Stock Price Analysis and Prediction

Trần Nguyễn Mỹ Anh
20235474

Nguyễn Khánh Ly
20235600

Trần Lê Hạ Đan
20235483

Nguyễn Thu Huyền
20235507

CQG NYSE and Nasdaq Performance Dashboard

11:00:47 AM

Description	Net Change	% Net Change	Open	High	Low
S&P 500 (SPC)	2159.75	3.06	0.13%	2156.81	2160.58	2152.56
DJ Industrial Average (DJI)	18337.50	22.40	0.13%	18313.00	18354.10	18283.20
NYSE Financial Index (NYK)	6100.08	42.96	0.71%	6057.27	6100.45	6057.27
Biotechnology Index (BTK)	3465.00	18.40	0.53%	3433.21	3467.29	3430.07
NYSE Healthcare Index (NYPID)	13057.99	-73.20	-0.56%	13131.19	13131.19	13043.57
S&P 500 Select Sector Energy Index (IXE)	666.71	7.37	1.12%	660.13	667.45	657.02
Bank of America Corp (BAC)	14.46	0.33	2.34%	14.11	14.47	14.09
General Electric Co (GE)	31.13	0.08	0.26%	31.03	31.24	31.02
SPDR S&P 500 ETF Trust (SPY)	215.86	0.33	0.15%	215.48	215.96	215.13
E-Mini S&P 500, Sep 16 (EP)	2154.25	1.75	0.07%	2152.25	2155.50	2145.25
Crude Light (Globex), Sep 16 (CLE)	40.64	1.13	2.86%	39.70	40.68	39.19
NY Harbor ULSD, Sep 16 (HOE)	1.29	0.03	2.40%	1.27	1.29	1.26
RBOB Gasoline (Globex), Sep 16 (RBE)	1.34	0.03	2.36%	1.31	1.34	1.30
Gold (Globex), Dec 16 (GCE)	1363.10	-9.50	-0.69%	1370.50	1373.40	1360.60
DAX Index, Sep 16 (DD)	10166.00	20.00	0.20%	10129.00	10187.00	10088.00
Euro STOXX 50, Sep 16 (DSX)	2906.00	0.00	0.00%	2902.00	2919.00	2888.00
10yr US Treasury Notes (Globex), Sep 16 (TYA)	132.39	-0.16	-0.12%	132.41	132.63	132.27

5-minute Charts

S&P 500 (SPC)

NASDAQ Composite Index (NQCOMP)

Description	Last Price	Net Change	% Net Change	Open	High	Low
NASDAQ Composite Index (NQCOMP)	5146.17	8.44	0.16%	5133.24	5149.22	5128.44
NASDAQ-100 Index (NDX)	4721.87	2.66	0.06%	4716.58	4730.34	4712.10
NASDAQ Financial 100 (IXF)	3344.86	25.12	0.76%	3323.75	3345.02	3322.53
NASDAQ Biotechnology (NBI)	3099.96	16.25	0.53%	3072.36	3101.62	3069.10
Nasdaq Health Care Index (NHC)	685.81	2.75	0.40%	680.88	686.15	680.22
PHLX Oil Service Sector (OSX)	157.54	1.66	1.06%	155.81	158.48	154.98
Sirius XM Holdings Inc. (SIRI)	4.25	0.02	0.47%	4.21	4.27	4.21
Apple Inc (AAPL)	105.45	0.97	0.93%	104.85	105.84	104.77
Tesla Motors, Inc. (TSLA)	226.58	-0.62	-0.27%	227.20	229.69	225.10
Micron Technology Inc (MU)	13.56	0.15	1.12%	13.28	13.57	13.28
Powershares QQQ Trust (QQQ)	115.06	0.10	0.09%	114.88	115.27	114.80
Cisco Systems Inc (CSCO)	30.63	0.01	0.03%	30.53	30.66	30.51
Facebook, Inc. (FB)	122.59	-0.50	-0.41%	123.09	123.92	122.31
iShares 20+ Year Treasury Bond ETF (TLT)	138.21	-0.12	-0.09%	138.60	138.78	137.91
Intel Corporation (INTC)	34.20	-0.36	-1.04%	34.10	34.28	34.10
QUALCOMM Inc (QCOM)	60.76	0.16	0.26%	60.59	61.11	60.50
American Airlines Group Inc. (AAL)	33.39	-0.12	-0.36%	33.49	33.96	33.17

Description	Last Price	Net Change	% Net Change	Tdy Vol 21d-Vol	Open	High	Low
CQG Top Ten Percentage Performers on the NYSE							
Genworth Financial Inc. (GNW)	3.56	0.81	29.45%	300%	3.43	3.58	3.32
3D Systems Corporation (DDD)	14.53	2.36	19.29%	182%	13.00	14.58	12.77
American Vanguard Corp (AVD)	17.50	2.58	16.74%	227%	16.05	17.79	15.61
Meritor Inc. (MTOR)	9.23	1.31	16.39%	174%	8.36	9.33	8.23
Nationstar Mortgage Holdings, Inc. (NSM)	13.71	1.93	16.28%	159%	12.81	13.84	12.61
Comstock Resources Inc (CRK)	3.37	0.45	15.41%	232%	2.90	3.47	2.79
Walker & Dunlop, Inc. (WD)	27.04	3.29	14.14%	127%	26.05	27.04	25.80
Fitbit, Inc. (FIT)	14.91	1.75	13.30%	266%	14.11	14.96	13.92
Cheetah Mobile Inc. (CMCM)	11.92	1.34	12.67%	187%	10.50	12.05	10.47
ENPRO Industries Inc (NPO)	50.70	4.60	9.98%	207%	46.50	52.00	46.50
CQG Bottom Ten Percentage Performers on the NYSE							
The Rubicon Project, Inc. (RUBI)	9.24	-4.43	-32.41%	1186%	9.50	9.70	9.00
Kate Spade Company (KATE)	15.92	-4.22	-20.95%	571%	16.00	16.50	15.66
ARC Document Solutions Inc (ARC)	3.39	-0.43	-11.26%	136%	3.61	3.61	3.35
Capital Senior Living Corp (CSU)	17.03	-2.05	-10.74%	324%	18.45	18.48	16.75
Stone Energy Corp (SGY)	10.06	-1.06	-9.53%	47%	8.50	10.47	8.42
Myers Industries Inc (MYE)	13.44	-1.36	-9.19%	45%	13.92	13.94	13.20
PROS Holdings, Inc. (PRO)	16.86	-1.94	-10.46%	112%	17.88	18.15	16.25
Tableau Software Inc (DATA)	51.49	-4.88	-8.71%	323%	52.87	53.50	50.27
Themon Group Holdings, Inc. (THR)	18.20	-1.40	-7.14%	212%	17.87	18.25	16.66
DRDGOLD Limited (DRD)	8.05	-0.60	-6.94%	55%	8.50	8.50	8.02

