# VolTS Augmented: An Improvement of a Volatility-based Trading System to Forecast Stock Markets Trends

Ivan Letteri[1,*]

[1] University of L'Aquila, Coppito (AQ), 67100, Italy

**Abstract**

In this study, we proposed an improvement to a previous strategy of the work called VolTS. It is a module of the AITA framework that integrates statistical analysis with machine learning techniques for forecasting stock market trends. The main goal is to verify the cointegration in volatility-based trading strategies within financial markets, and then to devise a methodology that leverages market dynamics to yield profits. This methodology improvement incorporates the Dynamic Time Warping to assess the similarity of trend sequences, even when they exhibit temporal misalignment coupling it. With the K-Nearest Neighbors algorithm, which identifies the most akin price patterns, we construct a sophisticated model spotlighting stocks among the nine largest ones listed on both the NYSE and Nasdaq exchanges that exhibit analogous price movements to our portfolio stocks. From volatility assets pointofview, it is applied the Granger Causality Test to the dataset containing the same mid-range volatility clusters of the chosen stocks, thus identifying them with robust predictive relationships. These "predictor" stocks were pivotal in shaping our trading strategy, serving as trend indicators to inform decisions on target stock trades. The empirical findings demonstrated the effectiveness of our method in identifying, small but not rare, profitable day-trading opportunities. This success was attributed to the predictive insights from volatility clusters, the Granger causality relationships, and Co-integration trends identified among the stocks. In conclusion, our research has significantly contributed to the realm of volatility-based trading strategies by introducing a methodology that mixes statistical techniques with machine learning.

**Keywords**

Dynamic Time Warping, K-Nearest Neighbour, Technical Analysis, Stock Market Prediction, Algorithmic Trading, Backtesting

## 1. Introduction

This article explores the emerging field of volatility-based trading strategies, which is encountering promising growth in the financial sector. Artificial intelligence (AI) has emerged as a pivotal player, offering robust tools for analysing market volatility and leveraging it for profitable outcomes. AI models, trained to estimate mean volatility, offer valuable insights into the inherent uncertainty and risk linked to individual securities or the overall market [1].

Our research was primarily driven by the following key questions: $RQ_1$: Can k-means clustering effectively determine the mean volatility of prominent stocks? $RQ_2$: defined the stocks with the same mid-term volatility, can the Granger Causality Test be employed to identify predictive influences between stocks? $RQ_3$: Established the subset of influenced stocks, can Dynamic Time Warping (DTW) [2] be paired with K-Nearest Neighbours (KNN) to spotlight stocks that "fluctuate" to the same tune exploiting the delay among them?

This study aimed to formulate trading strategies based on predictive connections affecting influential stocks as trend indicators. The proposed time series study, related to the prices and volatility-driven trading strategy was then rigorously evaluated through backtesting and performance analysis to validate its reliability. Through empirical evidence, this research has given an improvement to the previous volatility-based trading strategies [3].

The augmented AI trading strategy utilised k-means clustering of average volatility data [4]. This data encompassed nine major stock markets. Our initial objective was to identify distinct volatility patterns within the market and subsequently group assets accordingly. Following this, the Granger Causality Test (GCT) [5] was leveraged to pinpoint stocks that significantly predicted others within our analysis. Then, we pairwise compare two time series to find the closest match between them via DTW. Then, the future trend is predicted as the average of the trends of the $K$ neighbours. These predictive relationships were then utilised to establish buy, sell,

*Ital-IA 2024: 4th National Conference on Artificial Intelligence, organized by CINI, May 29–30, 2024, Naples, Italy*

*Corresponding author.

EMAIL: ivan.letteri@univaq.it (I. Letteri)

ORCID: 0000-0002-3843-386X (I. Letteri)



CEUR Workshop Proceedings (CEUR-WS.org)

or hold trading decisions.

Our previous research focused on technical trading strategies that emphasised technical indicators [1],[6]. This current exploration delves, first into Historical Volatility estimators as a dataset for identifying medium volatility selecting stocks for the Granger Causality Test asset cointegration approach [7]. Second, within the context of investment timing, we predict it via DTW paired with KNN.

