

Áp dụng hồi quy Ridge và mạng nơon nhân tạo để dự báo giá ICO sau sáu tháng

Applying ridge regression and ANN to predict ICO price after six months

Trần Kim Toại^{1*}, Võ Thị Xuân Hạnh¹, Võ Minh Huân¹

¹Trường Đại học Sư phạm Kỹ thuật Thành phố Hồ Chí Minh, Thành phố Hồ Chí Minh, Việt Nam

*Tác giả liên hệ, Email: toaitk@hcmute.edu.vn

THÔNG TIN	TÓM TẮT
<p>DOI:10.46223/HCMCOUJS.econ.vi.18.4.2104.2023</p>	<p>ICO (Initial Coin Offering) là phương thức huy động vốn đầu tư cho dự án tiền số. Những nhà đầu tư mua đồng tiền này ở thời điểm chưa phát hành với mức giá cực kỳ rẻ. Sau đó, các đồng tiền này được niêm yết lên sàn giao dịch, giá của chúng sẽ tăng lên cực kỳ nhanh nếu đồng tiền này tốt. Đánh giá ICO sau sáu tháng phát hành là khoảng thời gian nhà đầu tư mong đợi thu được lợi nhuận. Tập dataset gồm 109 ICO được xây dựng từ các website uy tín sau bước tiền xử lý dữ liệu. Phân tích tương quan giữa 12 đầu vào cho thấy dữ liệu gặp vấn đề đa cộng tuyến, điều này dẫn đến làm sai lệch kết quả của mô hình hồi quy bội. Hiện tượng quá khớp xảy ra khi sử dụng mô hình hồi quy bội. Để khắc phục những hạn chế của mô hình hồi quy bội, phương pháp hồi quy Ridge giải quyết được các vấn đề dữ liệu ICO này. Mô hình mạng neuron nhân tạo giải quyết mối quan hệ phi tuyến phức tạp giữa đầu vào và giá ICO. Bằng cách hiệu chỉnh tham số để đạt được hiệu năng tốt nhất theo ba thước đo hiệu năng Root Mean Square Error, Rsquares và Mean Absolute Error, kết quả chỉ ra thuật toán hồi quy ridge với tập kiểm tra với 3 ICO đạt được độ chính xác từ 63% tới 92% giá ICO, mô hình mạng neuron nhân tạo dự báo độ chính xác 98% giá trị thật tùy theo thước đo được sử dụng.</p>
<p>Ngày nhận: 27/11/2021</p>	
<p>Ngày nhận lại: 03/04/2022</p>	
<p>Duyệt đăng: 18/04/2022</p>	
<p>Mã phân loại JEL: G110; G170; G410</p>	<p>ABSTRACT</p>
<p><i>Từ khóa:</i> dự đoán ICO; hồi quy ridge; hồi quy tuyến tính; ICO; mạng neuron nhân tạo</p>	<p>ICO (Initial Coin Offering) is blockchain based crowdfunding venture to access capital. Investors buy this coin at an unissued time at an extremely cheap price. Then these coins are listed on the exchange, their price will go up extremely fast if the coin is good. Six months after ICO issuance is the period that investors expect to make a profit. The dataset of 109 ICOs is collected from reputable websites after data preprocessing. The correlation analysis between the 12 inputs shows that the data had multicollinearity problems, which led to skewed results of the multiple regression model. Overfitting occurs when using multiple regression models. To overcome the multiple regression model limitation, the Ridge regression method solves these ICO data problems of multiple regression models. The artificial neural</p>
<p><i>Keywords:</i> ICO prediction; ridge regression; linear regression; ICO; artificial neural network</p>	

network model solves the complex nonlinear relationships between inputs and ICO price. By tuning parameters to get the best performance according to three performance metrics Root Mean Square Error, Rsquares, and Mean Absolute Error, the ridge regression algorithm and artificial neural network achieved an accuracy of 63% to 92% and up to 98%, respectively in forecasting ICO prices with a test set of 3 ICOs that depend on the used metrics.

1. Giới thiệu

Một công ty muốn huy động vốn có thể phát hành đồng tiền điện tử riêng bằng cách tạo ra một lượng *token* như một dạng của cổ phiếu để bán ra thị trường nhằm thu hút các nhà đầu tư trong đợt mở bán đầu tiên. Đồng ICO ra đời nhằm giải quyết bài toán các công ty khởi nghiệp khi họ có ý tưởng kinh doanh tốt hoặc ý tưởng công nghệ đột phá nhưng gặp vấn đề thiếu vốn để triển khai các ý tưởng đi vào thực tế. Vì vậy họ tìm tới nhà đầu tư (Fisch, 2019; Panin, Kemell, & Hara, 2019).

Tuy nhiên, việc đảm bảo lợi nhuận của một ICO mang lại không phải tuyệt đối. Một ICO có thể thất bại hoặc thành công sau khi phát hành. Có những dự án ICO không tốt, lừa đảo vẫn xuất hiện trong cộng đồng, hoặc những dự án có lợi nhuận đầu tư không đạt kỳ vọng. Nhà đầu tư thường đợi từ 25 tuần đến 30 tuần để quyết định thoát khỏi kênh đầu tư nếu lợi nhuận không đạt kỳ vọng (Bugra & Bernat, 2018). Một khi ICO tiềm năng được phát hành, *token* được chấp nhận rộng rãi thì giá *token* sẽ tăng lên vượt bậc và mang lại lợi nhuận theo cấp số nhân cho nhà đầu tư so với giá lúc mua khi phát hành. Lúc này nhà đầu tư sẽ bán các *token* này để kiếm lời và đó chính là lợi nhuận (Hartmann, Grottole, Wang, & Lunesu, 2019). Các yếu tố ảnh hưởng đến sự không chắc chắn về sự thành công của ICO bao gồm chất lượng nhóm phát hành, thông tin về ICO, ý tưởng sản phẩm, phương tiện truyền thông, mạng xã hội và ý kiến của các chuyên gia trong lĩnh vực về tiền điện tử (Burns & Moro, 2018; Myalo & Glukhov, 2019). Chất lượng nhóm phát hành bao gồm nhóm làm việc với nhiều năm kinh nghiệm và kỷ luật tốt thì chất lượng sản phẩm cũng sẽ tốt hơn. Thông tin về ICO đề cập ngày bắt đầu và kết thúc bán token, giao dịch bằng cách nào, giá, tổng nguồn cung, vốn hóa thị trường, ... Ý tưởng sản phẩm đề cập công ty cần phải trình bày ý tưởng bằng video thuyết trình, công nghệ, nền tảng, dịch vụ sử dụng là gì, các cột mốc mà công ty muốn đạt được khi tung ra sản phẩm. Phương tiện truyền thông, mạng xã hội cho một dự án ICO nổi bật là khi nó được đề cập nhiều ở các mạng xã hội như Facebook, Twitter (Burns & Moro, 2018).

Thuật toán dự đoán giá ICO là công cụ hỗ trợ cho lợi ích của nhà đầu tư và tư vấn viên. Thuật toán dự đoán bằng máy học đem lại hiệu quả và tính chính xác cao vì vậy nó là xu hướng của các ứng dụng dự báo (Hartmann, Wang, & Lunesu, 2018). Vì vậy, những nhà nghiên cứu đã đưa ra rất nhiều thuật toán máy học ứng dụng trong dự báo như mô hình hồi quy tuyến tính bội, mô hình hồi quy Ridge, mạng nơron nhân tạo, support vector machine, ... (Chen, Challita, Saad, Yin, & Debbah, 2019; Deanna, 2018; Mingxi, Guowen, Jianping, Xiaoqian, & Yinhong, 2020; Yaohao & Pedro, 2018). Việc dự đoán giá ICO sắp phát hành trong tương lai gần là rất quan trọng. Khi các công ty, nhà đầu tư hoặc các tư vấn viên có được con số dự đoán cụ thể với độ chính xác cao, họ sẽ có những bước đi phù hợp bởi căn cứ vào kết quả dự đoán đó.