NYSE Total Volume	NASDAQ Total Volume
331,900	894,953
NYSE Up Volume	NASDAQ Up Volume
220,713	556,437
NYSE Down Volume	NASDAQ Down Volume
108,706	270,096
NYSE Unchanged Vol	NASDAQ Unchanged Vol
2,479	68,420
NYSE Adv Issues	NASDAQ Adv Issues
1,840	1,577
NYSE Dec Issues	NASDAQ Dec Issues
1,132	1,065
NYSE Unc Issues	NASDAQ Unc Issues
74	93
NYSE Tick	NASDAQ Tick
-15	11

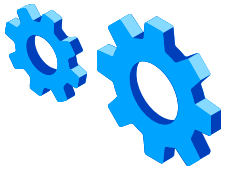
Description	Last Price	Net Change	% Net Change	Tdy Vol 21d-Vol	Open	High	Low
CQG Top Ten Percentage Performers on the NASDAQ							
Magellan Petroleum Corp (MPET)	2.92	1.72	143.33%	13422%	1.92	5.00	1.70
China Natural Resources, Inc. (CHNR)	2.45	0.82	47.59%	8815%	1.80	3.17	1.80
Sizmek Inc (SZMK)	3.86	1.20	45.11%	3922%	3.87	3.87	3.85
Cesca Therapeutics Inc. (KOOL)	5.75	1.30	29.21%	627%	6.70	7.39	5.63
Forbes Energy Services Ltd (FES)	0.16	0.03	23.08%	161%	0.14	0.19	0.14
Stanley Furniture Co Inc (STLY)	2.98	0.53	21.63%	4173%	3.25	3.25	2.95
Neovasc Inc. (NVCN)	0.66	0.12	22.22%	315%	0.54	0.72	0.54
Ocean Power Technologies (OPTT)	8.94	1.48	19.84%	42%	7.20	9.12	7.20
Memorial Production Partners LP (MEMP)	1.66	0.27	19.42%	161%	1.40	1.69	1.40
Big 5 Sporting Goods Co (BGFV)	12.44	1.97	18.82%	339%	11.80	12.82	11.54
CQG Bottom Ten Percentage Performers on the NASDAQ							
Cray Inc (CRAY)	22.04	-9.24	-29.54%	1046%	25.06	25.19	21.80
Electro Scntfc Inds Inc (ESIO)	4.85	-2.00	-29.20%	551%	5.81	5.89	4.74
The Chefs' Warehouse, Inc. (CHEF)	11.51	-4.64	-28.73%	1038%	14.00	14.39	10.78
Spherix Incorporated (SPEX)	1.52	-0.58	-27.62%	232%	1.60	1.69	1.45
Crocs Inc (CROX)	8.31	-2.72	-24.45%	605%	8.79	9.50	8.10
Sino Global Shipping America LTD (SINO)	1.79	-0.45	-20.09%	86%	2.24	2.26	1.77
Globus Maritime Limited (GLBS)	0.67	-0.15	-18.29%	65%	0.81	0.83	0.63
Cosi Inc (COSI)	0.23	-0.05	-17.86%	94%	0.26	0.26	0.22
USA Truck (USAK)	15.73	-3.62	-18.20%	83%	16.42	18.10	15.51
Odyssey Marine Exploration Inc (OMEX)	3.64	-0.75	-17.08%	469%	4.27	4.31	3.56

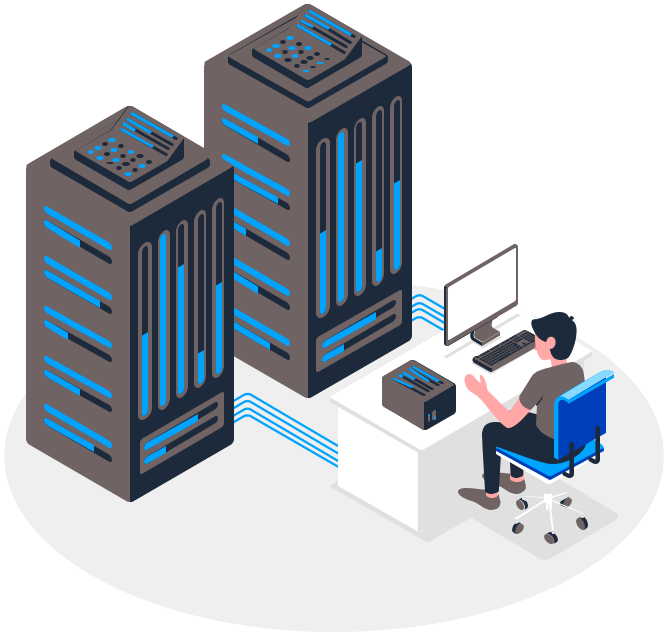
Designed by Thom Hartle

Rank

To better manage performance the dashboard ranking does not update on every quote, but once every second. Click "Rank" to force an update if the percent net change column is out of order.

Table of contents





01

Data crawling and analysis

We want to get data only for Vietnamese markets

Data Crawling

- Data collected from VCI Dashboard (Viet Capital securities)
- Contains data up to May 10th, 2025
- Features: Open, High, Low, Close, Volume, Trading Date, Stock Code

	Open	High	Low	Close	Volume	TradingDate	Code
0	16200.0	19800.0	15750.0	19350.0	278800	2010-07-06	LCS
1	19400.0	19400.0	19400.0	19400.0	30700	2010-07-07	LCS
2	20750.0	20750.0	18450.0	18900.0	192800	2010-07-08	LCS
3	18900.0	18900.0	18900.0	18900.0	1000	2010-07-09	LCS
4	17600.0	18900.0	17600.0	18230.0	133500	2010-07-12	LCS
5	18810.0	18950.0	18000.0	18450.0	50500	2010-07-13	LCS
6	18410.0	18410.0	17780.0	17780.0	52000	2010-07-14	LCS

