nhưng có ít tham số hơn LSTM, vì nó thiếu một cổng đầu ra. Hiệu suất của GRU đối với một số tác vụ nhất định về mô hình hóa âm nhạc đa âm, mô hình tín hiệu giọng nói và xử lý ngôn ngữ tự nhiên được cho là tương tự như hiệu suất của LSTM.

**Reset gate và update gate model:**

Cổng thiết lập lại kiểm soát bao nhiêu trạng thái trước đó mà chúng ta có thể vẫn muốn ghi nhớ. Tương tự như vậy, một cổng cập nhật sẽ cho phép kiểm soát bao nhiêu trạng thái mới chỉ là một bản sao của trạng thái cũ.

Về mặt toán học, trong một bước thời gian t nhất định, giả sử rằng đầu vào là một lô nhỏ (số ví dụ: n, số lượng đầu vào: d) và trạng thái ẩn của bước thời gian trước đó là (số đơn vị ẩn: h). Sau đó đặt lại cổng và cập nhật cổng được tính như sau:

,

,

Trong đó:  and , là thông số trọng lượng và  là các tham số sai lệch.

**Candidate Hidden State model:**

tích hợp cổng đặt lại, dẫn đến trạng thái ẩn ứng cử viên sautại bước thời gian ***t:***

Trong đó và là các tham số trọng lượng,

là độ lệch và ký hiệu là toán tử sản phẩm (theo phần tử) Hadamard. Ở đây chúng em sử dụng chức năng kích hoạt tanh.

**Hidden state model:**

kết hợp tác dụng của cổng cập nhật. Điều này xác định mức độ mà trạng thái ẩn mới phù hợp với trạng thái cũ so với mức độ nó giống với trạng thái ứng cử viên mới . Cổng cập nhật có thể được sử dụng cho mục đích này, đơn giản bằng cách sử dụng các tổ hợp lồi thông minh về phần tử của Điều này dẫn đến phương trình cập nhật cuối cùng cho GRU:

)

Tóm lại, GRU có hai đặc điểm phân biệt sau:

* Đặt lại cổng giúp nắm bắt các phụ thuộc ngắn hạn theo trình tự.
* Cổng cập nhật giúp nắm bắt các phụ thuộc dài hạn theo trình tự.

**Chương IV: Phương pháp**

1. **Giới thiệu**

Đầu vào là dữ liệu lịch sử chuỗi thời gian của tiền điện tử. Sau khi dữ liệu trải qua quá trình tiền xử lý, phân tách dữ liệu theo tập dữ liệu đào tạo, xác thực và thử nghiệm tương ứng, sau đó sẽ sử dụng các mô hình Hồi quy tuyến tính, ARIMA, GRUs, RNN, Fuzzy Time Series, BSTS, FCN, Holt-Winters Exponential Smoothing, GPR để xe lửa. Mô hình đã chọn được đào tạo bằng cách sử dụng tập dữ liệu đã chuẩn bị.

Trong quá trình đào tạo, mô hình học các mẫu, mối quan hệ và cấu trúc cơ bản trong dữ liệu. Quá trình đào tạo liên quan đến việc điều chỉnh các tham số của mô hình để giảm thiểu sự khác biệt giữa đầu ra dự đoán và đầu ra thực tế từ dữ liệu đào tạo. Sau khi mô hình được đào tạo, nó cần được đánh giá để đánh giá hiệu suất và khả năng khái quát hóa của nó.

Nếu sau khi đánh giá hiệu suất mô hình không hiệu quả thì quay lại hiệu chỉnh các siêu tham số rồi thực hiện huấn luyện lại.

1. **Các bước thực hiện**

- Bước 1: Thu thập dữ liệu chuỗi thời gian: dữ liệu có thuộc tính ngày tháng năm và các giá trị.

- Bước 2: Chia dữ liệu thành 3 phần để huấn luyện thực hiện: huấn luyện dữ liệu, kiểm tra dữ liệu, xác nhận dữ liệu với các tỉ lệ khác nhau 7:2:1, 6:2:2, 5:3:2

- Bước 3: Chọn một mô hình, xây dựng, hiệu chỉnh các siêu tham số để sẵn sàng huấn luyện.

- Bước 4: Huấn luyện mô hình.

- Bước 5: Dự đoán dựa trên mô hình.

- Bước 6: Nếu kết quả trả về là các giá trị lỗi thì quay lại bước 3, ngược lại thực hiện ước tính hiệu suất và lưu lại.