Những phương pháp máy học khác nhau đã được áp dụng để tiên đoán sự thành công của dự án ICO. Phân tích dựa trên cảm xúc được xem như một công cụ hữu ích để đánh giá độ hấp dẫn của ICO (Anchaya & ctg., 2020; Ahmed, 2021). Dựa trên các bình luận của tài khoản người dùng trên Twitter, các tác giả đã xây dựng dữ liệu cảm xúc của người dùng để đánh giá sự thành

công của một ICO và số tiền kêu gọi đầu tư thành công. Mô hình máy học hồi quy *logistic* và *random forest* được dùng để dự báo sự thành công ICO bởi tính chính xác cao trong việc phân tích cảm xúc của người dùng. Thêm vào đó, phân tích dựa trên các báo cáo trắng của ICO cũng được nghiên cứu để dự báo sự thành công ICO. Mô hình xử lý ngôn ngữ tự nhiên phân tích các thuật ngữ thường được sử dụng trong các báo cáo trắng của ICO thành công để đánh giá một ICO khác. Nghiên cứu kết luận rằng ICO thành công có các báo cáo trắng đầy đủ thông tin với các thuật ngữ được xem như các biến đầu vào của mô hình (Bugra & Bernat, 2018). Mục đích trong nghiên cứu này, chúng tôi dự báo giá ICO sau sáu tháng phát hành để đánh giá mức độ thành công ICO với sự xem xét nhiều yếu tố ảnh hưởng tới giá ICO có thể có kết quả là lợi nhuận hoặc bất lợi. Sau sáu tháng đầu tư, nhà đầu tư kỳ vọng giá trị lợi nhuận trả về thỏa mãn tỉ suất lợi nhuận. Dựa trên các trường dữ liệu đầu vào như các tham số của mô hình dự báo, nghiên cứu này phân tích tương quan các đầu vào so với đầu ra để đánh giá sự ảnh hưởng của các tham số tới kết quả giá ICO sau sáu tháng. Tập dữ liệu được xây dựng bao gồm 12 trường, bao gồm các yếu tố chính ảnh hưởng tới giá của một ICO. Các yếu tố này bao gồm giá đô la, giá bitcoin, tổng vốn cung ứng, vốn hóa thị trường, nguồn cung có sẵn, tiền nhận được, giá Ethereum tại thời điểm mở bán, giá bitcoin tại thời điểm mở bán, tháng bắt đầu mở bán ICO, ngày ICO mở bán, quốc gia mở bán ICO, thời lượng phát hành ICO. Sáu tháng sau phát hành ICO là khoảng thời gian mong đợi của nhà đầu tư đánh giá sự thành công của ICO. Hai phương pháp hồi quy ridge và mạng neuron nhân tạo được đề xuất và so sánh đánh giá độ chính xác dự báo trên tập cơ sở dữ liệu được xây dựng từ các website tin cậy về tiền điện tử ICO bởi quá trình mô phỏng sử dụng ngôn ngữ lập trình Python. Các bảng và hình được trình bày là kết quả trích từ quá trình mô phỏng bởi công cụ lập trình sử dụng ngôn ngữ Python này.

Phương pháp máy học trên chưa được vận dụng vào dự báo giá ICO theo những yếu tố khác nhau gây ra ảnh hưởng tới giá ICO. Nghiên cứu trước về dự báo sự thành công của ICO thường chỉ phân tích yếu tố cảm xúc người dùng ảnh hưởng tới sự thành công của ICO hoặc dựa vào yếu tố các báo cáo trắng để đưa ra dự đoán. Ngoài các yếu tố này, giá ICO còn phụ thuộc vào nhiều yếu tố khác mà chưa có nghiên cứu nào sử dụng để phân tích đánh giá khả năng ảnh hưởng tới sự thành công của giá ICO sau sáu tháng phát hành. Bài báo mô tả cơ sở lý thuyết xây dựng mô hình hồi quy Ridge dựa trên mô hình hồi quy tuyến tính và kiến trúc mạng neuron nhân tạo ở phần 2. Phần 3 trình bày quy trình xây dựng và tiền xử lý dữ liệu. Phần 4 chỉ ra các kết quả phân tích đánh giá mô hình vào việc dự báo giá ICO sau sáu tháng. Phần 5 đưa ra các kết luận với những đóng góp.

2. Mô hình thuật toán hồi quy Ridge

2.1. Bài toán hồi quy tuyến tính bội

Phương trình (1) chỉ ra mối quan hệ của mô hình hồi quy tuyến tính bội. Phương trình trình bày mối quan hệ giữa đầu ra và đầu vào của một hàm tuyến tính:

$$\hat{y} = h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n = \theta^T X \quad (1)$$

Trong đó \hat{y} được định nghĩa là giá trị dự đoán. Ma trận $\theta^T = [\theta_0 \ \theta_1 \ \dots \ \theta_n]$ mô tả các tham số mô hình và là ma trận chuyển vị của ma trận θ . Các biến $[\theta_0 \ \theta_1 \ \dots \ \theta_n]$ là tham số mô hình bài toán. Giá trị dự đoán \hat{y} và các tham số trình bày mối quan hệ tuyến tính. Ma trận X là ma trận chứa các biến của mô hình. $X = [x_1 \ x_2 \ \dots \ x_n]$.

Bài toán hồi quy tuyến tính có những hạn chế như rất nhạy cảm với nhiễu. Khi hệ thống gặp nhiễu gây ra sự không tuyến tính giữa đầu vào và đầu ra. Điều này sẽ ảnh hưởng đến kết quả dự báo khi dùng phương pháp hồi quy tuyến tính. Tiếp theo, hồi quy tuyến tính không biểu diễn được các mô hình có độ phức tạp lớn như mô hình bao gồm nhiều đầu vào. Khi mô hình trở nên phức tạp dẫn đến hiện tượng quá khớp.

2.2. Bài toán hồi quy Ridge

Hồi quy Ridge có sự thay đổi so với hồi quy tuyến tính bởi thêm vào một thành phần điều chỉnh vào hàm mất mát. Về bản chất, hồi quy Ridge tối ưu song song hai thành phần bao gồm tổng bình phương phần dư và thành phần điều chỉnh. Phương trình của hàm tổn thất của hồi quy Ridge có dạng trong phương trình (2).

$$J(\theta) = \frac{1}{2N} \left(\sum_{i=1}^N (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{i=1}^N \theta_i^2 \right) \quad (2)$$

Ở đây, x^i đặc trưng cho các đầu vào thứ i của mô hình. N gọi là kích cỡ mẫu. y^i đại diện cho giá trị thực tế. λ được gọi là tham số điều chỉnh độ phức tạp của mô hình. Tham số λ này được dùng để kiểm soát độ lớn của thành phần điều chỉnh tác động lên hàm mất mát. Hồi quy Ridge áp dụng kỹ thuật kiểm soát thành phần hiệu chỉnh này để giúp giải quyết các hiện tượng đa cộng tuyến hoàn hảo và quá khớp của dữ liệu. Trong trường hợp λ rất lớn, hầu như tất cả các tham số mô hình suy giảm về 0 và được gọi là hiện tượng phù hợp dưới mức (*underfitting*). Khi λ rất nhỏ, hồi quy Ridge trở thành hồi quy tuyến tính thông thường. Điều này dẫn đến hiện tượng quá khớp (*overfitting*). Trường hợp λ nhỏ, vai trò thành phần điều chỉnh ít quan trọng, mức độ kiểm soát hiện tượng quá khớp trở nên kém hơn. Thật quan trọng khi tham số λ được thiết kế phù hợp. Chúng tôi đã thực nghiệm với nhiều giá trị λ khác nhau trong quá trình huấn luyện. Mô hình huấn luyện tốt nhất là mô hình cần giảm thiểu hiện tượng quá khớp, khi đó giá trị tham số λ tối ưu.

