



# Machine learning

## Cryptocurrency price forecasting

ML & DL

Presenter: Nguyen Xuan Mong

Instructor: Nguyen Van Hau



# Overview

## I. Giới thiệu

## II. Dữ liệu và Quy trình

1. Tiền xử lý dữ liệu

2. Mô hình Machine learning

3. Mô hình Deep learning

## III. Metrics đánh giá và Kết quả



# I. Giới thiệu





## II. Dữ liệu và Quy trình

### 1. Dữ liệu

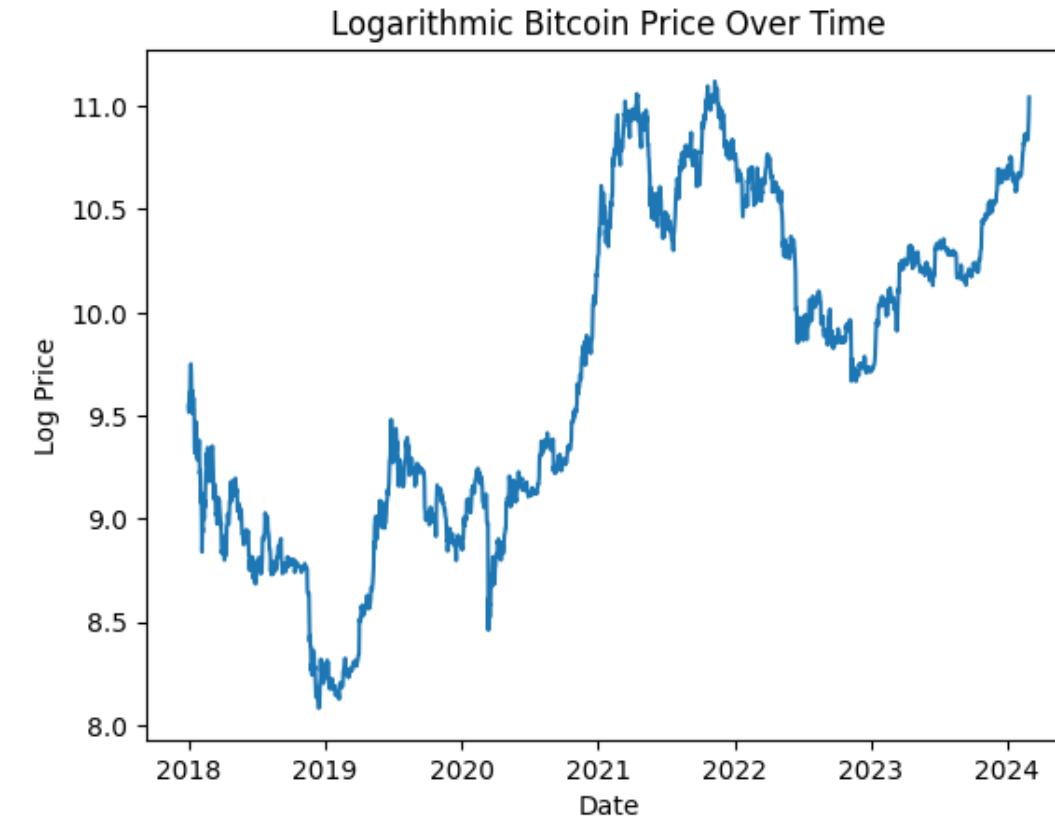
Data columns (total 7 columns):			
#	Column	Non-Null Count	Dtype
0	Date	2250 non-null	object
1	Open	2250 non-null	float64
2	High	2250 non-null	float64
3	Low	2250 non-null	float64
4	Close	2250 non-null	float64
5	Volume	2250 non-null	float64
6	Market Cap	2250 non-null	float64

	Open	High	Low	Close	Volume	Market Cap
count	2250.000000	2250.000000	2250.000000	2250.000000	2.250000e+03	2.250000e+03
mean	22343.137322	22854.586331	21805.439057	22365.854563	4.247632e+10	4.208112e+11
std	16248.746436	16647.585143	15827.388275	16269.696713	3.449772e+10	3.120267e+11
min	3239.810072	3269.710130	3188.542617	3236.700910	2.039620e+09	5.632816e+10
25%	8442.585971	8645.723472	8197.471602	8446.854302	1.577712e+10	1.482500e+11
50%	18209.195785	18682.766050	17652.921135	18293.466325	3.493959e+10	3.400000e+11
75%	34568.277500	35310.321970	33510.794255	34568.167500	6.222844e+10	6.547500e+11
max	67510.919420	68769.955370	66358.933600	67500.024740	2.120000e+11	1.270000e+12