This paper is organised as follows: Section 2 summarize fundamental concepts within our AITA framework [8], highlighting the Volatility Trading System (VolTS) module [3]. Section 3 outlines the improvement with the VolTS Augmented (in short *VolTS-Aug*) module, which analyses securities' volatility averages and price trends establishing predictive relationships. It then delves into the implementation of the trading strategy and includes a thorough empirical analysis of its performance and robustness. Section 4 presents practical findings achieved through backtesting followed by a discussion. Finally, Section 5 concludes the study by summarising the effectiveness and applicability of the proposed method.

# 2. BACKGROUND

## 2.1. Price Action

The price action (PA) influences Historical Volatility (HV), and in turn, HV can provide insights into future PA. When the PA exhibits strong price movements, such as wide trading ranges, breakouts, or rapid directional changes, it tends to increase.

VolTS-Aug, as an improvement of the VolTS module within the AITA framework, adheres to these principles. Low HV signifies a period of consolidation or low price volatility, indicating a potential upcoming spike in volatility or a shift in the PA. On the other side, high HV suggests a higher probability of sharp market movements or trend changes. Also into VolTS-Aug, the PA is encoded as OHLC, i.e., the open, high, low, and close prices of the assets.

For each timeframe $t$, the OHLC of an asset is represented as a 4-dimensional vector $X_t = (x_t^{(o)}, x_t^{(h)}, x_t^{(l)}, x_t^{(c)})^T$, where $x_t^{(l)} > 0$, $x_t^{(l)} < x_t^{(h)}$ and $x_t^{(o)}, x_t^{(c)} \in [x_t^{(l)}, x_t^{(h)}]$.

## 2.2. Historical Volatility Time Series

The construction of the dataset is composed by time series from the following HV estimators:

The **Parkinson** (PK) estimator incorporates the stock's daily *high* and *low* prices as follow:

$$PK = \sqrt{\frac{1}{4Nln(2)} \sum_{i=1}^{N} \left( ln \frac{x_t^{(h)}}{x_t^{(l)}} \right)^2}.$$

It is derived from the assumption that the true volatility of the asset is proportional to the logarithm ($ln$) of the ratio of the high $x_t^{(h)}$ and low $x_t^{(l)}$ prices of $N$ observations.

The *Garman-Klass* (GK) estimator assumes that price movements are log-normally distributed calculated as follows: $\sqrt{\frac{1}{N} \sum_{i=1}^{N} \frac{1}{2} \left( ln \frac{x_t^{(h)}}{x_t^{(l)}} \right)^2 - \sum_{i=1}^{N} (2ln(2) - 1) \left( ln \frac{x_t^{(c)}}{x_t^{(o)}} \right)^2}$

The *Rogers-Satchell* (RS) estimator uses the range of prices within a given time interval as a proxy for the volatility of the asset as follows: $RS = \sqrt{\frac{1}{N} \sum_{t=1}^{N} ln \left( \frac{x_t^{(h)}}{x_t^{(c)}} \right) ln \left( \frac{x_t^{(h)}}{x_t^{(o)}} \right) + ln \left( \frac{x_t^{(l)}}{x_t^{(c)}} \right) ln \left( \frac{x_t^{(l)}}{x_t^{(o)}} \right)}$

RS assumes that the range of prices within the interval is a good proxy for the volatility of the asset, additionally, the estimator may be sensitive to outliers and extreme price movements.

The *Yang-Zhang* (YZ) estimator [9] incorporates OHLC prices as follows: $YZ = \sqrt{\sigma^2_{OvernightVol} + k\sigma^2_{OpenToCloseVol} + (1-k)\sigma^2_{RS}}$, where $k = 0.34/1.34 + \frac{N+1}{N-1}$, $\sigma^2_{OpenToCloseVol} = \frac{1}{N-1} \sum_{i=1}^{N} \left( ln \frac{x_t^{(c)}}{x_t^{(o)}} - ln \frac{x_t^{(c)}}{x_t^{(o)}} \right)^2$, and $\sigma^2_{OvernightVol} = \frac{1}{N-1} \sum_{i=1}^{N} \left( ln \frac{x_t^{(o)}}{x_{t-1}^{(c)}} - ln \frac{x_t^{(o)}}{x_{t-1}^{(c)}} \right)^2$. YZ exhibits notable performance across a broad spectrum of scenarios, including those characterised by jumps and non-normality in the data. However, this estimator has limitations, and its effectiveness may be constrained in certain contexts.