- Bước 7: Chọn model khác và quay lại bước 2.

A picture containing diagram, sketch, line, plan

Description automatically generated

Hình 21 Các bước thực hiện

**Chương V: Kết quả dự đoán**

1. **Kết quả các độ đo:**

***Các độ đo sử dụng:***

+ Mean Absolute Percentage Error (MAPE):

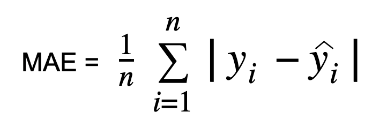
* MAPE đo lường độ lỗi trung bình giữa dự đoán và giá trị thực tế theo tỷ lệ phần trăm tuyệt đối.
* Giá trị MAPE càng nhỏ, tức là dự đoán càng gần với giá trị thực tế.
* MAPE phù hợp khi các quan sát có giá trị thực tế gần nhau.

Ảnh có chứa Phông chữ, màu trắng, biểu đồ, thuật in máy

Mô tả được tạo tự động

+ Mean Absolute Error (MAE):

* MAE đo lường độ lỗi trung bình giữa dự đoán và giá trị thực tế dưới dạng giá trị tuyệt đối.
* Giá trị MAE càng nhỏ, tức là dự đoán càng gần với giá trị thực tế.
* MAE phù hợp khi các quan sát có phân bố không đối xứng hoặc chứa các giá trị ngoại lệ (outliers).



+ Root Mean Square Error (RMSE):

* RMSE đo lường độ lỗi trung bình giữa dự đoán và giá trị thực tế dưới dạng độ lỗi bình phương.
* RMSE lớn hơn MAE vì nó tính toán lỗi theo đơn vị gốc.
* Giá trị RMSE càng nhỏ, tức là dự đoán càng gần với giá trị thực tế.
* RMSE phù hợp khi các quan sát có phân bố gần với phân phối chuẩn.

Ảnh có chứa Phông chữ, văn bản, màu trắng, biểu đồ

Mô tả được tạo tự động

Trong đó:

* n: kích thước mẫu tập dữ liệu
* yi: giá trị thực tế tại thời điểm t
* : giá trị trung bình tại thời điểm t
* : giá trị dự đoán tại thời điểm t

- Độ chính xác:

a) ETH:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | 7 : 2 : 1 | | | | 6:2:2 | | | | 5:3:2 | | |
| **RMSE** | **MAE** | **MAPE** | **RMSE** | | **MAE** | **MAPE** | **RMSE** | | **MAE** | MAPE |
| LN | 2084.42 | 2079.87 | 0.5543 | 620.71 | | 584.92 | 0.276 | 1167.06 | | 1140.04 | 2.8163 |
| ARIMA | 2285.98 | 2275.48 | 0.5753 | 677.18 | | 626.21 | 0.2886 | 2270.49 | | 1901.25 | 0.5056 |
| HW | 67.968 | 4619.62 | 0.734 | 39.503 | | 1560.5 | 0.504 | 28.481 | | 811.153 | 0.344 |
| FTS | 167.153 | 133.191 | 0.0774 | 143.76 | | 143.76 | 0.0749 | 862.77 | | 823.36 | 1.1431 |
| RNN | 1875.841 | 1873.635 | 0.9998 | 1582.6 | | 1561.74 | 0.9998 | 1571.67 | | 1552.02 | 0.9998 |
| BSTS | 1447.67 | 1411.09 | 7.2837 | 1357.4 | | 1325.7 | 7.6214 | 1357.37 | | 1325.8 | 7.6124 |
| GRU | 1855.403 | 1853.44 | 0.9997 | 1597.4 | | 1576.71 | 0.9998 | 1611.92 | | 1590.79 | 0.9998 |
| GPR | 1137.587 | 987.089 | 3.3778 | 1171.07 | | 1010.01 | 3.3013 | 1171.07 | | 1010.01 | 3.3013 |
| FCN | 1650.429 | 1639.3 | 0.9998 | 1522.5 | | 1504.68 | 0.9998 | 1304.35 | | 1292.19 | 0.9998 |