Từ công thức (2), có thể được viết lại công thức (3).

$$J(\theta) = \frac{1}{2N} (|\theta^T X - y|^2 + \lambda \theta^2) \quad (3)$$

Với $y = \begin{bmatrix} y^1 \\ y^2 \\ \vdots \\ y^n \end{bmatrix}$ mô tả ma trận chứa các giá trị thực tế của các biến phụ thuộc, $X = \begin{bmatrix} x^1 \\ x^2 \\ \vdots \\ x^n \end{bmatrix}$ là

ma trận chứa các biến ngõ vào. Ma trận tham số mô hình $\theta^T = [\theta_0 \quad \theta_1 \quad \dots \quad \theta_n]$.

Một cách tương tự như vấn đề đặt ra của bài toán hồi quy bội, tham số mô hình $J(\theta)$ cần được thiết kế với giá trị cực tiểu.

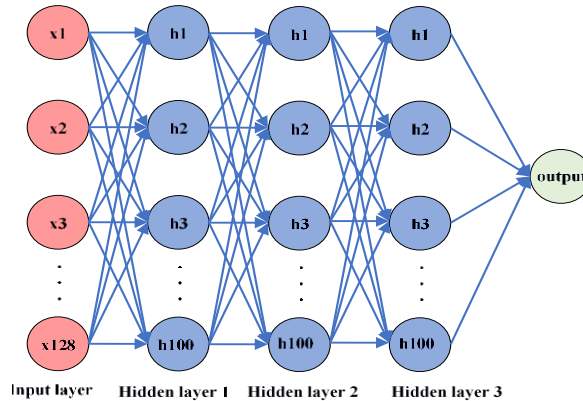
❖ Nghiệm bài toán hồi quy Ridge (Deanna, 2018):

$$\theta = (X^T X + \lambda I)^{-1} X^T y \quad (4)$$

Trong đó, ma trận X chứa các biến ngõ vào và có ma trận chuyển vị là X^T . Ma trận đơn vị I được thiết kế với kích thước phù hợp. Dễ dàng nhận ra nghiệm hồi quy Ridge bao gồm thêm một lượng λI khác với hồi quy tuyến tính.

Hồi quy Ridge đưa thêm số hạng (λI) vào hàm tổn thất. Kỹ thuật này được gọi là sử dụng phương pháp hiệu chỉnh. Kỹ thuật hiệu chỉnh này làm cho hàm $(X^T X + \lambda I)$ luôn tồn tại sự khả nghịch vì vậy phương trình (2) luôn có nghiệm.

2.3. Thuật toán mạng neuron nhân tạo (ANN)



Hình 1. Mô hình mạng neuron cho dự báo giá ICO

Nguồn: Chen và cộng sự (2019)

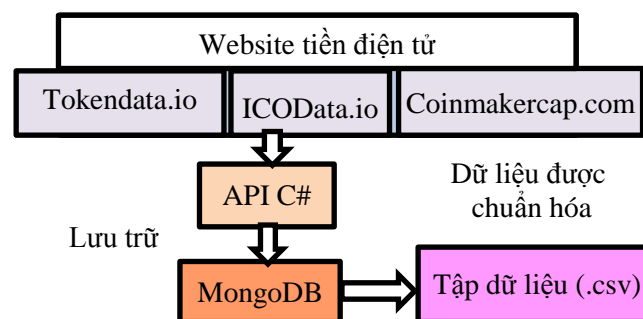
Mạng neuron nhân tạo là kỹ thuật dự báo dựa trên mô hình dùng để mô phỏng hoạt động của bộ não người. Mạng ANN trình diễn mối quan hệ phi tuyến giữa các biến đầu vào và đầu ra. Thuật toán mạng Neuron nhân tạo thực hiện theo hai giai đoạn bao gồm giai đoạn lan truyền tiến và giai đoạn lan truyền ngược. Giải thuật tính toán với lan truyền tiến bao gồm tổng các tích giữa ngõ vào với trọng số w . Kết quả tính toán này được lan truyền đi qua lớp ẩn. Giá trị ở lớp ẩn và lớp ngõ ra được tính toán sử dụng hàm *Tanh*. Giải thuật tính toán bước lan truyền ngược bao gồm việc cập nhật lại trọng số bởi thuật toán *Gradient Descent* để giảm mất mát. Việc cập nhật trọng số được thực hiện liên tục tới khi giá trị mất mát đạt tới một sai số chấp nhận được của mô hình.

Hình 1 trình bày kiến trúc mạng neuron cho hệ thống dự báo giá ICO sau sáu tháng. Để xác định số lớp ẩn và số node trong lớp ẩn, các thực nghiệm được nghiên cứu nhiều lần. Từ đó, các tham số phù hợp với kết quả chính xác nhất được xác định. Nghiên cứu thiết kế mô hình mạng neuron bao gồm lớp ngõ vào với 128 trường dữ liệu, 03 lớp ẩn, 100 node trong mỗi lớp ẩn, một neuron lớp đầu ra. Hàm *Tanh* được dùng làm hàm tích cực trong các lớp ẩn.

3. Thiết kế hệ thống dự báo giá ICO

3.1. Quá trình thu thập dữ liệu

Hệ thống dự báo giá ICO này có 02 thành phần, bao gồm giải thuật và tập *dataset*. Tập *dataset* được xây dựng từ nhiều website liên quan đến giá ICO. Thuật toán sử dụng phương pháp hồi quy Ridge để thực thi với tập *dataset* này. Kết quả giải thuật chỉ ra giá của một ICO sau khi phát hành 06 tháng.



Hình 2. Sơ đồ các bước thu thập dữ liệu

Nguồn: Maria và Omar (2020)

Tập dữ liệu dataset được xây dựng từ các nguồn dữ liệu Internet ICO tin cậy về lĩnh vực tiền điện tử như tokendata.io, icodata.io, coinmarketcap.com bởi giao diện API được lập trình bằng C# (Maria & Omar, 2020). Sau đó, bước tiền xử lý dữ liệu được thực thi để lọc bỏ và xử lý những dữ liệu không hợp lệ hoặc bị trống do thiếu dữ liệu. Định dạng file .csv được dùng để lưu trữ. Bước chuẩn bị dữ liệu này nhằm làm sạch và chuyển đổi dữ liệu thô chưa định dạng thành dữ liệu hữu dụng để có thể dùng phân tích chúng. Dữ liệu không hợp lệ là dữ liệu có các trường dữ liệu *start_date* và *end_date* với giá trị không hợp lệ. Tập dữ liệu hoàn chỉnh có 109 ICO bao gồm 109 hàng và 12 cột.