In this research, our attention is centred on mid-volatility. This focus allows us to either close open positions or refrain from entering a position when the anticipated volatility coefficient is high, thereby mitigating the risk of losses. On the other hand, if the expected volatility is too low, it does not offer any potential for gains.

## 2.3. Trading Strategy

In this experiment, we used the *Trend Following* (TF) strategy. It is one way to engage in trend trading, where a trader initiates an order in the direction of the breakout after the price surpasses the resistance line as follows: let $P_t$ the price at time $t$, and let $MA$ denote the Moving Average of the asset price over a certain period. If $P_t \geq MA_t$ indicates an upward trend to take a long position otherwise it is a downward trend to take a short position.

Then, our strategy is compared only with the *Buy and Hold* (B&H) strategy considering it as a benchmark.

## 2.4. Backtesting Metrics

AITA framework considers the following profit and risk metrics to evaluate the potential profitability of investments and manage risk exposure.

**Drawdown (DD).** It is a measure of the peak-to-trough decline in the value of a trading account before a new peak is attained. DD is defined as follows: $DD = \frac{P-T}{T}$, where $P$ is the highest value or peak of the portfolio. $T$ is the lowest value or trough after the peak. Maximum Drawdown (MDD) is the most significant loss from peak to trough during a specific period calculated as follows: $MDD = max_i \left( \frac{P_i - min_{j:i \leq j \leq N} T_j}{P_i} \right)$, where $P_i$ is the highest value or peak of the portfolio time $i$. $T_j$ is the lowest value or trough after the peak up to time $j$. $N$ is the total number of data points.

**Sortino ratio (SoR).** It is a risk-adjusted profit measure, which refers to the return per unit of deviation as follows: $SoR = \frac{R_p - R_f}{\sigma_d}$, where $R_p$ is the expected portfolio return, $R_f$ the risk-free rate of return, and $\sigma_d$ denotes the downside deviation of the portfolio returns.

**Sharpe ratio (SR).** It is a variant of the risk-adjusted profit measure, which applies $\sigma_p$ as a risk measure: $SR = \frac{R_p - R_f}{\sigma_p}$ where $\sigma_p$ is the standard deviation of the portfolio return.

**Calmar ratio (CR)** is another variant of the risk-adjusted profit measure, which applies MDD as risk measure: $CR = \frac{R_p - R_f}{MDD}$.

To check the goodness of trades, we mainly focused on the **Total Returns** $TR_k(t)$ for each stock $(k = 1, ..., p)$ in the time interval $(t = 1, ..., n)$ with the price $P$, defined as follows: $TR_k(t) = \frac{P_k(t+\Delta t) - P_k(t)}{P_k(t)}$.

Furthermore, we analyzed the standardized returns $r_k = (TR_k - \mu_k)/\sigma_k$, with $(k = 1, ..., p)$, where $\sigma_k$ is the standard deviation of $TR_k$, e $\mu_k$ denote the average overtime for the studied period.

# 3. THE EXPERIMENT

## 3.1. Asset Collections

AITA automatically downloads the OHLC prices via an internal Python library connected to an API, using the MetaTrader5 (MT5)[1]. The data collected for this study includes the OHLC prices of the stocks listed in Table 1.