## II. Dữ liệu và Quy trình

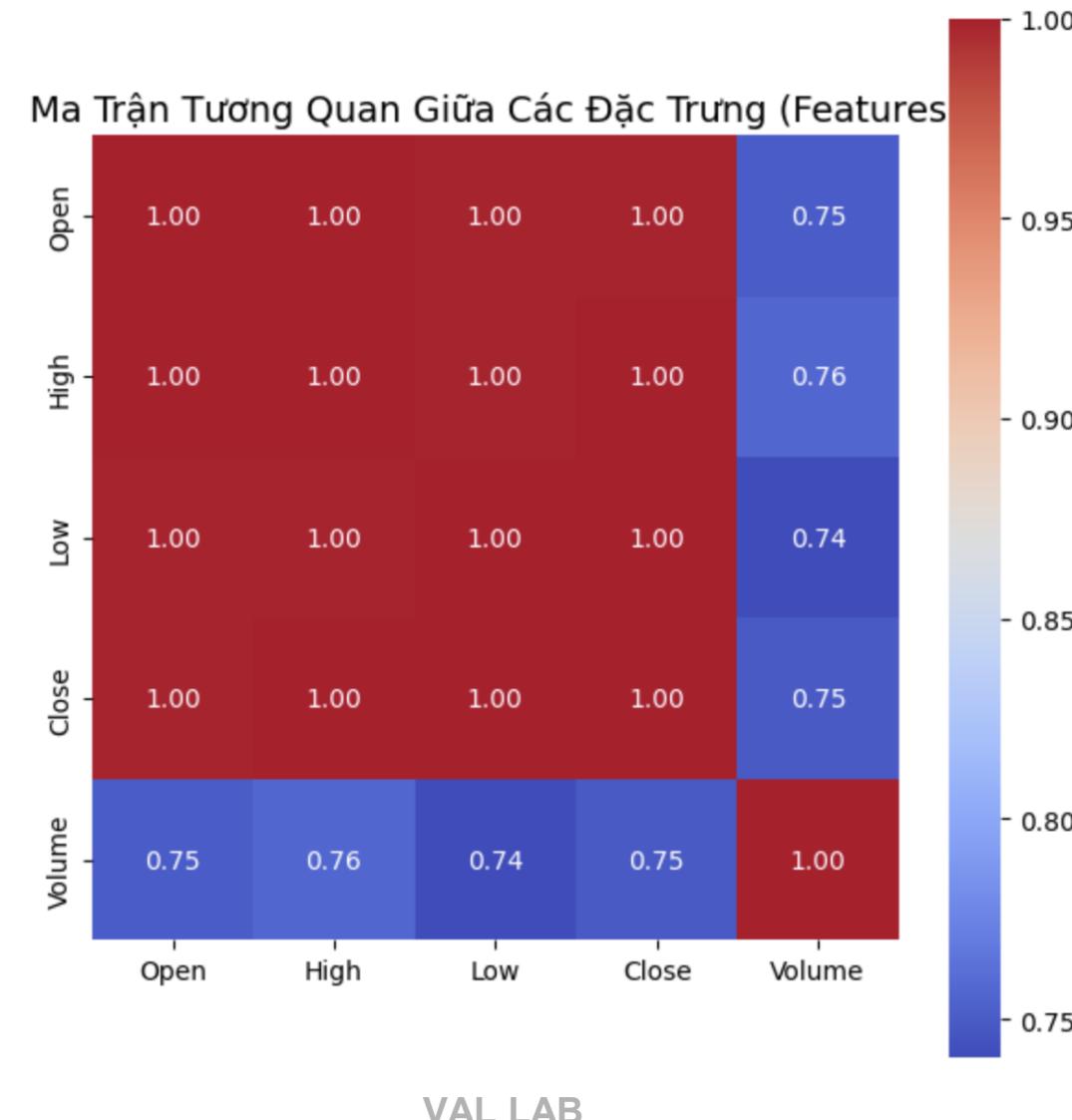
### 1. Dữ liệu





## II. Dữ liệu và Quy trình

### 1. Dữ liệu





## II. Dữ liệu và Quy trình

### 1. Tiền xử lý dữ liệu

- + Cách tiếp cận 1 sử dụng các đặc trưng: **open,close,high,low**
- Diền dữ liệu còn thiếu bằng dữ liệu ngày trước đó
- Chuẩn hoá dữ liệu Min-Max Scaler
- Tạo dữ liệu lag cho training model