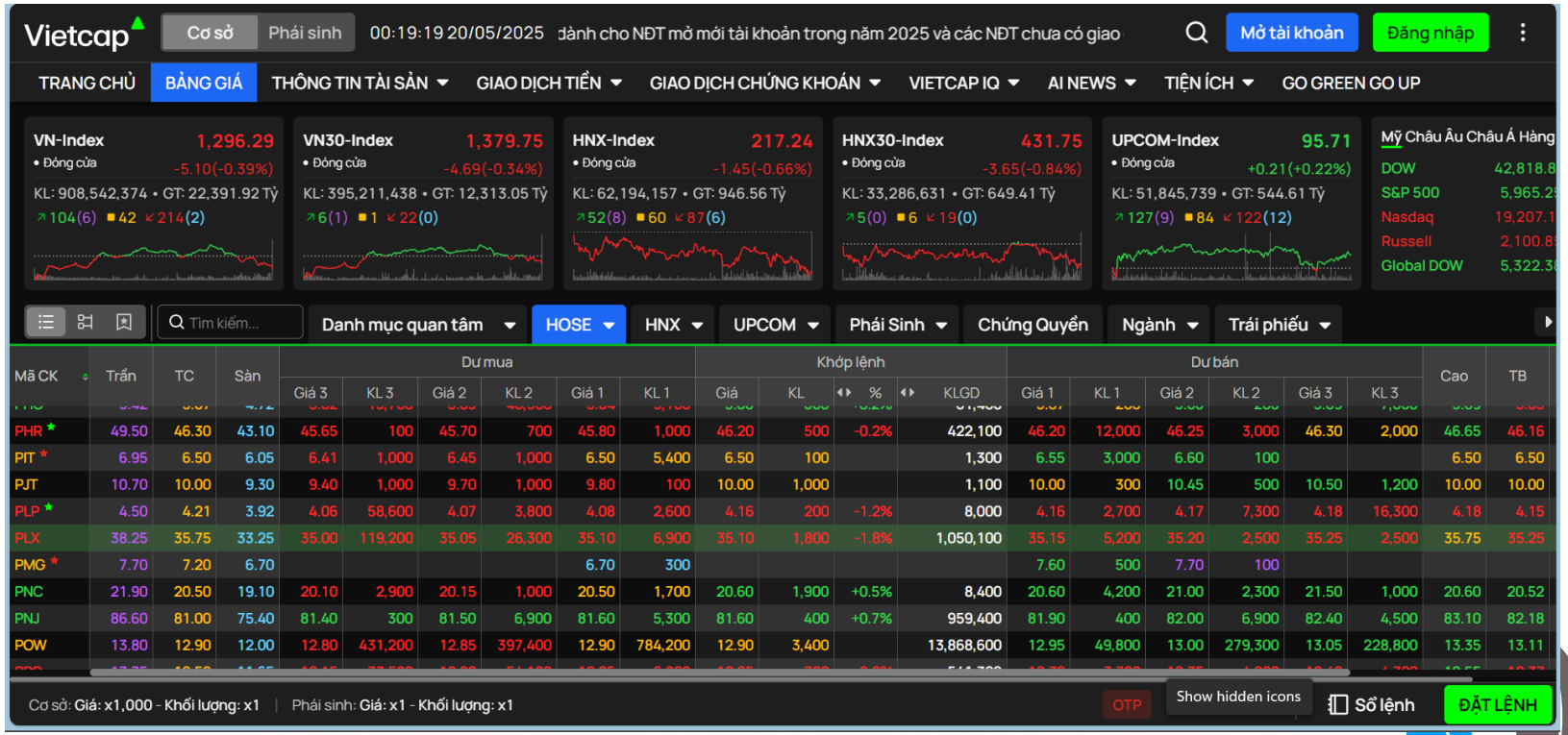
Data Crawling

```
Vnstock.stock(symbol, source).quote.history(
                                    start, end, interval)
```

vnstock library to handle the API with:

- symbol: stock code
- interval: time range
- source: "VCI" or "TCBS"
- start, end: the starting time and ending time

Data Crawling



Mã CK	Trần	TC	Sàn	Dư mua						Khớp lệnh				Dư bán						Cao	TB
				Giá 3	KL 3	Giá 2	KL 2	Giá 1	KL 1	Giá	KL	%	KLGD	Giá 1	KL 1	Giá 2	KL 2	Giá 3	KL 3		
PHR	49.50	46.30	43.10	45.65	100	45.70	700	45.80	1,000	46.20	500	-0.2%	422,100	46.20	12,000	46.25	3,000	46.30	2,000	46.65	46.16
PIT	6.95	6.50	6.05	6.41	1,000	6.45	1,000	6.50	5,400	6.50	100		1,300	6.55	3,000	6.60	100			6.50	6.50
PJT	10.70	10.00	9.30	9.40	1,000	9.70	1,000	9.80	100	10.00	1,000		1,100	10.00	300	10.45	500	10.50	1,200	10.00	10.00
PLP	4.50	4.21	3.92	4.06	58,600	4.07	3,800	4.08	2,600	4.16	200	-1.2%	8,000	4.16	2,700	4.17	7,300	4.18	16,300	4.18	4.15
PLX	38.25	35.75	33.25	35.00	119,200	35.05	26,300	35.10	6,900	35.10	1,800	-1.8%	1,050,100	35.15	5,200	35.20	2,500	35.25	2,500	35.75	35.25
PMG	7.70	7.20	6.70					6.70	300					7.60	500	7.70	100				
PNC	21.90	20.50	19.10	20.10	2,900	20.15	1,000	20.50	1,700	20.60	1,900	+0.5%	8,400	20.60	4,200	21.00	2,300	21.50	1,000	20.60	20.52
PNJ	86.60	81.00	75.40	81.40	300	81.50	6,900	81.60	5,300	81.60	400	+0.7%	959,400	81.90	400	82.00	6,900	82.40	4,500	83.10	82.18
POW	13.80	12.90	12.00	12.80	431,200	12.85	397,400	12.90	784,200	12.90	3,400		13,868,600	12.95	49,800	13.00	279,300	13.05	228,800	13.35	13.11