**Table 1**
List of the main 9 stocks selected for the experimentation.

| Ticker | Company | Market |
|---|---|---|
| MSFT | Microsoft Corporation | Nasdaq |
| GOOGL | Alphabet Inc. | Nasdaq |
| MU | Micron Technology, Inc. | Nasdaq |
| NVDA | NVIDIA Corporation | NYSE |
| AMZN | Amazon.com, Inc. | NYSE |
| META | Meta Platforms, Inc. | NYSE |
| QCOM | QUALCOMM Incorporated | Nasdaq |
| IBM | Int. Business Machines Corp. | NYSE |
| INTC | Intel Corporation | NYSE |

## 3.2. Historical Volatility Time Series

The History Volatility Clustering process of our approach determines the stocks with intermediate volatility. First calculate the average of historical volatility time series among the aforementioned estimators (see sect. 2.2). Next, the resulting volatility series are clusterized using the KMeans++ algorithm. In particular, we split into three clusters $(K = 3)$ *high*, *middle*, and *low* volatility.

Figure 1 shows the results displayed through a plot of the time series belonging to the middle cluster where we are focused on our strategy. It is worth noting that, the main region is in the time window from 1st January 2021 to 1st March 2024.

So, we use this interval as the dataset, and then from the intermediate cluster (confined between the two red dashed lines in fig 1). The candidate assets selected are NVDA, META, AMZN, MU and QCOM.

## 3.3. Causality Analysis

Co-integration refers to the long-term stable linear combination between two or more time series, although individual series may be non-stationary.

In the context of volatility-based trading, the *VolTS-Aug* module performs the GCT to examine this relationship between the lagged volatility of

[1] https://www.metatrader5.com/

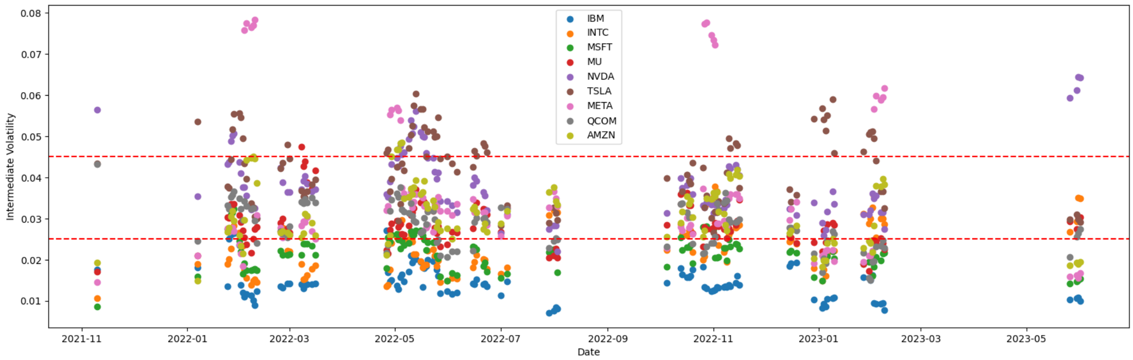


**Figure 1:** Stocks selected in the range, from 1st January 2021 to 1st March 2024, are *MU*, *NVDA*, *AMZN*, *QCOM*, *META* from the Historical Volatility estimators dataset.

one asset and the future volatility of another asset by applying the following steps:

**Step 1.** *Significant Granger causality:* Let $X$ and $Y$ be the pair stocks time series volatility to check, where $X$ represents the potential causal variable and $Y$ represents the potential effect variable. The null hypothesis (H0) states that $X$ does not Granger cause $Y$, while the alternative hypothesis (H1) states that $X$ does Granger cause $Y$. The F-test is defined as follows:

$$F - test = \frac{[(RSS_{Y(t)} - RSS_{YX_{(t)}})/p]}{[(RSS_{YX_{(t)}})/(n-p-k)]},$$

where $RSS$ is the *Residual Sum of Squares* for the two AutoRegressive models: $Y(t) = c_Y + \beta_{Y_1} * Y(t-1) + \beta_{Y_2} * Y(t-2) + \cdots + \beta_{Y_p} * Y(t-p) + \epsilon_{Y(t)}$, and $X: Y(t) = c_{YX} + \beta_{YX_1} * X(t-1) + \beta_{YX_2} * X(t-2) + \cdots + \beta_{YX_p} * X(t-p) + \epsilon_{YX(t)}$, with $p$ the lag order, $n$ the number of observations, and $k$ the number of parameters in the models. It indicates how much the regression coefficients of the lagged time series help to explain the variation in the target time series.