## II. Dữ liệu và Quy trình

### 1. Tiền xử lý dữ liệu

Ngày	Giá thực tế (USD)
Thứ 2	40,000
Thứ 3	42,000
Thứ 4	41,000
Thứ 5	45,000
Thứ 6	44,000

Đặc trưng (Input)	Mục tiêu (Target)
[40,000, 42,000]	<b>41,000</b>
[42,000, 41,000]	<b>45,000</b>
[41,000, 45,000]	<b>44,000</b>



## II. Dữ liệu và Quy trình

### 1. Tiền xử lý dữ liệu

+ Cách tiếp cận 2 tạo thêm đặc trưng (features engineering):

- Vẫn như cách tiếp cận 1 thêm bước sau:

• **VVR (Volume-to-Volatility Ratio)**: \* *Cách tính:* Lấy Khối lượng giao dịch chia cho biên độ dao động giá (High - Low).

• *Ý nghĩa:* Nó đo lường cường độ giao dịch so với sự biến động giá. Một chỉ số VVR cao cho thấy thị trường có tính thanh khoản tốt, giá ổn định hơn dù khối lượng giao dịch lớn.

$$VVR_t = \frac{Volume_t}{(High_t - Low_t + \epsilon)}, \text{ where } \epsilon = 10^{-6}$$



## II. Dữ liệu và Quy trình

### 1. Tiền xử lý dữ liệu

+ Cách tiếp cận 2 tạo thêm đặc trưng (features engineering):

- Vẫn như cách tiếp cận 1 thêm bước sau:

**VWAP (Volume-Weighted Average Price):**

• *Cách tính:* Trung bình giá giao dịch trong ngày được trọng số theo khối lượng giao dịch tại từng mức giá.

• *Ý nghĩa:* Nó cung cấp một mức giá "công bằng" hơn so với giá đóng cửa đơn thuần, giúp mô hình nhận diện được các vùng giá có sự tham gia lớn của các tổ chức hoặc "cá voi".

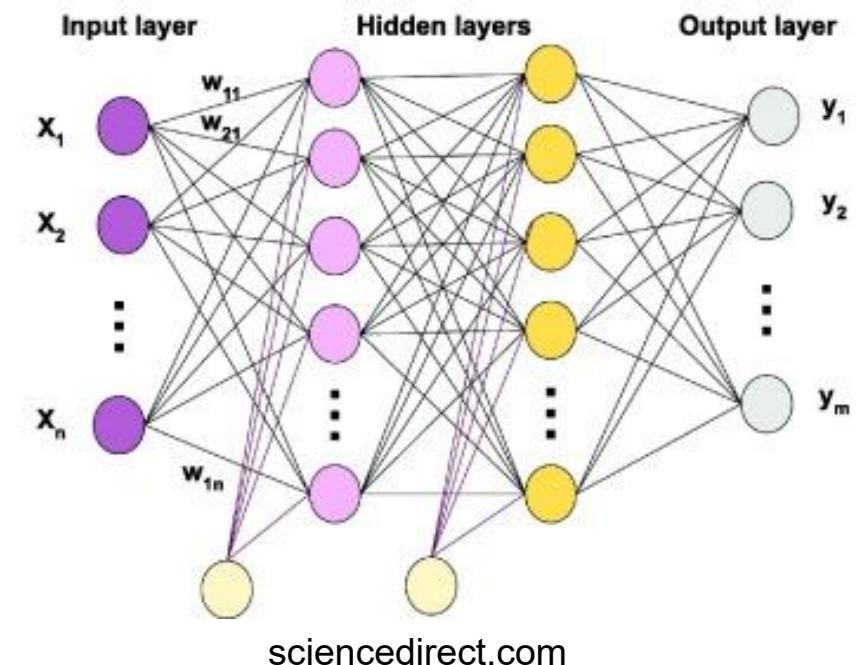
$$VWAP_t = \frac{\sum (TP_t \cdot Volume_t)}{\sum Volume_t}, \text{ where } TP_t = \frac{High_t + Low_t + Close_t}{3}$$