Data Analysis: 4 sectors—2 most traded stocks

Construction

- LCS: Licogi JSC
- PTC: Icapital Investment JSC

Technology

- VGI: Viettel Global Investment
- ITD: Tien Phong Technology

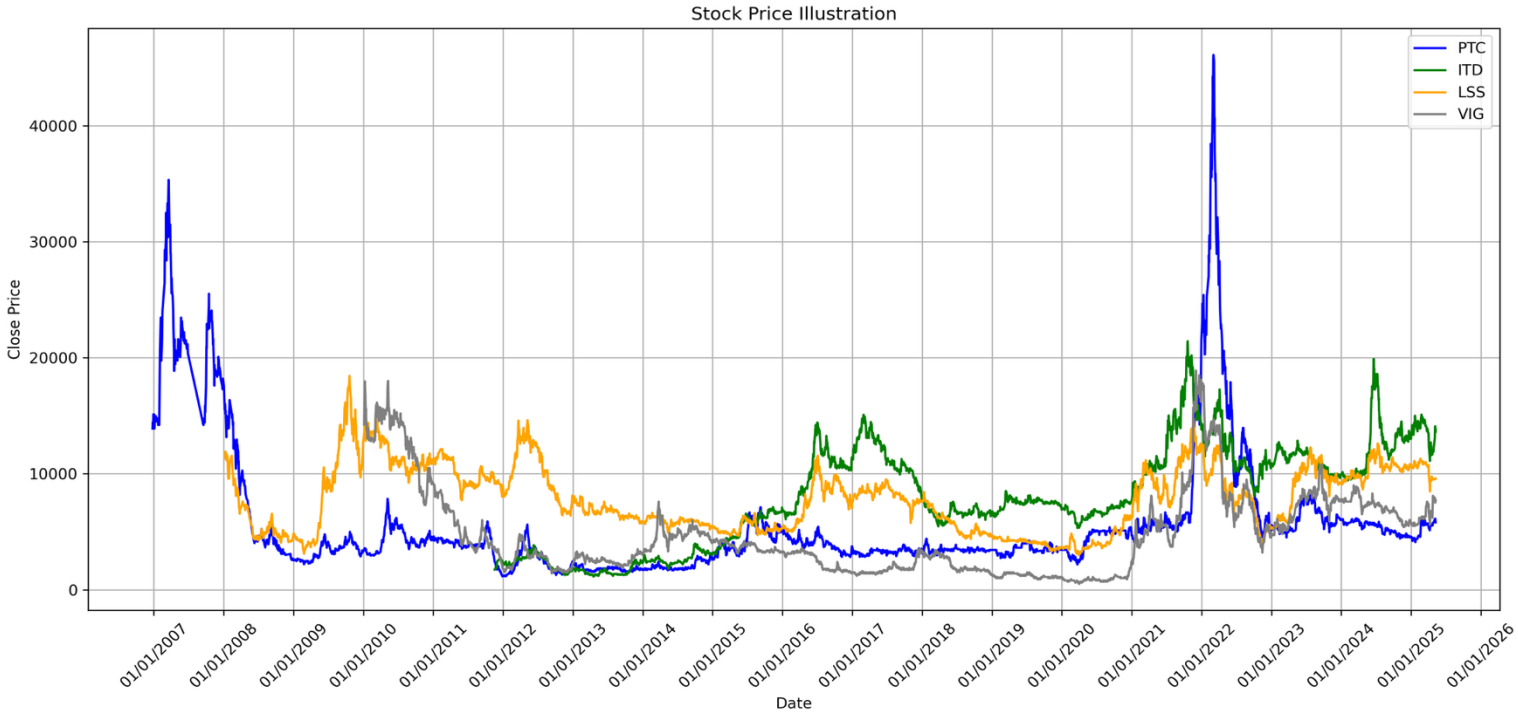
Consumer

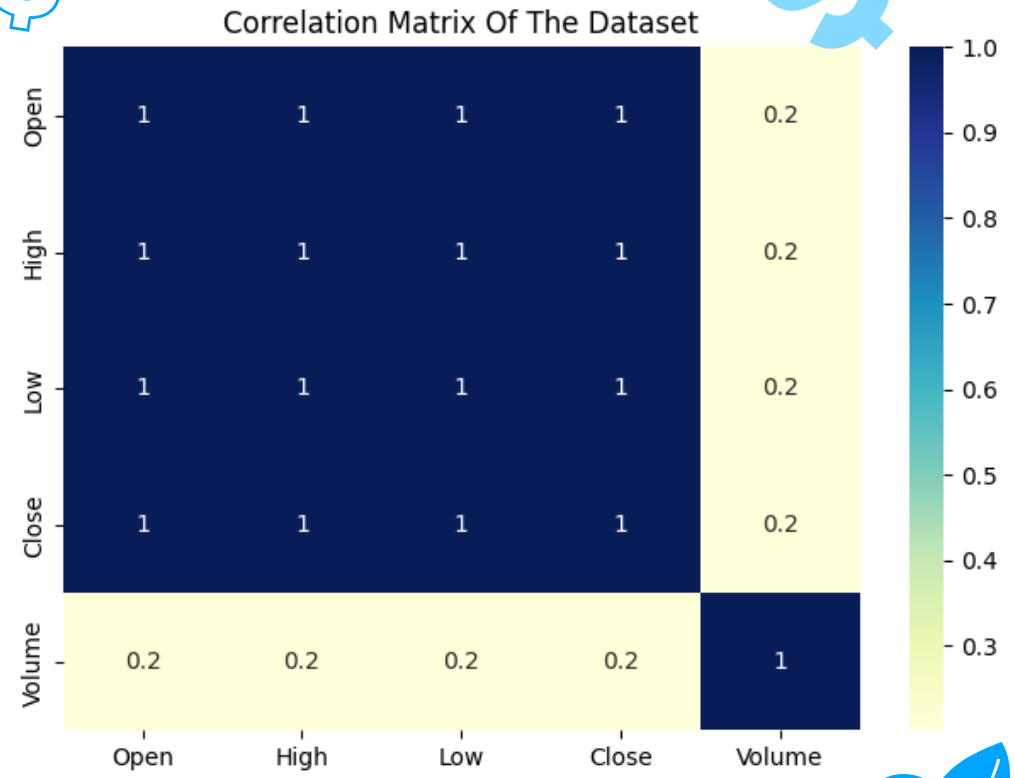
- LSS: Lam Son Sugar
- PLX: Vietnam National Petroleum Group

Finance

- TCB: Vietnam Technological and Commercial Joint Stock Bank
- VIG: Vietnam Financial Investment Securities Corporation

Data Analysis





Feature Selection

- We only choose Close feature since the importance of all four first features are equal.
- Volume seems unrelated. However, we will use Volume in another way.

RSI Score and relation to the volume

- The relative strength index (RSI) is a momentum indicator used in technical analysis. RSI measures the speed and magnitude of a security's recent price changes to evaluate overvalued or undervalued conditions in the price of that security.
- To calculate RSI, we define upward (U) and downward (D) indicators, which are:

$$U_t = \begin{cases} P_t - P_{t-1} & if\ P_t > P_{t-1}, \\ 0 & otherwise. \end{cases}$$

$$D_t = \begin{cases} -P_t + P_{t-1} & if\ P_t < P_{t-1}, \\ 0 & otherwise. \end{cases}$$

RSI Score and relation to the volume

- Then up_t and down_t are average numbers of upward moves and downward moves of closing price of the past n days:

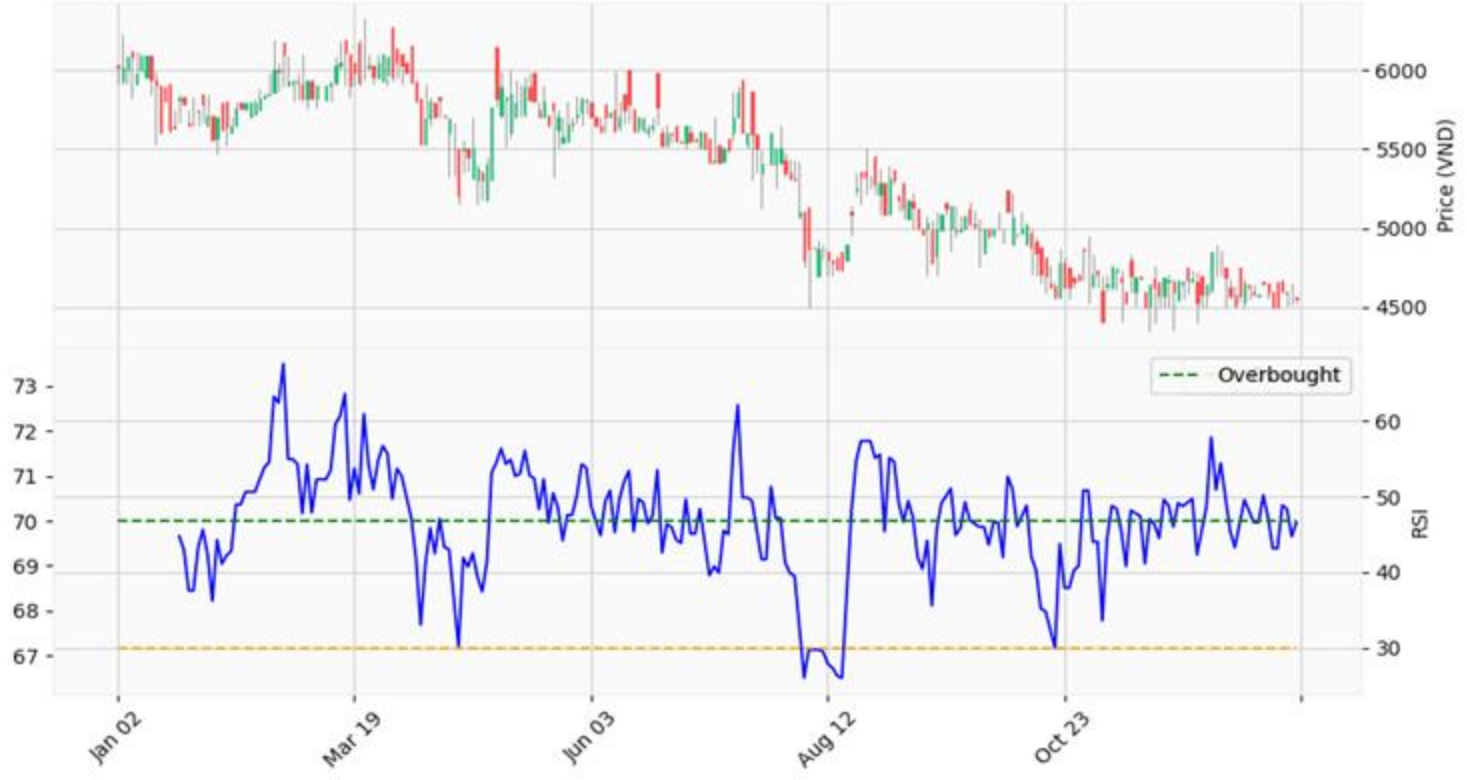
$$\text{up}_t = \frac{s}{n+1} \times U_t + \left(1 - \frac{s}{n+1}\right) \times up_{t-1}$$

$$\text{down}_t = \frac{s}{n+1} \times D_t + \left(1 - \frac{s}{n+1}\right) \times down_{t-1}$$

- Finally, we get the RSI score calculated by:

$$\mathbf{RSI}_t = 100 - \frac{100}{1 + \frac{\text{up}_t}{\text{down}_t}}$$

RSI Score and relation to the volume

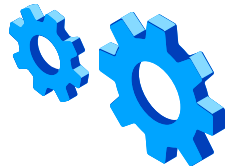


02

Algorithms

A. Random Forest Regressor (RFR)

- The RFR algorithm is trained using historical stock data
- Implemented using the scikit-learn library in Python
- The random forest means data about data estimator. It fits a number decision trees on various sub samples of the given data.
  - ➢ Control over-fitting.
  - ➢ Improve the predictive accuracy.



A. Random Forest Regressor (RFR)

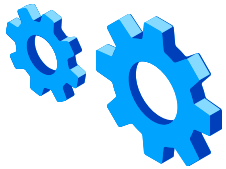
Split data into **Features (X)** (all input columns except 'Close' and 'Code') and **Target (y)** (Close), and further split into training and testing sets(8:2)

- Step 1: From the dataset pick N random records.
- Step 2: Based on N records, build a decision tree.
- Step 3a: Choose the number of trees and repeat steps 1, 2.
- Step 3b: In case of a regression problem, for a new record, each tree in the forest predicts a value for Y (output).

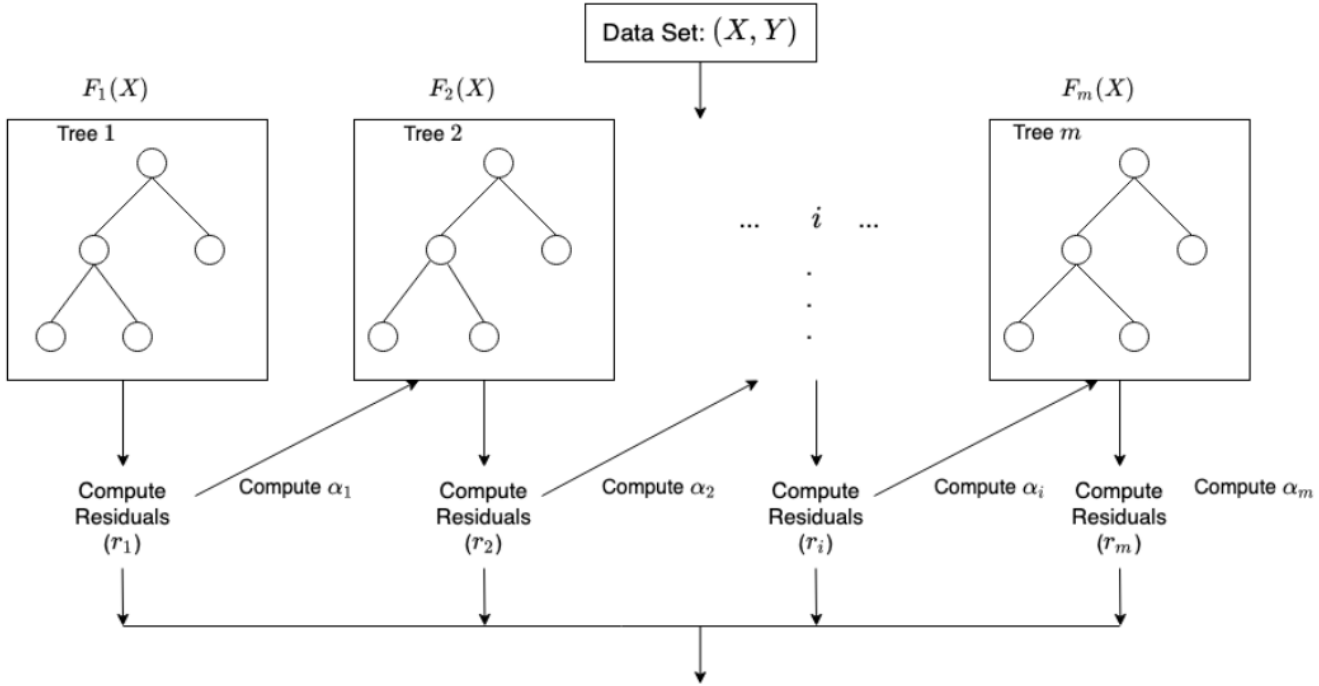
B. Extreme Gradient Boosting (XGBoost)