Bảng 1

Bộ dữ liệu dataset chia thành 03 phần

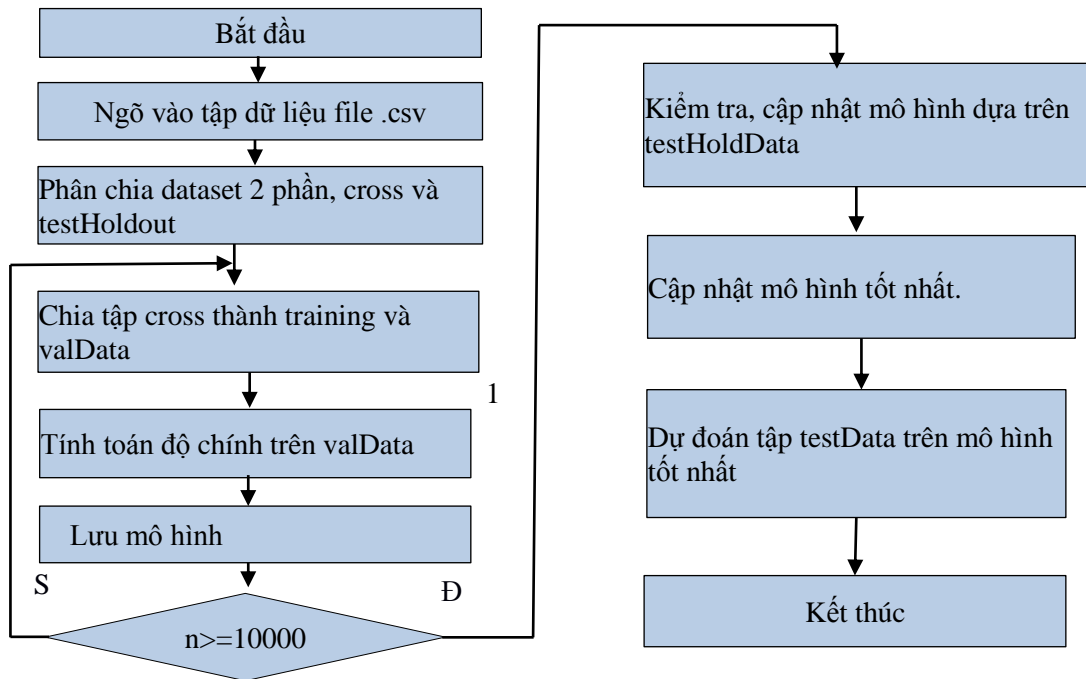
Cross data		Test holdout data	Test data
Training data	Validation data	21 ICO	3 ICO
68 ICO	17 ICO		

Nguồn: Tác giả xây dựng (2022)

Để thuật toán máy học thực hiện dự đoán chính xác hơn, mã hóa *One Hot* được sử dụng để chuyển đổi các trường dữ liệu trong định dạng chữ thành dữ liệu nhị phân. Áp dụng mã hóa *One hot* để phân loại cho các trường dữ liệu. Tuy nhiên, kỹ thuật này gây ra tăng số lượng đầu vào. Áp dụng mã hóa *One hot* để phân loại cho bốn trường dữ liệu cuối bao gồm: *ICO duration day*, *Date ICO was lunched*, *Month ICO was lunched* và *County an ICO was lunched from*. Dữ liệu các trường cuối này phải chuyển từ định dạng chữ sang định dạng số. Do đó, tập dataset sẽ không còn là 12 cột mà lên đến 128 cột, số hàng vẫn 109 hàng.

Tập dữ liệu dataset gồm 109 đồng ICO được tách ra thành 106 đồng ICO dùng để huấn luyện mô hình, và 3 đồng ICO được dùng để làm tập kiểm tra (tập *testData*) đồng thời nhằm đánh giá độ chính xác thuật toán đề xuất. Tiếp tục tách 106 đồng ICO ra làm 02 tập nhỏ hơn, bao gồm tập *cross* với hai trục tọa độ *XCross* và *yCross*, và tập *testHoldout* bao gồm hai trục tọa độ *XTestHoldout* và *yTestHoldout*. Một cách tương ứng, tập dataset được phân chia với tỷ lệ 80% cho tập *cross* (bao gồm 85 ICO) và 20% cho tập *testHoldout* (bao gồm 21 ICO). Sử dụng tập *cross* này để huấn luyện thuật toán hồi quy Ridge. Từ tập dữ liệu *cross* này (bao gồm 85 ICO) tiến hành tách ra 02 tập con: tập huấn luyện bao gồm hai trục tọa độ *XTrain* và *yTrain*, và tập đánh giá (tập *validation*) gồm hai trục tọa độ *XVal* và *yVal* với tỷ lệ phân chia 80% cho tập huấn luyện cùng với 20% cho tập *validation* để thực hiện đánh giá tránh hiện tượng quá khớp và tìm các siêu tham số tốt hơn như được trình bày trong Bảng 1. Hồi quy Ridge dùng tập huấn luyện này để huấn luyện cho mô hình. Mô hình sau khi được huấn luyện sẽ đánh giá độ chính xác bằng cách sử dụng tập (*XVal*, *yVal*). Trong nghiên cứu này, các thước đo hiệu năng RMSE, MAE và R^2 được sử dụng để đánh giá sai số giữa giá trị dự đoán so với giá trị thực tế.

Tập dữ liệu *cross* dùng để huấn luyện thuật toán và lưu mô hình (model) dưới định dạng file .sav. Kết quả của việc huấn luyện sẽ tạo ra ba mô hình với ba giá trị RMSE, MAE và R^2 trong vòng lặp đầu tiên. Tập *testHoldoutData* dùng để theo dõi độ chính xác bởi vòng lặp để điều chỉnh thông số và tìm mô hình huấn luyện tối ưu dựa trên các thước đo đánh giá hiệu năng của mô hình. Mô hình thực hiện tối ưu dựa trên một vòng lặp với tập mẫu được chọn khác nhau qua mỗi vòng lặp. Kết quả là giá trị RMSE, MAE và R^2 khác nhau sau mỗi lần lặp. Giá trị thước đo hiệu năng được đánh giá và so sánh nhau qua mỗi lần lặp. Mô hình tối ưu chọn giá trị RMSE, MAE và R^2 tốt nhất. Tập *testData* đưa vào mô hình tối ưu này để tính sai số giữa giá trị thực tế và giá trị dự báo như được mô tả trong lưu đồ Hình 3.



Hình 3. Lưu đồ các bước thực hiện dự báo áp dụng cho cả mô hình hồi quy Ridge và mạng ANN

Nguồn: Tác giả xây dựng (2022)

3.2. Phân tích sự tương quan các biến

Căn cứ vào sự ảnh hưởng đầu vào đối với đầu ra, nghiên cứu này phân tích mối tương quan mạnh yếu giữa các yếu tố đầu vào đầu ra này. Một biểu đồ phân tán với các hệ số tương quan tương quan giữa đầu vào và đầu ra được trình bày.

Hình 4 diễn tả sự tương quan giữa các đầu vào tương ứng đầu ra, và giữa các đầu vào với nhau. Mối tương quan mạnh thể hiện giá trị hệ số tương quan lớn và được in đậm để làm nổi bật lên. Như được mô tả trong Hình 4, đường chéo chính là tên các đầu vào và đầu ra.

- **Giá đô la:** trình bày trong biến đầu vào price_usd. Đây là giá của một token được quy đổi bằng giá đô la. Hệ số tương quan là 0.88 giữa hai biến price_usd và output. Sự tương quan này thể hiện mối tương quan dương mạnh tức là nếu price_usd tăng thì giá ngõ ra ICO cũng tăng.

- **Giá bitcoin:** trình bày giá của token được quy đổi theo giá bitcoin. Hệ số tương quan là 0.88 giữa hai biến price_btc và output. Mối tương quan là mạnh giữa hai biến btc_price và output này.