## II. Mô hình Machine learning

### Multi layer Perceptron

Mô hình	Số lớp ẩn	Chi tiết các tầng (Units)
MLP_Level1	1	32
MLP_Level2	2	64 → 32
MLP_Level3	3	128 → 64 → 32

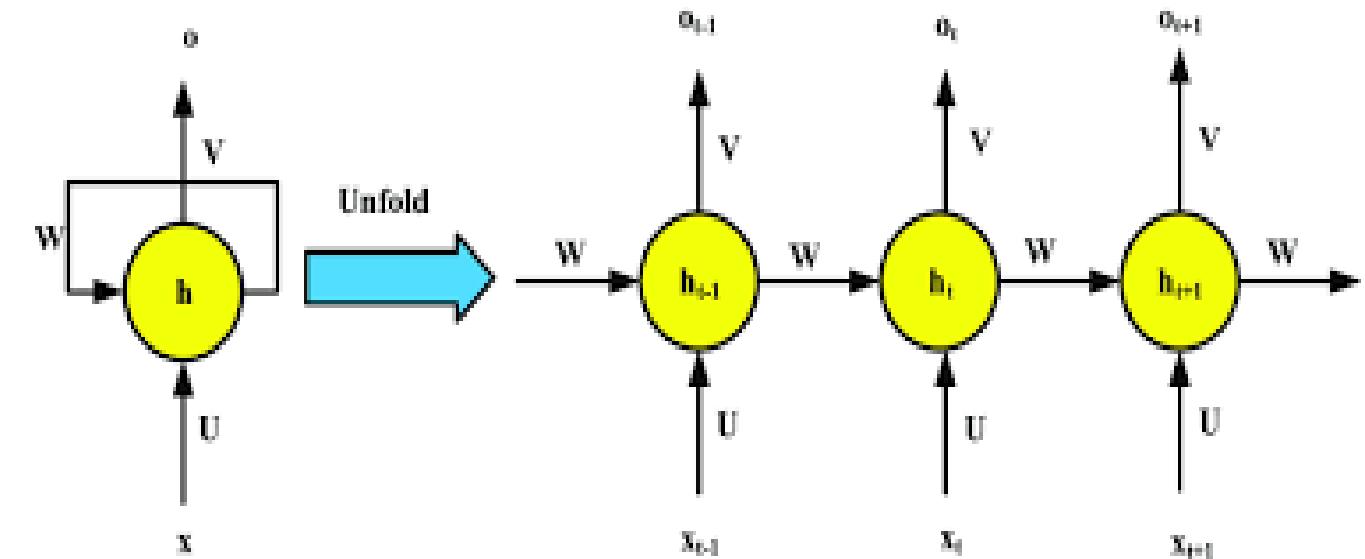




## II. Mô hình Deeplearning

### 1. Recurrent Neural Network

Mô hình	Số lớp RNN	Số lớp ẩn FC	Chi tiết cấu trúc (Units)
RNN_Level1	1	1	$\text{RNN}(64) \rightarrow \text{FC}(32) \rightarrow \text{Out}$
RNN_Level2	2	2	$\text{RNN}(96) \rightarrow \text{FC}(64) \rightarrow \text{FC}(32) \rightarrow \text{Out}$
RNN_Level3	3	3	$\text{RNN}(128) \rightarrow \text{FC}(96) \rightarrow \text{FC}(64) \rightarrow \text{FC}(32) \rightarrow \text{Out}$



ResearchGate



# Metrics và Kết quả

## 1. Metrics

- RMSE
- MAE
- R^2

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

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n}$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$



# Metrics và Kết quả

## Kết quả

### 1. Với dữ liệu gốc chưa sử dụng chuẩn hoá

Metrics	Model (MLP - lag 5)		
	1	2	3
RMSE	13367.831	14080.356	14781.496
MAE	12762.544	13453.518	14134.001
R^2	-1.388	-1.649	-1.919

Metrics	Model (SimpleRNN - lag 5)		
	1	2	3
RMSE	13567.214	13682.152	14754.249
MAE	10944.022	11067.614	12192.581
R^2	-1.459	-1.501	-1.909



# Metrics và Kết quả

## 1. Metrics

- Kết quả sau khi chuẩn hoá MinMaxScaler

	Model (MLP - lag 5)		
Metrics	1	2	3
RMSE	<b>0.0227</b>	0.0406	0.0476
MAE	0.0228	<b>0.0273</b>	0.0356
R <sup>2</sup>	<b>0.9575</b>	0.9086	0.8746