- Boosting:
  - Boosting is an ensemble model
  - Build a strong model from the number of weak models
- Gradient boosting:
  - In gradient boosting, each predictor corrects its predecessor's error

B. Extreme Gradient Boosting (XGBoost)



- Extreme Gradient Boosting:
  - An implementation of Gradient Boosted decision trees
  - Decision trees are created in sequential form
  - Weights play an important role in XGBoost



$$F_m(X) = F_{m-1}(X) + \alpha_m h_m(X, r_{m-1}),$$

where α_i, and r_i are the regularization parameters and residuals computed with the i^{th} tree respectfully, and h_i is a function that is trained to predict residuals, r_i using X for the i^{th} tree. To compute α_i we use the residuals computed, r_i and compute the following: $arg \min_{\alpha} = \sum_{i=1}^{m} L(Y_i, F_{i-1}(X_i) + \alpha h_i(X_i, r_{i-1}))$ where $L(Y, F(X))$ is a differentiable loss function.

C. Support Vector Regression (SVR)

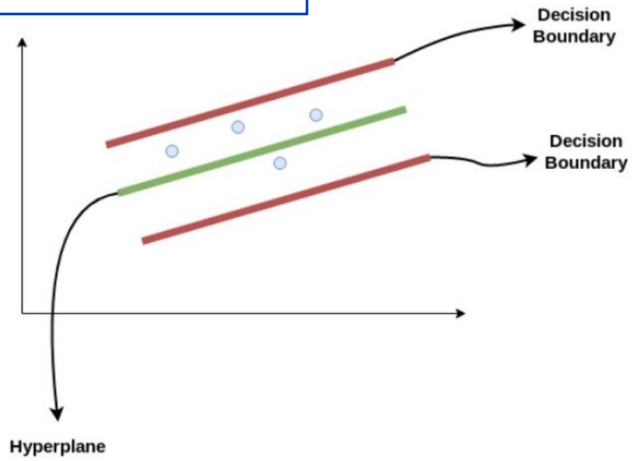
- SVR is an improved model of Support Vector Machine (SVM)
- Find an appropriate line (or hyperplane in higher dimensions) to fit the data
- Can handle high-dimensional features

C. Support Vector Regression (SVR)

- The function of SVR method is: $y = f(x) = \omega^{\mathrm{T}}\varphi(x) + b$
- ω and b are estimated by minimizing the following optimization problem:

$$\min \frac{1}{2} \|\omega\|^2$$

  - subjects to $\begin{cases} y_i - \omega^{\mathrm{T}}\varphi(x_i) - b \leq \varepsilon, \\ b + \omega^{\mathrm{T}}\varphi(x_i) - y_i \leq \varepsilon + \xi_i^* \end{cases}$



  - Here, ξ_i and ξ_i^* are slack variables introduced to cope with training data possibly violating the condition $|f(x_i) - y_i| \leq \varepsilon$

C. Support Vector Regression (SVR)

$$\min \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

- subjects to $\begin{cases} y_i - \omega^T \varphi(x_i) - b \leq \varepsilon, \\ b + \omega^T \varphi(x_i) - y_i \leq \varepsilon + \xi_i^*, \\ \xi_i, \xi_i^* \geq 0, i = 1, \ldots, n, \end{cases}$

- where C is a constant known as the penalty factor, ε is the insensitive loss parameter and the slack variables ξ_i and ξ_i^* measure the amount of difference between the estimated value and the target value beyond ε

D. Auto Regressive Integrated Moving Average (ARIMA)

- ARIMA is a class of models that explains a given time series based on its own past values, that is, its own lags and the lagged forecast errors, so that the equation can be used to forecast future values.
- An ARIMA model is characterized by 3 terms: p, d, q, where:
  - p : order of the AR term,
  - q : order of the MA term,
  - d: number of differencing required to make the time series stationary.

D. ARIMA

- AutoRegressive Integrated Moving Average (ARIMA) is a time series forecasting model that incorporates autocorrelation measures to model temporal structures within the time series data to predict future values.
- A pure Auto Regressive (AR only) model is one that depends only on its own lags. That is, Y_t is a function of the 'lags of Y_t'.

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_1$$

- where, Y_{t-1} is the lag-1 of the series, β_1 is the coefficient of lag-1 that the model estimates and α is the intercept term, also estimated by the model.

D. ARIMA

- A pure Moving Average (MA only) model is one where Y_t depends only on the lagged forecast errors

$$Y_t = \alpha + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \varphi_2\varepsilon_{t-2} + \cdots + \varphi_q\varepsilon_{t-q}$$

where the error terms are the errors of the autoregressive models of the respective lags. The errors ε_t and ε_{t-1} are the errors from the following equations:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_t Y_0 + \varepsilon_t$$
$$Y_{t-1} = \beta_1 Y_{t-2} + \beta_2 Y_{t-3} + \cdots + \beta_{t-1} Y_0 + \varepsilon_{t-1}$$

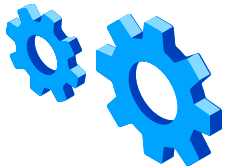
- An ARIMA model is one where the time series was differenced at least once to make it stationary, and you combine the AR and the MA terms. So, the equation becomes:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_p Y_{t-p} + \varepsilon_t + \varphi_1\varepsilon_{t-1} + \varphi_2\varepsilon_{t-2} + \cdots + \varphi_q\varepsilon_{t-q}$$

03

Model Tuning

A. Hyperparameter Tuning



- Tuning is to find the optimal parameter
- Improve prediction power and running speed of models
- Tune the parameters of the RFR, SVR, XGBoost and models

Choosing hyperparameters

Random forest regressor:

- **GridSearchCV** : Used to find the best hyperparameters for the RFR model, with the hyperparameters of **n_estimators, min_samples_leaf, max_depth** being tuned.
    - The best combination of hyperparameters selected through 5-fold cross-validation was: n_estimators=100, min_samples_leaf=2, max_depth=None.

➢ The model was then fit to the training data using the best hyperparameters found.

Choosing hyperparameters

Support Vector Regression:

- C: determines the penalty for data points that fall outside the margin or violate the regression tolerance
- Kernel: determines the type of kernel function used in SVR to transform the feature space
  - linear: used when the data can be well separated by a hyperplane.
  - rbf (Radial Basis Function): used when the data is not linearly separable in the original space.
  - sigmoid : used when the data is not linearly separable and has similar characteristics to the sigmoid function in logistic regression.
  - γ: adjusts the influence of a training data point on other data points.