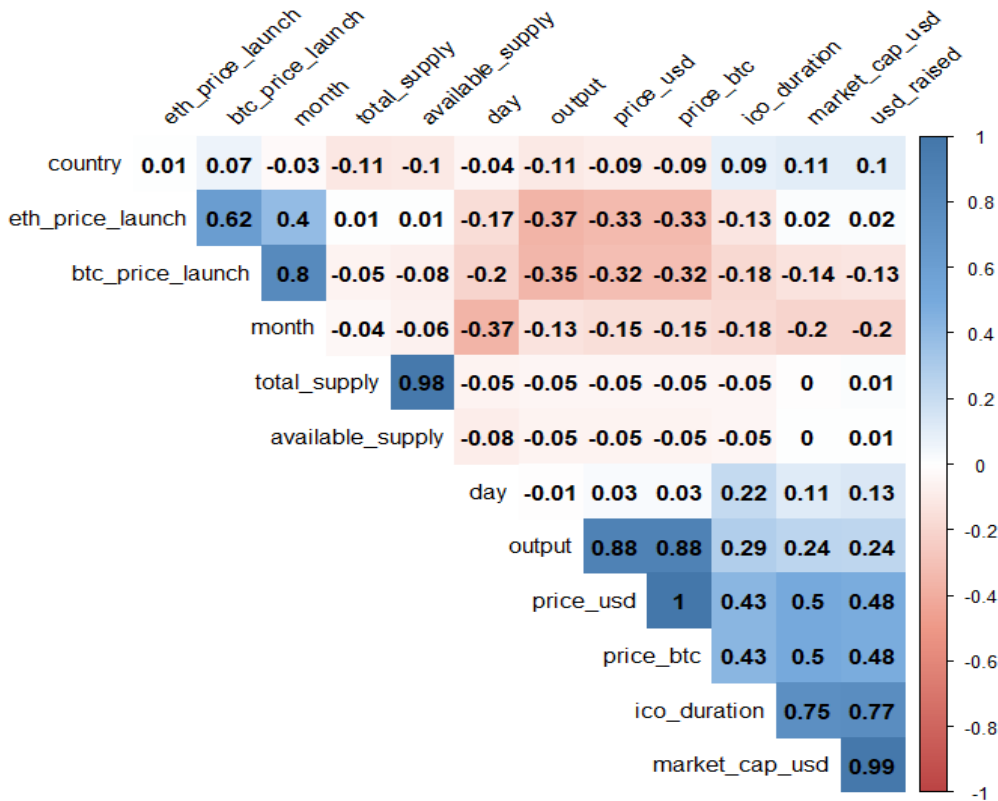
- **Tổng cung ứng:** sử dụng biến total_supply, là tổng số token có thể cung ứng cho thị trường. Ví dụ như đồng 0x cung ứng ra thị trường với một tỷ token. Đồng aeron cung ứng ra thị trường với 20 triệu token. Hai biến này gần như không có tương quan với nhau vì hệ số tương quan là -0.05.

- **Vốn hóa thị trường:** biểu diễn bằng biến ngõ vào market_cap_usd và output là 0.24. Giá trị tương quan này có nghĩa rằng hai biến có mối tương quan yếu.

- **Nguồn cung có sẵn:** đặc trưng cho lượng token đang lưu thông. Hình 4 chỉ ra giá trị tương quan là -0.05 giữa biến ngõ vào và ngõ ra, chứng tỏ chúng không có sự tương quan với nhau.

- **Tiền nhận được:** là tổng số tiền nhận được kể từ lúc công ty phát hành đến lúc công ty kết thúc mở bán tính bằng đô la. Hệ số tương quan là 0.24 giữa hai biến usd_raised và output dựa trên Hình 4. Mối liên hệ giữa hai biến này là yếu.

- Giá Ethereum tại thời điểm mở bán: giá của Ethereum tại thời điểm phát hành ảnh hưởng không đáng kể đến đầu ra. Hệ số tương quan là -0.366 giữa hai biến eth_price_launch và output. Hai biến này có mối liên hệ yếu.



Hình 4. Biểu đồ phân tích sự tương quan giữa các biến ngõ vào với ngõ ra và giữa các ngõ vào với nhau sau khi mô phỏng tương quan

Nguồn: Tác giả mô phỏng (2022)

- Giá Bitcoin tại thời điểm mở bán: một cách tương tự giá Ethereum, giá Bitcoin tại thời điểm phát hành ảnh hưởng nhỏ đến đầu ra. Hệ số tương quan là -0.36565 giữa hai biến btc_price_launch và output. Mối liên hệ yếu giữa hai biến này.

- Tháng bắt đầu mở bán ICO: mô tả tháng mà ICO được phát hành. Biến này ít ảnh hưởng đến đầu ra. Hệ số tương quan là -0.37 giữa hai biến month và ngõ ra (output). Hai biến có mối liên hệ yếu.

- Ngày ICO mở bán: mô tả thời gian ngày mà ICO được phát hành trong tháng. Biến day ảnh hưởng nhỏ đến biến ngõ ra output. Hệ số tương quan được trình bày trong Hình 4 là 0.03. Hai biến này có mối liên hệ yếu.

- Quốc gia mở bán ICO: mô tả quốc gia mà ICO được phát hành. Hệ số tương quan là -0.11 giữa hai biến country và output. Mối liên hệ yếu giữa hai biến này.

- Thời lượng phát hành ICO: là khoảng thời gian phát hành ICO bao nhiêu ngày. Ví dụ: 0x phát hành trong 09 ngày. Dựa vào biểu đồ có thể thấy hai biến này gần như độc lập với nhau bởi có hệ số tương quan gần 0.2.

Kết quả phân tích ở trên có thể nhận thấy rằng tương quan mạnh giữa hai đầu vào giá đô la và giá bitcoin tương ứng hai biến Price_usd và Price_btc. Mối tương quan giữa các biến còn lại tương đối thấp hoặc tương quan yếu.

Một câu hỏi đặt ra là phân tích tương quan mang lại điều gì? Dễ nhận thấy nếu các biến đầu vào và biến đầu ra có mối tương quan cao, mô hình hồi quy tuyến tính được đề nghị để dự báo giá ICO và đạt được kết quả dự đoán tốt. Mặc dầu hai biến price_usd và price_btc có hệ số tương quan cao với ngõ ra nhưng chúng ta không thể xây dựng mô hình chỉ xem xét hai biến này mà bỏ qua tất cả các biến đầu vào khác. Trong khi đó, các biến đầu vào có tương quan yếu với biến đầu ra nhưng chúng vẫn có mối tương quan nhau. Trong nghiên cứu này, mô hình xem xét tất cả 12 biến đầu vào vào mô hình dự báo giá ICO.

Dựa theo mô tả Hình 4 về sự tương quan giữa các biến, chọn hai biến đầu vào market_cap_usd và available_supply vì 02 biến này có hệ số tương quan gần bằng 0 sau khi làm tròn (hệ số tương quan của hai biến là 0.00066). Có nghĩa rằng hai biến độc lập với nhau. Giả sử hệ số hồi quy của market_cap_usd là X_1 và hệ số hồi quy của available_supply là X_2 . Se(X_1) là sai số chuẩn hệ số hồi quy của X_1 , Se(X_2) là sai số chuẩn hệ số hồi quy của X_2 . Sai số chuẩn hồi quy là hệ số dùng để đo độ chính xác của hệ số hồi quy ước tính. Dựa theo kết quả Bảng 2 có thể thấy hệ số hồi quy của các đầu vào, sai số chuẩn hồi quy và tổng bình phương gần như là không thay đổi. Từ đó kết luận, với các đầu vào độc lập với nhau thì các giá trị hệ số hồi quy, sai số chuẩn hệ số hồi quy và giá trị tổng bình phương gần như không thay đổi.