Bảng 1 Bảng độ chính xác dự đoán tập dữ liệu ETH

b) BNB:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | 7 : 2 : 1 | | | | 6:2:2 | | | | 5:3:2 | | |
| **RMSE** | **MAE** | **MAPE** | **RMSE** | | **MAE** | **MAPE** | **RMSE** | | **MAE** | MAPE |
| LN | 216.89 | 214.57 | 0.42 | 33.61 | | 24.96 | 0.08 | 253.6 | | 251.83 | 6.8 |
| ARIMA | 229.43 | 227.6 | 0.4317 | 56.656 | | 48.011 | 0.143 | 259.76 | | 257.95 | 8.337 |
| HW | 21.092 | 444.89 | 0.597 | 18.093 | | 327.35 | 0.53 | 15.685 | | 246.03 | 5.738 |
| FTS | 31.879 | 25.84 | 0.07 | 29.31 | | 23.89 | 0.0778 | 248.08 | | 246.18 | 5.766 |
| RNN | 318.591 | 317.32 | 0.999 | 313.12 | | 311.83 | 0.999 | 287.54 | | 286.436 | 0.999 |
| BSTS | 291.01 | 289.54 | 31.803 | 276.97 | | 275.26 | 26.71 | 276.97 | | 275.26 | 26.71 |
| GRU | 305.95 | 304.69 | 0.999 | 305.824 | | 304.62 | 0.999 | 293.1 | | 291.89 | 0.999 |
| GPR | 214.73 | 183.37 | 9.64 | 204.93 | | 175.233 | 9.275 | 204.93 | | 175.23 | 9.275 |
| FCN | 301.55 | 300.58 | 0.999 | 285.85 | | 284.63 | 0.999 | 148.34 | | 147.88 | 0.997 |

Bảng 2 Bảng độ chính xác dự đoán tập dữ liệu BNB

c) XRP:

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | 7 : 2 : 1 | | | | 6:2:2 | | | | 5:3:2 | | |
| **RMSE** | **MAE** | **MAPE** | **RMSE** | | **MAE** | **MAPE** | **RMSE** | | **MAE** | MAPE |
| LN | 0.571 | 0.569 | 0.57 | 0.249 | | 0.244 | 0.373 | 0.336 | | 0.327 | 4.344 |
| ARIMA | 0.36 | 0.355 | 0.45 | 0.4904 | | 0.487 | 0.544 | 0.12 | | 0.105 | 0.205 |
| HW | 0.705 | 0.497 | 0.536 | 0.682 | | 0.465 | 0.531 | 0.459 | | 0.211 | 0.34 |
| FTS | 0.116 | 0.098 | 0.188 | 0.1061 | | 0.0822 | 0.16 | 0.032 | | 0.028 | 0.063 |
| RNN | 0.2791 | 0.278 | 0.578 | 0.2611 | | 0.259 | 0.605 | 0.252 | | 0.251 | 0.597 |
| BSTS | 0.0982 | 0.081 | 0.227 | 0.0995 | | 0.084 | 0.245 | 0.0995 | | 0.084 | 0.245 |
| GRU | 0.30834 | 0.306 | 0.6013 | 0.28 | | 0.272 | 0.617 | 0.265 | | 0.257 | 0.604 |
| GPR | 0.26122 | 0.176 | 0.374 | 0.3142 | | 0.215 | 0.459 | 0.314 | | 0.215 | 0.459 |
| FCN | 0.255 | 0.254 | 0.596 | 0.242 | | 0.24 | 0.6 | 0.25 | | 0.248 | 0.61 |

Bảng 3 Bảng độ chính xác dự đoán tập dữ liệu XRP

1. **Rút ra kết luận thực nghiệm**

Dựa vào kết quả đánh giá trên, có thể thấy với mỗi tập dữ liệu sẽ có mô hình riêng và tỷ lệ dữ liệu phân chia phù hợp. Để có hiệu suất dự đoán tốt nhất, nhóm thực hiện dự đoán giá của từng bộ dữ liệu trong 30 ngày tới dựa trên các mô hình và tỉ lệ đã chọn. Các mô hình được chọn sẽ có giá trị MAE và MAPE và MAE nhỏ.

1. **ETH**

Đối với tập dữ liệu ETH: hai mô hình tốt nhất là .

* FTS với tỉ lệ Train/Test/Validate as 7/2/1, MAE là 133.1906, RMSE là 167.1525 và MAPE là 0.0774
* FCN với tỉ lệ Train/Test/Validate as 6/2/2, RMSE là 1522.477, MAE là 1504.679 và MAPE là 0.9998

Vì vậy, chúng em sẽ chọn 2 mô hình này để dự báo giá “Close” trong 30 ngày tới.

1. **BNB**

Đối với tập dữ liệu BNB: hai mô hình tốt nhất là .