**Step 2.** *Causality Direction:* If the volatility of Stock $X$ Granger causes the volatility of Stock $Y$, it suggests that changes in Stock $X$ volatility can be used to predict changes in Stock $Y$ volatility. A low p-value suggests the presence of a causal relationship between the time series.

**Step 3.** *Delta Time Trends: VolTS-Aug* performs the DTW paired with KNN to examine the interval time necessary for profitable trades: (i) The *DTW* distance between two time series is the sum of differences between their corresponding points, optimally aligned. (ii) The *KNN* classifier, aimed at finding the most similar neighbours for each observation based on DTW distance.

## 3.4. The Algorithm

Three are the main steps followed by *VolTS-Aug*:

**Regression step**: For each pair of time series $(X_i, Y_j)$, where $i \neq j$, we construct a linear regression model: $X_i = \beta_{0,ij} + \beta_{1,ij} Y_j + \epsilon_{ij}$, where $\beta_{0,ij}$ is the intercept, $\beta_{1,ij}$ is the regression coefficient, and $\epsilon_{ij}$ is the error term. We calculate the F-statistic to evaluate the overall adequacy of the model.

**GCT step**: For each pair of time series $(X_i, Y_j)$, we perform the Granger causality test. The model for the Granger test can be expressed as $X_i(t) = \alpha_{ij} + \sum_{k=1}^{n} \beta_{k,ij} X_i(t-k) + \sum_{k=1}^{n} \gamma_{k,ij} Y_j(t-k) + \epsilon_{ij}(t)$, where $X_i(t)$ is the current value of $X_i$, $X_i(t-k)$ and $Y_j(t-k)$ are the lagged values of $X_i$ and $Y_j$, respectively, and $\epsilon_{ij}(t)$ is the error term. If the coefficients $\gamma_{k,ij}$ are statistically different from zero, we reject the null hypothesis and conclude that $Y_j$ Granger causes $X_i$.

**DTW-KNN step**: (i) For each pair of time series $(X_i, Y_j)$, where $X_i$ has length $n$ and $Y_j$ has length $m$, the DTW distance $d(X_i, Y_i)$ is given by $d(X_i, Y_i) = min_{alignment} \sum_{i=1}^{n} \sum_{j=1}^{m} c(i,j)$, where $c(i,j)$ is the distance between points $X_i[i]$ and $Y_j[j]$, and the optimization is performed overall possible alignments. (ii) The process of finding the best parameters delta time $\delta_t$ involves the KNN whit $k$ optimization through grid search.

## 3.5. Metrics observed

Profit and risk metrics are pivotal considerations in trading AITA framework evaluates the following, for the potential profitability of the investments and to manage the risk exposure.

(i) The *Maximum drawdown (MDD)* measures the largest decline from the peak in the whole trading period, to show the worst case, as follows: $MDD =$

$max_{\tau \in (0,t)}[max_{t \in (0,\tau)} \frac{n_t - n_\tau}{n_t}]$.

(ii) The *Sharpe ratio (SR)* is a risk-adjusted profit measure, which refers to the return per unit of deviation as follows: $SR = \frac{\mathbb{E}[r]}{[r]}$.

(iii) The *Sortino ratio (SoR)* is a variant of the risk-adjusted profit measure, which applies DD as risk measure: $SoR = \frac{\mathbb{E}[r]}{DD}$.

(iv) The *Calmar ratio (CR)* is another variant of the risk-adjusted profit measure, which applies MDD as risk measure: $CR = \frac{\mathbb{E}[r]}{MDD}$.

To check the goodness of trades, we mainly focused on the (v) *Total Returns* $R_k(t)$ for each stock $(k = 1, ..., p)$ in the time interval $(t = 1, ..., n)$, where $TR = R_k(t) = \frac{Z_k(t+\Delta t) - Z_k(t)}{Z_k(t)}$, and furthermore analysing the standardized returns $r_k = (R_k - \mu_k)/\sigma_k$, with $(k = 1, ..., p)$, where $\sigma_k$ is the standard deviation of $R_k$, e $\mu_k$ denote the average overtime for the studied period.