Tương tự, Hình 4 mô tả tương quan giữa các biến, chọn hai biến tiên lượng usd_raised và ico_duration vì hai biến này có hệ số tương quan gần bằng 0.77, xem xét với hai biến có sự tương quan lớn với nhau ảnh hưởng như thế nào. Giả sử hệ số hồi quy của market_cap_usd là X_1 và hệ số hồi quy của available_supply là X_2 và Se(X_1) là sai số chuẩn hệ số hồi quy của X_1 , Se(X_2) là sai số chuẩn hệ số hồi quy của X_2 . Hệ số hồi quy thay đổi đáng kể khi sử dụng hai biến tiên lượng có giá trị tương quan mạnh như trong Bảng 3.

Bảng 2

Đánh giá tương quan các biến market_cap_usd và available_supply

Mô hình	X_1	Se(X_1)	X_2	Se(X_2)	Sum of squares
market_cap_usd	6.7e-10	2.6e-10	NA	NA	73.9
available_supply	NA	NA	-2.3e-12	4.2e-12	3.4
Model gồm 2 biến trên	6.7e-10	2.6e-10	-2.3e-12	4.1e-12	73.9/3.4

Nguồn: Tác giả mô phỏng (2022)

Bảng 3

Phân tích tương quan các biến ico_duration và usd_raised

Mô hình	X_1	Se(X_1)	X_2	Se(X_2)	Sum of squares
usd_raised	2.9e-09	1.0e-09	NA	NA	70
ico_duration	NA	NA	2.4e-02	7.7e-3	104.4
Model gồm 2 biến trên	3.8e-10	1.6e-09	2.2e-02	1.2e-02	70/34.9

Nguồn: Tác giả mô phỏng (2022)

Khác với hai biến gần như độc lập thì hai biến có mối tương quan cao khi đưa vào thì kết quả trả về hệ số hồi quy khác biệt rất nhiều so với khi sử dụng mô hình hồi quy từng biến độc lập. Cụ thể hệ số hồi quy của usd_raised và ico_duration đã giảm đi khi đưa vào mô hình. Hệ số hồi

quy sẽ đi cùng với giá trị các biến độc lập. Phương trình hồi quy tuyến tính là một hàm của các biến của mô hình và hệ số hồi quy. Giá trị các biến mô hình lớn thì hệ số hồi quy sẽ nhỏ để đạt giá trị kết quả phù hợp với các biến và hệ số hồi quy khác. Dựa trên phân tích hệ số hồi quy và sai số chuẩn của hệ số hồi quy, đánh giá dữ liệu có hiện tượng đa cộng tuyến xảy ra trong mô hình hồi quy tuyến tính. Để vượt qua hiện tượng đa cộng tuyến của tập dữ liệu này, nghiên cứu đề xuất sử dụng mô hình hồi quy Ridge có khả năng vượt qua hiện tượng đa cộng tuyến này.

Sự thay đổi sai số chuẩn của hệ số hồi quy có thể ảnh hưởng đến độ chính xác của hệ số. Cụ thể khi sử dụng mô hình hồi quy đơn biến nếu chỉ sử dụng biến `usd_raised` thì sai số (Standard Error) là $1.0e-9$, nhưng khi đưa vào mô hình hồi quy bội thì lại tăng lên thành $1.6e-09$. Tương tự với tham số `ico_duration` thì sai số cũng tăng, từ $7.7e-3$ tới $1.2e-02$. Hiện tượng đa cộng tuyến xảy ra.

Bảng 4

Bảng phân tích quá khớp trong mô hình hồi quy bội với tập `testHoldout`

	Lỗi huấn luyện	Lỗi kiểm tra
Mô hình	0.0027	11.5

Nguồn: Tác giả mô phỏng (2022)

Dựa theo hai thông số lỗi huấn luyện và lỗi kiểm tra, chúng ta kết luận hiện tượng quá khớp đang tồn tại khiến mô hình dự đoán trở nên sai lệch. Bảng 4 cho thấy lỗi huấn luyện lúc này là 0.0027. Kết quả này cho thấy mô hình dự đoán trên tập huấn luyện rất tốt. Kết quả này đạt được sau khi đưa tập dữ liệu vào mô hình hồi quy tuyến tính với các hệ số để tối thiểu sự sai biệt giữa giá trị thực tế và giá trị dự đoán bởi mô hình tuyến tính. Sai số này chứng tỏ dự đoán rất tốt trên tập huấn luyện. Tuy nhiên, giá trị dự đoán và giá trị thật không khớp với nhau ở tập kiểm tra (tập `test`). Giá trị dự đoán đã thay đổi khá xa so với giá trị thật. Dựa trên kết quả Bảng 4 về lỗi huấn luyện và lỗi kiểm tra, ta kết luận rằng hiện tượng quá khớp đang tồn tại. Để tránh hiện tượng quá khớp này, kỹ thuật thêm vào một đại lượng chính quy hóa vào hàm tổn thất sẽ giải quyết vấn đề quá khớp này.

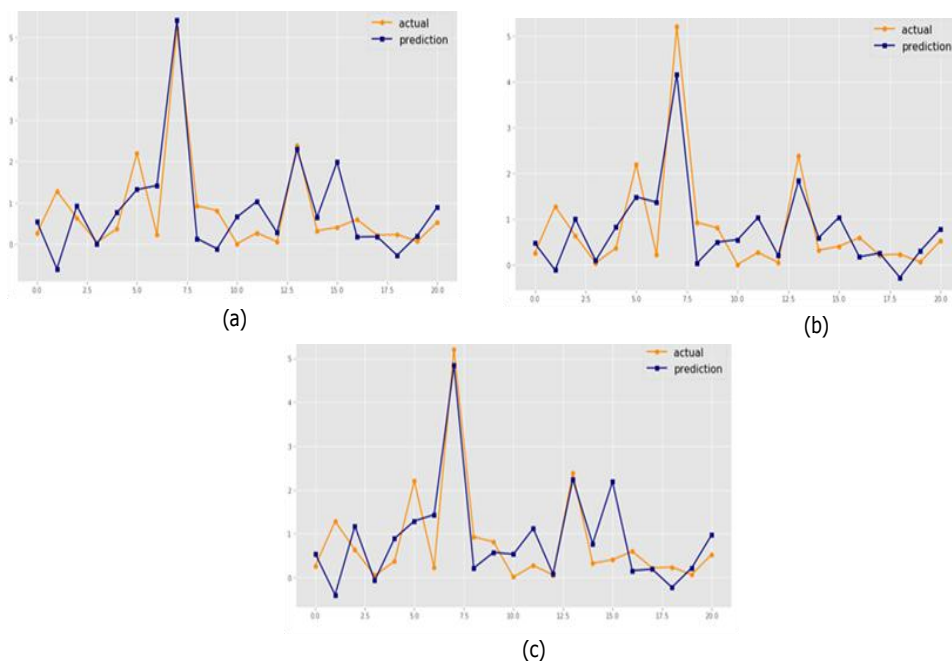
4. Kết quả mô phỏng

Để đánh giá độ chính xác của mô hình thuật toán, các thước đo hiệu năng được sử dụng thông qua đại lượng RMSE được trình bày công thức (5), R^2 được trình bày công thức (6) và MAE trong công thức (7) (Jigar, Sahil, Priyank, & Kotecha, 2015). Ở đây, \hat{y} định nghĩa giá trị dự đoán của y . y được định nghĩa là giá trị thực tế. \bar{y} được xem là giá trị trung bình.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y - \hat{y})^2} \quad (5)$$