• FTS với tỷ lệ phân tách của Train/Test/Validate là 7/2/1, giá trị MAE là 28,8333, RMSE là 318791 và MAPE là 0,07.

• FCN với tỷ lệ phân chia của Train/Test/Validate là 6/2/2, giá trị RMSE là 301,55, MAE là 300,58 và MAPE là 0,999.

Vì vậy, chúng em sẽ sử dụng 2 mô hình này để dự báo giá “Đóng” trong 30 ngày tới.

1. **XRP**

Với tập dữ liệu XRP: hai mô hình tốt nhất là FTS và FCN.

* FTS với tỉ lệ Train/Test/Validate là 6/2/2, MAE là 0.0822, RMSE là 0.1061 và MAPE là 0.1603
* GRU với tỉ lệ Train/Test/Validate là 5/3/2, RMSE là 0.265, MAE là 0.257 và MAPE là 0.6035

Vì vậy, chúng em sẽ chọn 2 mô hình này để dự báo giá “Close” trong 30 ngày tới.

1. **Trực quan hóa dự báo:**

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

Mô tả được tạo tự động

Hình 22Kết quả của mô hình FTS với dataset chia theo 7/2/1

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

Mô tả được tạo tự động

Hình 23. Kết quả của mô hình FCN với dataset chia theo 6/2/2

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 24 kết quả của mô hình FTS với dataset chia theo 6/2/2

A picture containing text, screenshot, plot, font

Description automatically generated

Hình 25 kết quả của mô hình GRU với dataset chia theo 5/3/2

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 26 Kết quả của mô hình FCN với dataset BNB chia theo 7/2/1

A picture containing text, screenshot, plot, diagram

Description automatically generated

Hình 27 Kết quả của mô hình FTS với dataset BNB chia theo 7/2/1

**Chương VI: Tổng kết**

- Trong quá trình thực hiện đề tài nghiên cứu, bên cạnh những kết quả đạt được, nhóm thực hiện gặp một số khó khăn, thách thức:

• Khó khăn trong việc xây dựng các mô hình dự đoán: Các mô hình dự đoán tiền điện tử thường phức tạp và đòi hỏi kiến ​​thức chuyên sâu về lĩnh vực này. Nhóm đã xem xét nhiều loại tài liệu bao gồm các bài báo khoa học và đưa ra các quyết định quan trọng trong việc lựa chọn và xây dựng các mô hình dự đoán.

• Khó khăn trong việc đánh giá hiệu quả của mô hình: Để đánh giá hiệu quả của các mô hình dự đoán, nhóm đã sử dụng một số thước đo đánh giá khác nhau nhưng kết quả cho thấy độ chính xác của các mô hình vẫn chưa cao.

- Trong tương lai, nhóm thực hiện sẽ cố gắng giải quyết những thách thức trên và cung cấp các giải pháp tốt hơn để dự đoán giá tiền điện tử bằng cách:

• Nâng cao kỹ năng chọn lọc và xử lý dữ liệu: tiếp tục nghiên cứu và áp dụng các phương pháp tiên tiến nhất trong việc chọn lọc và xử lý dữ liệu để đảm bảo tính khả thi và chính xác của các mô hình dự đoán.

• Kết hợp các mô hình: tiếp tục nghiên cứu và kết hợp các mô hình với nhau để nâng cao hiệu suất.

• Tăng cường hợp tác và chia sẻ kinh nghiệm: Tiếp tục tìm kiếm tài liệu, công trình nghiên cứu trên các diễn đàn học thuật và tạp chí khoa học.

Với các giải pháp trên, tin rằng nhóm có thể cải thiện hiệu quả và độ chính xác của các mô hình dự đoán giá tiền điện tử trong tương lai