# 4. RESULTS AND DISCUSSIONS

## 4.1. The Experiment

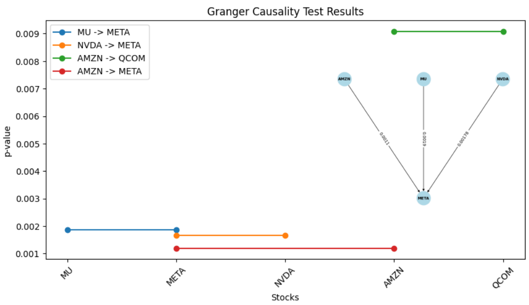


**Figure 2:** Best Granger Causality Test with 21 lags.

The VolTS algorithm iterates the daily lags in a range from 2 to 30 days to determine the best testing range time. The best result is achieved with lags=21, where '*best*' is considered when there is direction coherency among the stocks with the lowest p-value, with the maximum cardinality of the set of stocks previously filtered, and not create cyclic graph in the connections. On other words, the GCT suggests buying AMZN, MU and NVDA when META has a positive trend and vice versa (see fig. 2).

The experiment results indicate that the strategy resulted in a total gain of 241.57$ in 21 days of market opening, starting with an initial budget of 1000$ per stock.

Tab. 2 contains further details about the performance metrics of the strategy and shows how the total amount in the portfolio is increased to 3241.57$ (8.05%), which is a positive sign of profitable trading, also considering the fixed commission of 9$ per trade. Notice that, the managing of the budget is set in compounded mode, so the full amount is reused for each trade.

## 4.2. Backtesting

The analysis of individual stocks' performance is presented in figure 3 about META co-integration. The trades of AMZN bought following the META trend given a profit of 59.98$, with a winrate of 75%, a MDD of 0.035%, and a return of 5.98%, which outperforms the B&H strategy with return of 0.998%. The trades of MU bought following the META trend given a profit of 146.68$, with a winrate of 100%, a MDD of 0%, and a return of 34.91%, which outperforms the B&H strategy with return of 3%. The trades of NVDA bought following the META trend given a profit of 39.11$, with a winrate of 75%, a MDD of 0.11%, and a return of 3.98%, which is similar to the B&H strategy with return of 4.02%.

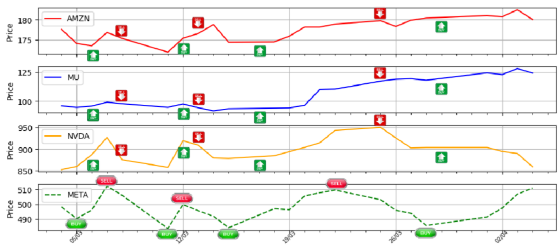


**Figure 3:** Trades during the 21 testing days (1st March - 5th April).

Figure 3 shows all the 7 positions applied at the same time on NVDA, MU and AMZN with buying/selling trades following the co-integration with META trend determined with *VolTS-Aug* algorithm. The trades on MU stock outperforms all the others stocks increasing notably the portfolio gain (see Table 2).

# 5. Conclusions

In this paper, we propose *VolTS-Aug* an improvement of the AITA framework module calle VolTS. *VolTS-Aug* handle volatility in trading strategy combining causality by the Historical Volatility Granger Causality Test and DTW & KNN to determine

**Table 2**
Results of the backtesting in the experiment.

| Stock | #Trades | $\mu$ WinRate (%) | $\mu$ TR ($) | $\mu$ SR | $\mu$ SoR | CR |
|---|---|---|---|---|---|---|
| *META ->AMZN, NVDA, MU* | 7 | 83.33 | 241.57 | 3.03 | 8.75 | 15.941 |

a profitable stock pairings improving on previous work [10, 1].