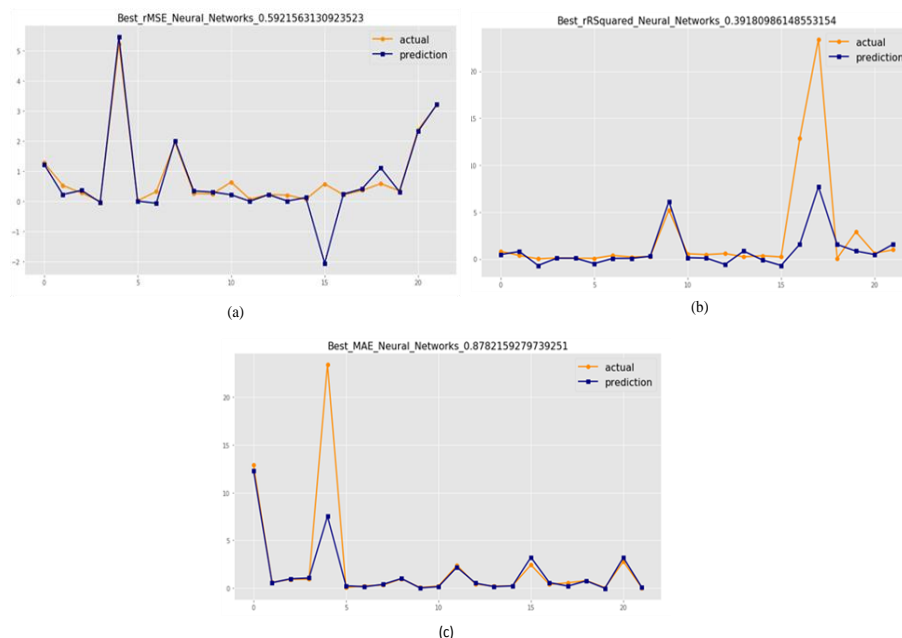
$$R^2 = 1 - \frac{\sum_{i=1}^N (y - \hat{y})^2}{\sum_{i=1}^N (y - \bar{y})^2} \quad (6)$$

$$MAE = \frac{\sum_{i=1}^N abs(y - \hat{y})}{N} \quad (7)$$



Hình 5. Đánh giá kết quả so sánh giữa giá trị $y_{TestHoldout}$ & $y_{PredHoldout}$ sử dụng mô hình hồi quy Ridge (a) RMSE = 0.75; (b) Rsquared = 0.71; (c) MAE = 0.56

Nguồn: Tác giả mô phỏng (2022)



Hình 6. Đánh giá kết quả so sánh giữa giá trị $y_{TestHoldout}$ & $y_{PredHoldout}$ trong mô hình ANN (a) Rmse = 0.59; (b) Rsquared = 0.39; (c) MAE = 0.88

Nguồn: Tác giả mô phỏng (2022)

Một bộ dữ liệu ngẫu nhiên được chạy thực nghiệm để có được ba mô hình huấn luyện tốt nhất. Một mô hình đạt được chỉ số thước đo hiệu năng RMSE tốt nhất, một mô hình có chỉ số thước đo hiệu năng R^2 tốt nhất và một mô hình đạt chỉ số thước đo hiệu năng MAE tốt nhất. Để đánh giá thuật toán hồi quy Ridge, quá trình huấn luyện thực thi chạy 10,000 vòng lặp để tạo ra mô hình. Đánh giá kết quả huấn luyện trả về từng mô hình theo từng thước đo đánh giá RMSE, R^2 , MAE.

Sau quá trình huấn luyện, dữ liệu XTestHoldout được đưa vào từng mô hình được huấn luyện tốt nhất ở trên để thực thi bước kiểm tra. Kết quả của quá trình kiểm tra sẽ trả về là dữ liệu dự đoán yPredHoldout. Sau đó, nó được kết hợp với dữ liệu yTestHoldout để tính toán thước đo hiệu năng của mô hình. Hình 5 và 6 (a), (b), (c) tương ứng kết quả đánh giá với ba mô hình tốt nhất của hồi quy Ridge và ANN trên.

Bảng 5

Đánh giá ba đồng ICO trong tập dữ liệu chưa nhìn thấy để dự đoán độ chính xác trong dự báo giá dùng mô hình hồi quy Ridge và ANN

Tên ICO	Best RMSE		Best R ²		Best MAE	
	Hồi quy Ridge	ANN	Hồi quy Ridge	ANN	Hồi quy Ridge	ANN
0x	0.94 (86%)	1.05 (98%)	1.179 (92%)	1.32 (82%)	1.173 (92%)	0.98 (90%)
Modum	4.377 (63%)	3.42 (85%)	2.264 (76%)	5.31 (52%)	2.19 (73%)	3.2 (87%)
Crypto20	1.267 (75%)	0.94 (66%)	1.09 (87%)	1.29 (74%)	1.1756 (81%)	1.02 (94%)

Nguồn: Tác giả mô phỏng (2022)

Bảng 5 trình bày kết quả dự đoán với tập kiểm tra (*testData*) gồm ba ICO được chọn từ tập dữ liệu 109 ICO. Tập dữ liệu 109 ICO này được thu thập là giá ICO sau 06 tháng phát hành được sử dụng trong quá trình áp dụng hai phương pháp dự báo. Dựa trên tập dataset này, một tập con với giá 85 ICO sau 06 tháng được sử dụng làm tập huấn luyện để tìm mô hình tốt nhất. Tập đánh giá gồm 21 ICO được chọn trong tập dataset này. Ở đây, ba đồng ICO được chọn ngẫu nhiên đã phát hành với giá sau 06 tháng được chọn làm tập kiểm tra để đánh giá độ chính xác của các phương pháp dự báo. Nếu kết quả dự báo có độ chính xác gần như giá ICO sau 06 tháng phát hành, có nghĩa rằng ICO được dự báo có độ chính xác cao. Từ đó, nhà đầu tư có thể dự đoán được việc đầu tư vào đồng ICO nào có lợi nhuận, thỏa mãn được sự kỳ vọng của nhà đầu tư. Tập kiểm tra này được dùng để đánh giá hiệu năng của ba mô hình tốt nhất dùng ba thước đo hiệu năng RMSE, R², và MAE. Cả ba mô hình thuật toán đều có kết quả khá chính xác. Kết quả mô phỏng dùng ngôn ngữ Python cho kết quả rằng, thuật toán hồi quy Ridge với mô hình có thước đo hiệu năng R², tối ưu đưa kết quả dự báo với độ chính xác lên tới 92% giá trị thật của ICO với trường hợp đồng 0x ICO trong tập kiểm tra.

Trong trường hợp áp dụng thuật toán mạng nơron nhân tạo (ANN), kết quả mô phỏng trình bày mô hình có MAE tối ưu cho dự báo chính xác nhất lên 98% giá trị thật của đồng Crypto20 ICO. Thuật toán ANN tạo ra kết quả dự báo chính xác cao hơn hồi quy Ridge nhưng yêu cầu thời gian huấn luyện lâu hơn. So sánh hai phương pháp dự báo này là thông tin hữu ích để nhà đầu tư lựa chọn phương pháp dự báo phù hợp giữa mô hình hồi quy Ridge và mô hình thuật toán ANN. Khi nhà đầu tư cần xem xét lựa chọn phương pháp dự báo giá ICO cần xem xét tài nguyên phần cứng đang có để lựa chọn phương pháp dự báo phù hợp. Hồi quy Ridge dự báo với sự chính xác 63% tới 92% thấp hơn mô hình thuật toán ANN từ 52% tới 98%. Tuy nhiên, mô hình hồi quy Ridge đơn giản, tốn ít tài nguyên phần cứng để huấn luyện hơn mô hình thuật toán ANN.