	Model (MLP - lag 5)		
Metrics	1	2	3
RMSE	0.096	0.204	0.253
MAE	0.094	0.164	0.214
R <sup>2</sup>	0.49	-1.312	-2.541

VAL LAB



# Metrics và Kết quả

## 1. Metrics

- Kết quả sau khi chuẩn hoá MinMaxScaler - Dữ liệu trước dịch covid 19

	Model (MLP - lag 10)		
Metrics	1	2	3
RMSE	0.033	<b>0.032</b>	0.043
MAE	0.025	<b>0.024</b>	0.036
R^2	0.841	<b>0.852</b>	0.7396

	Model (MLP - lag 5)		
Metrics	1	2	3
RMSE	0.028	<b>0.029</b>	0.043
MAE	0.020	<b>0.020</b>	0.036
R^2	0.857	<b>0.884</b>	0.742

	Model (SimpleRNN - lag 10)		
Metrics	1	2	3
RMSE	0.032	<b>0.027</b>	0.122
MAE	0.024	<b>0.019</b>	0.105
R^2	0.853	<b>0.897</b>	-1.103

	Model (SimpleRNN - lag 5)		
Metrics	1	2	3
RMSE	<b>0.027</b>	0.032	0.052
MAE	<b>0.019</b>	0.026	0.044
R^2	<b>0.898</b>	0.851	0.625



# Metrics và Kết quả

## 1. Metrics

- Kết quả sau khi chuẩn hoá MinMaxScaler - Dữ liệu sau dịch covid 19

Metrics	Model (MLP - lag 10)		
	1	2	3
RMSE	<b>0.063</b>	0.084	0.079
MAE	<b>0.050</b>	0.069	0.059
R^2	<b>0.752</b>	0.564	0.606

Metrics	Model (SimpleRNN - lag 10)		
	1	2	3
RMSE	<b>0.022</b>	0.149	0.158
MAE	<b>0.016</b>	0.104	0.109
R^2	<b>0.969</b>	-0.389	-0.560



# Metrics và Kết quả

## 1. Metrics

- Kết quả với feature engineering:

Metrics	Model (MLP - lag 10)		
	1	2	3
RMSE	0.197	0.207	0.206
MAE	0.155	0.167	0.165
R^2	-1.179	-1.382	-1.372

Metrics	Model (SimpleRNN - lag 10)		
	1	2	3
RMSE	0.250	0.209	0.183
MAE	0.212	0.168	0.140
R^2	-2.508	--1.440	-0.884



# Metrics và Kết quả

## 1. Metrics

- Kết quả với feature engineering:

Metrics	Model (MLP - lag 10)		
	1	2	3
RMSE	0.197	0.207	0.206
MAE	0.155	0.167	0.165
R^2	-1.179	-1.382	-1.372

Metrics	Model (SimpleRNN - lag 10)		
	1	2	3
RMSE	0.250	0.209	0.183
MAE	0.212	0.168	0.140
R^2	-2.508	--1.440	-0.884



# Metrics và Kết quả

## 1. Metrics

- Kết quả với feature engineering:

Metrics	Model (MLP - lag 10)		
	1	2	3
RMSE	0.197	0.207	0.206
MAE	0.155	0.167	0.165
R^2	-1.179	-1.382	-1.372

Metrics	Model (SimpleRNN - lag 10)		
	1	2	3
RMSE	0.250	0.209	0.183
MAE	0.212	0.168	0.140
R^2	-2.508	--1.440	-0.884



# References

- [1] M. Z. Islam *et al.*, "Cryptocurrency price forecasting using machine learning: Building intelligent financial prediction models," *Int. J. Account. Econ. Stud.*, vol. 12, no. 3, pp. 255–268, Jul. 2025, doi: 10.14419/s0pktr58.
- [2] H. Sebastião and P. Godinho, "Forecasting and trading cryptocurrencies with machine learning under changing market conditions," *Financ. Innov.*, vol. 7, no. 1, pp. 1–30, 2021, doi: 10.1186/s40854-020-00217-x.
- [3] A. Mazinani, L. Davoli, and G. Ferrari, "Deep learning algorithms for cryptocurrency price prediction: A comparative analysis," *Distrib. Ledger Technol.: Res. Pract.*, vol. 4, no. 1, Art. no. 5, pp. 1–38, Feb. 2025, doi: 10.1145/3699966.



**Thank You !**  
**Question?**