# **References**

|  |  |
| --- | --- |
| [1] | M. Poongodi, Ashutosh Sharma, V. Vijayakumar, Vaibhav Bhardwaj, Abhinav Parkash Sharma, Razi Iqbal and Rajiv Kumar,, "Prediction of the price of Ethereum blockchain cryptocurrency in an industrial finance system," in *International journal Computers & Electrical Engineering*, 2020. |
| [2] | Siti Saadah, A.A Ahmad Whafa, "Monitoring Financial Stability Based on Prediction of Cryptocurrencies Price Using Intelligent Algorithm," in *2020 International Conference on Data Science and Its Applications (ICoDSA)*, Bandung, Indonesia, 2020. |
| [3] | V. Derbentseva, V. Babenko, K. Khrustalev, H. Obruch, S. Khrustalova, "Comparative Performance of Machine Learning Ensemble Algorithms for Forecasting Cryptocurrency Prices," *International Journal of Engineering,* vol. 34, no. 01, pp. 140-148, 2021. |
| [4] | Kiki Ramadani, Dodi Devianto, "The forecasting model of Bitcoin price with fuzzy time series Markov chain and chen logical method," *NTERNATIONAL CONFERENCE ON SCIENCE AND APPLIED SCIENCE (ICSAS2020),* vol. 2296, no. 1, 2020. |
| [5] | Mohammad J. Hamayel, Amani Yousef Owda, "A Novel Cryptocurrency Price Prediction Model Using GRU, LSTM and bi-LSTM Machine Learning Algorithms," vol. 2, no. 4, pp. 477-496, 2021. |
| [6] | Dian Utami Sutiksno, Ansari Saleh Ahmar, Nuning Kurniasih, Eko Susanto, Audrey Leiwakabessy, "Forecasting Historical Data of Bitcoin using ARIMA and α-Sutte Indicator," in *2nd International Conference on Statistics, Mathematics, Teaching, and Research*, 2018. |
| [7] | L. T. Lương, "MÔ HÌNH DỰ BÁO CHUỖI THỜI GIAN MỜ SỬ DỤNG KỸ THUẬT PHÂN CỤM DỰA TRÊN ĐỒ THỊ," *Tạp chí Khoa học và Công nghệ Thái Nguyên,* vol. 226, no. 11, pp. 177-184, 2021. |
| [8] | S.-M. Chen, "“Forecasting enrollments based on fuzzy time series,”," *Fuzzy Sets and System,* no. 81, pp. 311-319, 1996. |
| [9] | Qiang Song, Brad S. Chissom, "Forecasting Enrollments with Fuzzy Time Series part 1," in *The Anuual Meeting of the Mid-South Educational Reseach Association,*, Lexington, KY, 1991. |
| [10] | Jizhong Wu, Bo Liu, Hao Zhang ,Shumei He, Qianqian Yang, "Fault Detection Based on Fully Convolutional Networks (FCN)," *Journal of Marine Science and Engineering,* vol. 9, no. 3, p. 259, 2021. |
| [11] | Wikipedia, "Exponent Smoothing," [Online]. Available: https://en.wikipedia.org/wiki/Exponential\_smoothing. [Accessed 10 06 2023]. |
| [12] | Simplilearn, "Recurrent Neural Network(RNN) Tutorial: Types, Examples, LSTM and More," [Online]. Available: https://www.simplilearn.com/tutorials/deep-learning-tutorial/rnn. [Accessed 10 06 2023]. |
| [13] | SAP, "Triple Exponent Smoothing," [Online]. Available: https://bom.so/2zt4EL. [Accessed 10 06 2023]. |

[14] “Autoregressive Integrated Moving Average (ARIMA) Prediction Model,” *Investopedia*. https://www.investopedia.com/terms/a/autoregressive-integrated-moving-average-arima.asp (accessed Jun. 21, 2023).

[15] “Gated recurrent unit,” *Wikipedia*. Jun. 04, 2023. Accessed: Jun. 21, 2023. [Online]. Available: <https://en.wikipedia.org/w/index.php?title=Gated_recurrent_unit&oldid=1158552114#cite_note-Ravalli2018-4>

[16] “10.2. Gated Recurrent Units (GRU) — Dive into Deep Learning 1.0.0-beta0 documentation.” https://d2l.ai/chapter\_recurrent-modern/gru.html (accessed Jun. 21, 2023).

[17] C. E. Rasmussen and C. K. Williams, “Gaussian processesfor machine learning. 2006,” The MIT Press, Cambridge,MA, USA, vol. 38, pp. 715–719, 2006.

[18] “(PDF) Dự đoán xu thế chỉ số chứng khoán Việt Nam VN-Index sử dụng phân tích hồi quy Gaussian Process và mô hình tự hồi quy trung bình động ARMA.” https://www.researchgate.net/publication/329123511\_Du\_doan\_xu\_the\_chi\_so\_chung\_khoan\_Viet\_Nam\_VN-Index\_su\_dung\_phan\_tich\_hoi\_quy\_Gaussian\_Process\_va\_mo\_hinh\_tu\_hoi\_quy\_trung\_binh\_dong\_ARMA (accessed Jun. 21, 2023).