5. Kết luận

Nhà đầu tư mong đợi tỉ suất lợi nhuận đạt được kỳ vọng khi đầu tư vào đồng tiền ICO. Tuy nhiên, giá đồng tiền ICO phụ thuộc vào nhiều yếu tố đặc biệt khi mà đồng ICO thường được đầu tư từ lúc chưa phát hành. Vì vậy, tính chất rủi ro và việc quản lý các rủi ro để đầu tư thành công ICO là cần thiết cho nhà đầu tư. Phân tích giá ICO với sự phân tích tương quan mười hai yếu tố tác động giá ICO cho thấy có hai yếu tố là price_usd và price_btc có độ tương quan cao với ngõ ra, các biến còn lại có tương quan yếu hơn nhưng giữa các yếu tố đầu vào lại có tương quan với nhau, vì vậy nghiên cứu xem xét hết tất cả các yếu tố đầu vào này là yếu tố chính tác động giá ICO. Sử dụng mười hai yếu tố ảnh hưởng tới giá ICO làm cơ sở để đánh giá so sánh hai phương pháp dự báo, cho ta thấy được các yếu tố này đóng vai trò quan trọng cho việc phân tích dự báo. Việc phân tích tương quan giữa các yếu tố tác động giá ICO đưa ra kết luận về hiện tượng đa cộng tuyến trong mô hình hồi quy tuyến tính. Điều này dẫn đến làm sai lệch kết quả của mô hình hồi quy bội. Hiện tượng quá khớp khi sử dụng mô hình hồi quy bội cần sử dụng kỹ thuật chính quy hóa, là kỹ thuật giúp giảm lỗi quá khớp bằng cách thêm một thành phần chính quy hóa vào hàm lỗi của mô hình hồi quy bội. Thuật toán hồi quy Ridge là phương thức hồi quy phi tuyến có thể vượt qua các thách thức của vấn đề dữ liệu đang gặp phải mà hồi quy bội không giải quyết được. Hồi quy ridge dùng hệ số lamda để điều chỉnh hệ số hồi quy. Tập dữ liệu 109 ICO được thu thập và tiền xử lý dữ liệu được xem như tập dữ liệu cho hai mô hình dự báo. Trong đó, 85 ICO được sử dụng cho tập huấn luyện và đánh giá, 21 ICO được chọn ngẫu nhiên để ước lượng khả năng của mô hình, tập kiểm tra được chọn ngẫu nhiên 3 ICO. Sau quá trình huấn luyện, ba tiêu chí RMSE tốt nhất, Rsquare tốt nhất, MAE tốt nhất được dùng để tìm mô hình dự báo tối ưu nhất. Kết quả mô phỏng dự báo cho thấy, độ chính xác dự báo giá ICO sau 6 tháng tới 92% giá trị thực tế dùng mô hình hồi quy ridge với tập kiểm tra trong trường hợp đồng 0x ICO. Kết quả mô phỏng dự báo dùng thuật toán ANN cho thấy, độ chính xác dự báo đạt 98% giá trị thực tế với tập kiểm tra trong trường hợp đồng 0x ICO.

LỜI CẢM ƠN

Nghiên cứu này thuộc đề tài được cấp bởi số: T2021-77TĐ được hỗ trợ kinh phí bởi Trường đại học Sư Phạm Kỹ Thuật Thành phố Hồ Chí Minh, Việt Nam.

Tài liệu tham khảo

- Ahmed, I. (2021). Forecasting the early market movement in Bitcoin using Twitter's sentiment analysis: An ensemblebased prediction model. *2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*. doi:10.1109/IEMTRONICS52119.2021.9422647
- Anchaya, C., Nathee, N., Somsak, C., Piyachat, U., Jutharut, C., & Nopasit, C. (2020). Can Tweets predict ICO success? Sentiment analysis for success of ICO whitepaper: Evidence from Australia and Singapore Markets. In *15th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP)* (pp. 01-05). doi:10.1109/iSAI-NLP51646.2020.9376810
- Bugra, C., & Bernat, S. D. (2018). *Analysis of success factors for initial coin offering and automatisaton of whitepaper analysis using text-mining algorithm*. Truy cập ngày 10/09/2021 tại https://upcommons.upc.edu/bitstream/handle/2117/169757/MEMOIRE%20MAJEURE%20DIGITAL_BugraCetingok_BernatSerraDeola.pdf?sequence=1&isAllowed=y

- Burns, L., & Moro, A. (2018). *What makes an ICO Successful? An investigation of the role of ICO characteristics, team quality and market sentiment*. Truy cập ngày 10/09/2021 tại <https://deliverypdf.ssrn.com/delivery.php?ID=963020092123086109012084120066026073033078047010022006094075126002102126011024115007006058039044111113028125000083121067004101123082069048092106127069094099089003125073013044098100124103075010126126100123085122100001090025082089121109027024083125109117&EXT=pdf&INDEX=TRUE>
- Chen, M., Challita, U., Saad, W., Yin, C., & Debbah, M. (2019). Artificial neural networks - Based machine learning for wireless networks: A tutorial. *IEEE Communications Surveys & Tutorials*, 21(4), 3039-3071.
- Deanna, N. G. (2018). Ridge regression and multicollinearity: An in-depth review. *SAS Global Forum*. doi:10.3233/MAS-180446
- Fisch, C. (2019). Initial Coin Offerings (ICOs) to finance new ventures: An exploratory study. *Journal of Business Venturing*, 34(1), 1-22.
- Hartmann, F., Grotto, G., Wang, X., & Lunesu, M. I. (2019). Alternative fundraising, success factors for blockchain-based vs. conventional crowdfunding. *IEEE International Workshop on Blockchain Oriented Software Engineering (IWBOSE)*. doi:10.1109/IWBOSE.2019.8666515
- Hartmann, F., Wang, X., & Lunesu, M. I. (2018). A hierarchical structure model of success factors for (blockchain-based) crowdfunding. In *Blockchain and web 3.0* (pp. 270-308). doi:10.4324/9780429029530-16
- Jigar, P., Sahil, S., Priyank, T., & Kotecha, K. (2015). Predicting stock market index using fusion of machine learning techniques. In *Expert systems with applications* (pp. 2162-2172). doi:10.1016/j.eswa.2014.10.031
- Maria, I. L., & Omar, D. (2020). ICO evaluation websites analysis. *IWBOSE*, 48-56. doi:10.1109/IWBOSE50093.2020.9050259
- Mingxi, L., Guowen, L., Jianping, L., Xiaoqian, Z., & Yinhong Y. (2020) Forecasting the price of Bitcoin using deep learning. *Finance Research Letters*, 40. doi:10.1016/j.frl.2020.101755
- Myalo, A. S., & Glukhov, N. Y. (2019). Factors of success of initial coin offering. *Empirical Evidence from 2016-2019*, 23, 30-49. doi:10.26794/2587-5671-2019-23-5-30-49
- Panin, A., Kemell, K.-K., & Hara, V. (2019). Initial Coin Offering (ICO) as a fundraising strategy: A multiple case study on success. *ICSOB 2019: 10th International Conference of Software Business*. doi:10.1007/978-3-030-33742-1_19
- Yaohao, P., & Pedro, H. M. (2018). The best of two worlds: Forecasting high frequency volatility for cryptocurrencies and traditional currencies with Support Vector Regression. *Expert Systems With Applications*, 97, 177-192. doi:10.1016/j.eswa.2017.12.004