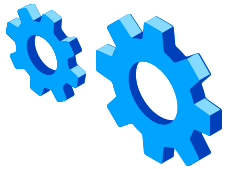
Choosing hyperparameters

XGBoost:

- n_estimators: represents the number of individual decision trees (weak learners) to be built in the XGBoost model.
- max_depth: defines the maximum depth of each decision tree in the XGB mode
- learning_rate: controls the contribution of each tree in the ensemble

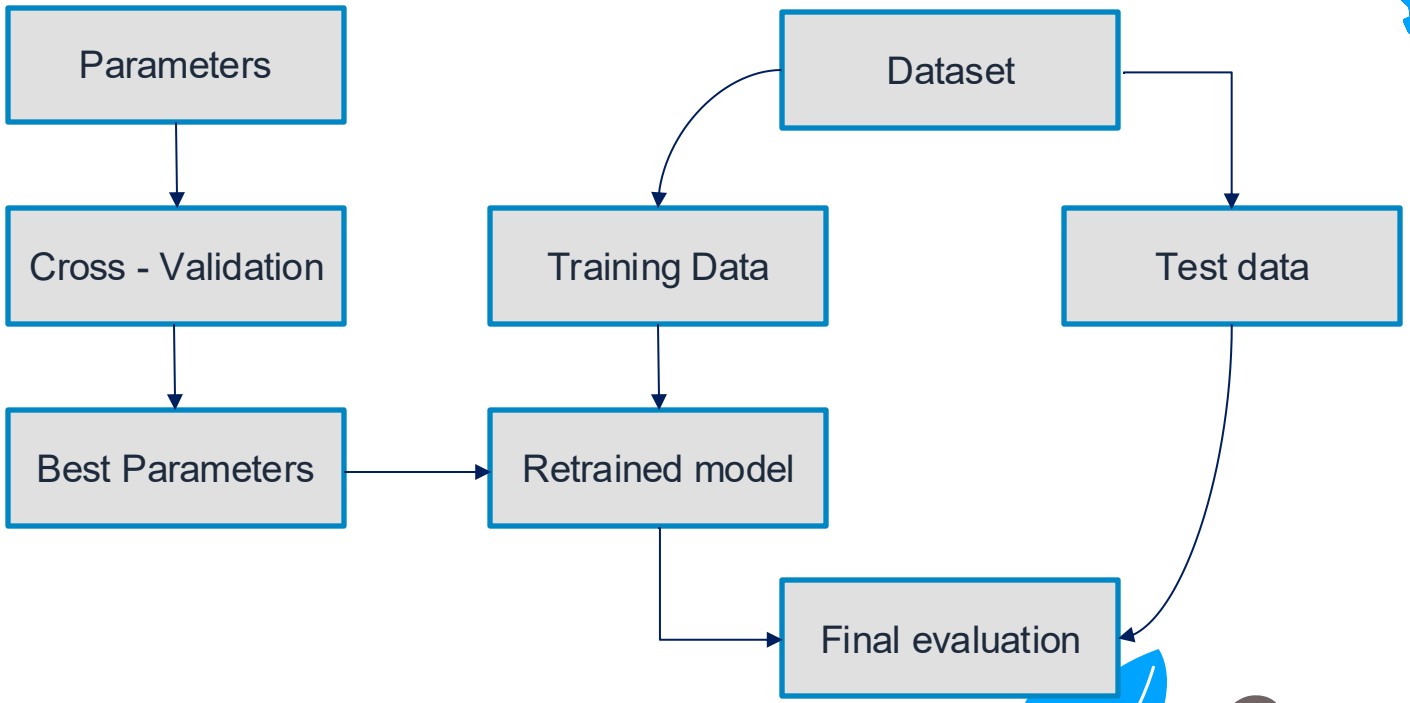
B. Cross Validation For Time Series

- Can’t use Cross-Validation in this project
- Alter by Time Series Cross Validation



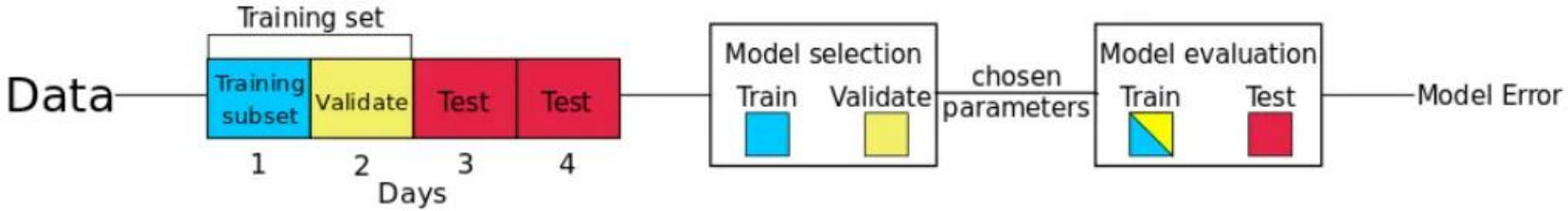
- The rules of splits :
  - Every test set contains unique observations
  - Observations from the training set occur before their corresponding test set
- In this project, we introduce two kind of Nested Cross Validation:
  - Predict Second Half
  - Day Forward Training

B. Cross Validation For Time Series



B. Cross Validation - Predict Second Half

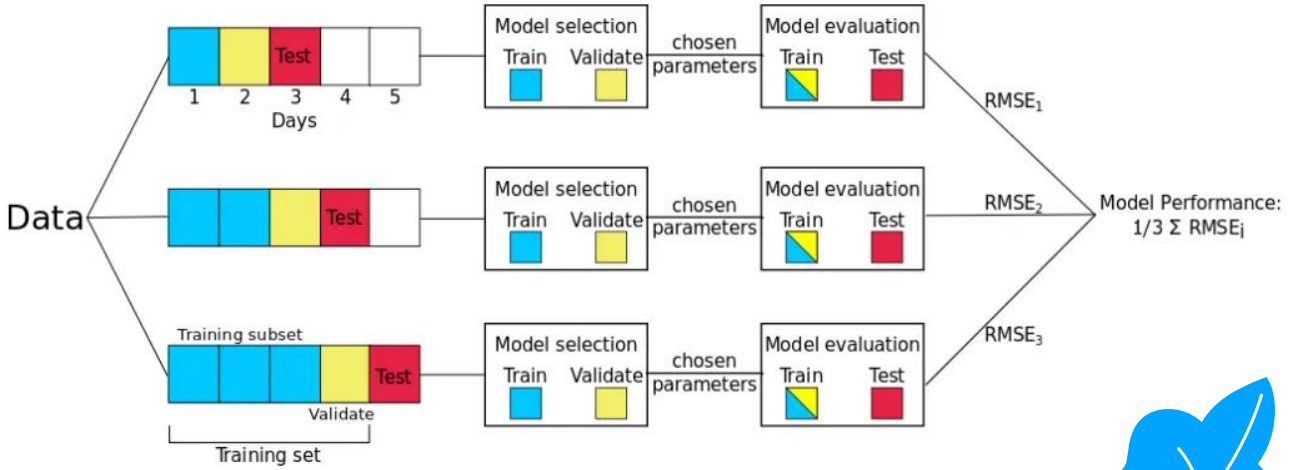
- 50% : Training subset/ Validate
- 50% : Test



- **Simple but pay attention to chronological order of the data !!!**

B. Cross Validation - Day Forward-Chaining

- Each day is the test set and all previous data is the training set.
- Each split represents a distinct evaluation scenario.