The novelty of the approach implemented in *VolTS-Aug* lies in a better trades timing. To validate this claim, we applied our methodology to nine assets reduce to four after filtering by our methodology. The results shows a promising potential of this approach with a gain of 241.57$ (8.05%) in 21 days. In future works, we will further test its reliability with more refined assets selection (e.g., [11, 12]) and balancing (buy, sell and hold trades) strategies (e.g., [13, 14]).

## References


[1] I. Letteri, G. D. Penna, G. D. Gasperis, A. Dyoub, Dnn-forwardtesting: A new trading strategy validation using statistical timeseries analysis and deep neural networks, 2022. arXiv:2210.11532.

[2] V. Niennattrakul, C. A. Ratanamahatana, On clustering multimedia time series data using k-means and dynamic time warping, in: 2007 International Conference on Multimedia and Ubiquitous Engineering (MUE'07), 2007, pp. 733–738. doi:10.1109/MUE.2007.165.

[3] I. Letteri, Volts: A volatility-based trading system to forecast stock markets trend using statistics and machine learning, 2023. arXiv:2307.13422.

[4] D. Arthur, S. Vassilvitskii, K-means++: The advantages of careful seeding, volume 8, 2007, pp. 1027–1035. doi:10.1145/1283383.1283494.

[5] G. Kirchgässner, J. Wolters, Granger Causality, Springer Berlin Heidelberg, Berlin, Heidelberg, 2007, pp. 93–123. doi:10.1007/978-3-540-73291-4_3.

[6] I. Letteri, G. D. Penna, G. D. Gasperis, A. Dyoub, Trading strategy validation using forwardtesting with deep neural networks, in: M. Arami, P. Baudier, V. Chang (Eds.), Proceedings of the 5th International Conference on Finance, Economics, Management and IT Business, FEMIB 2023, Prague, Czech Republic, April 23-24, 2023, SCITEPRESS, 2023, pp. 15–25. doi:10.5220/0011715300003494.

[7] R. F. Engle, C. W. J. Granger, Co-integration and error correction: Representation, estimation, and testing, Econometrica 55 (1987) 251–276. URL: http://www.jstor.org/stable/1913236.

[8] I. Letteri, AITA: A new framework for trading forward testing with an artificial intelligence engine, in: F. Falchi, F. Giannotti, A. Monreale, C. Boldrini, S. Rinzivillo, S. Colantonio (Eds.), Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-located with the 3rd CINI National Lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023, volume 3486 of *CEUR Workshop Proceedings*, CEUR-WS.org, 2023, pp. 506–511.

[9] D. Yang, Q. Zhang, Drift-independent volatility estimation based on high, low, open, and close prices, The Journal of Business 73 (2000) 477–492. URL: http://www.jstor.org/stable/10.1086/209650.

[10] I. Letteri, G. D. Penna, G. D. Gasperis, A. Dyoub, A stock trading system for a medium volatile asset using multi layer perceptron, 2022. arXiv:2201.12286.

[11] I. Letteri, G. D. Penna, L. D. Vita, M. T. Grifa, Mta-kdd'19: A dataset for malware traffic detection 2597 (2020) 153–165. URL: http://ceur-ws.org/Vol-2597/paper-14.pdf.

[12] I. Letteri, G. D. Penna, P. Caianiello, Feature selection strategies for HTTP botnet traffic detection (2019) 202–210. doi:10.1109/EuroSPW.2019.00029.

[13] I. Letteri, A. D. Cecco, G. D. Penna, Dataset optimization strategies for malware traffic detection, CoRR abs/2009.11347 (2020). URL: https://arxiv.org/abs/2009.11347. arXiv:2009.11347.

[14] I. Letteri, A. D. Cecco, A. Dyoub, G. D. Penna, Imbalanced dataset optimization with new resampling techniques, in: K. Arai (Ed.), Intelligent Systems and Applications - Proceedings of the 2021 Intelligent Systems Conference, IntelliSys 2021, Amsterdam, The Netherlands, 2-3 September, 2021, Volume 2, volume 295 of *Lecture Notes in Networks and Systems*, Springer, 2021, pp. 199–215. doi:10.1007/978-3-030-82196-8_15.