B. Cross Validation - Comparison and Conclusion

The Predict Second Half

- May not capture the full variability of the data or provide a robust estimate of model error
- May not effectively evaluate the model's ability to generalize to unseen future data

The Day forward-chaining

- More robust estimate of the model's error
- Better assessment of its generalization capabilities.

04

Summary & proposals

Summary

The number in each cells are

- Train mean squared error,
- Train R2 score,
- Test mean squared error,
- Test R2 score respectively

Algorithm	LCS	PTC	LSS	PLX	TCB	VIG	ITD	VGI
ARIMA	0.0158	0.0111	0.0170	0.0224	0.0211	0.0179	0.0110	0.0263
	0.9864	0.9937	0.9915	0.9790	0.9932	0.9912	0.9976	0.9782
	0.0080	0.0327	0.0340	0.0270	0.0322	0.0380	0.0474	0.1577
	0.9687	0.9676	0.9613	0.9212	0.9747	0.9370	0.9065	0.9584
SVR	0.0052	0.0039	0.0054	0.0152	0.0044	0.0033	0.0028	0.0215
	0.9948	0.9961	0.9946	0.9848	0.9956	0.9967	0.9972	0.9785
	0.0032	0.0064	0.0169	0.0209	0.0489	0.0356	0.0275	0.0384
	0.9968	0.9936	0.9831	0.9791	0.9511	0.9644	0.9725	0.9616
SVR & DFC	0.0061	0.0037	0.0061	0.0173	0.0032	0.0030	0.0021	0.0019
	0.9938	0.9960	0.9941	0.9771	0.9938	0.9968	0.9974	0.9766
	0.0012	0.0050	0.0060	0.0114	0.0064	0.0044	0.0053	0.0078
	0.9794	0.9779	0.9882	0.9633	0.9668	0.9774	0.9804	0.9542
SVR & PSH	0.0064	0.0036	0.0057	0.0259	0.0036	0.0028	0.0018	0.0207
	0.9936	0.9964	0.9943	0.9741	0.9964	0.9972	0.9982	0.9793
	0.0105	0.0054	0.0070	0.0133	0.0069	0.0051	0.0113	0.0061
	0.9895	0.9946	0.9930	0.9867	0.9931	0.9949	0.9887	0.9939
XGB	0.0002	0.0002	0.0004	0.0003	0.0002	0.0002	0.0002	0.0004
	0.9997	0.9997	0.9995	0.9996	0.9998	0.9998	0.9998	0.9996
	0.1385	0.0447	0.0208	0.0371	0.1868	0.0443	0.0677	1.0000
	0.8614	0.9552	0.9791	0.9628	0.8132	0.9557	0.9323	0.0000
XGB & DFC	0.0026	0.0007	0.0053	0.0117	0.0009	0.0021	0.0014	0.0010
	0.9973	0.9990	0.9948	0.9845	0.9988	0.9977	0.9982	0.9870
	0.0210	0.0294	0.0065	0.0497	0.1812	0.0114	0.0338	0.7232
	0.7147	0.9660	0.9870	0.8941	0.3127	0.9310	0.9065	-0.8025
XGB & PSH	0.00003	0.00009	0.00003	0.00003	0.00003	0.00002	0.00008	0.00002
	0.99997	0.99999	0.99997	1.00000	1.00000	0.99998	1.00000	0.99998
	0.21508	0.03499	0.00819	0.12182	0.02055	0.01409	0.09098	0.21134
	0.78492	0.96501	0.99181	0.87818	0.97945	0.98591	0.90902	0.78866
RFR	0.0004	0.0005	0.0005	0.0009	0.0003	0.0003	0.0002	0.0003
	0.9996	0.9995	0.9995	0.9991	0.9997	0.9997	0.9998	0.9997
	0.0014	0.0024	0.0018	0.0041	0.0010	0.0014	0.0008	0.0016
	0.9986	0.9976	0.9982	0.9959	0.9990	0.9986	0.9992	0.9984

Proposal of Improvements

- **Enhancing Data Integration**: To improve the overall performance of the system, we recommend integrating additional relevant data sources. By incorporating financial news, social media sentiment analysis, and economic indicators, we can capture a broader range of factors that influence stock prices.

Proposal of Improvements

- **User-Friendly Interface and Visualization**: To ensure an intuitive and user-friendly experience, we propose developing a web-based interface that allows users to interact with the system effortlessly. The interface should provide clear visualizations of predicted stock prices, historical data, and relevant performance metrics. Additionally, incorporating features like customizable alerts and notifications will empower users to make timely and informed investment decisions

References

1. Chương, P. H. 2020. Tác động của đại dịch Covid-19 đến nền kinh tế Việt Nam. Tạp chí Kinh tế và Phát triển, 274:12.
2. Nguyễn, T. N. Q. and Võ, T. H. L. 2019. Tác động của một số yếu tố kinh tế vĩ mô đến chỉ số giá chứng khoán tại Việt Nam. TẠP CHÍ KHOA HỌC ĐẠI HỌC MỞ THÀNH PHỐ HỒ CHÍ MINH - KINH TẾ VÀ QUẢN TRỊ KINH DOANH, 14(3):47–63.
3. Nguyễn, T. T. V. 2010. Cuộc khủng hoảng kinh tế thế giới năm 2008–2009 và tác động đối với kinh tế - xã hội Việt Nam. Luận văn ThS. Quan hệ quốc tế: 6031 40, Trường Đại học Khoa học Xã hội và Nhân văn.
4. M. Ghosh and R. Gor. 2022. Stock price prediction us-ing support vector regression and k-nearest neighbors: A comparison. International Journal of Engineering Science Technologies, 6(4):1–9.

References

5. S. Gumparthi. 2017. Relative strength index for de-veloping effective trading strategies in constructing optimal portfolio. International Journal of Applied Engineering Research, 12(19):8926–8936.

6. Sarika Keswani, Veerma Puri, and Rimjhim Jha. 2024. Relationship among macro economic factors andstock prices: cointegration approach from the in-dian stock market. Cogent Economics & Finance,12(1):2355017.

7. Dr Polamuri, Kudipudi Srinivas, and A. Mohan. 2019.Stock market prices prediction using random forestand extra tree regression. International Journal ofRecent Technology and Engineering, 8:1224 – 1228.

8. S. Prabhakaran. 2023. Arima model – completeguide to time series forecasting in python.https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/.Machine Learning Plus.

Thank you for your attention!

Do you have any questions?