International Journal of Advanced Smart Convergence Vol.12 No.4 111-118 (2023)
http://dx.doi.org/10.7236/IJASC.2023.12.4.111

IJASC 23-4-13

P-Triple Barrier Labeling: Unifying Pair Trading Strategies and Triple Barrier Labeling Through Genetic Algorithm Optimization

Ning Fu, Suntae Kim

Dept. of Software Engineering, Jeonbuk National University, Korea
funing230@jbnu.ac.kr, stkim@jbnu.ac.kr

Abstract

In the ever-changing landscape of finance, the fusion of artificial intelligence (AI)and pair trading strategies has captured the interest of investors and institutions alike. In the context of supervised machine learning, crafting precise and accurate labels is crucial, as it remains a top priority to empower AI models to surpass traditional pair trading methods. However, prevailing labeling techniques in the financial sector predominantly concentrate on individual assets, posing a challenge in aligning with pair trading strategies. To address this issue, we propose an inventive approach that melds the Triple Barrier Labeling technique with pair trading, optimizing the resultant labels through genetic algorithms. Rigorous backtesting on cryptocurrency datasets illustrates that our proposed labeling method excels over traditional pair trading methods and corresponding buy-and-hold strategies in both profitability and risk control. This pioneering method offers a novel perspective on trading strategies and risk management within the financial domain, laying a robust groundwork for further enhancing the precision and reliability of pair trading strategies utilizing AI models.

Keywords: *Pair Trading, Triple Barrier Labeling, AI, Genetic Algorithm, Labeling Method.*

1. Introduction

As artificial intelligence (AI) undergoes continuous advancements, researchers are increasingly merging it with the economic sphere, striving for amplified profits. In the field of AI, especially within the specialized domain of supervised learning, the crucial role of labels cannot be underestimated. AI models trained with precise labels can more accurately seize market opportunities, surpassing traditional trading strategies. Among numerous trading strategies, pair trading is highly regarded. Its fundamental principle is rooted in mean reversion theory, ensuring stable profits through the analysis of price differences between paired assets [1]. Researchers like Chen et al. [2] and Mudchanatongsuk et al. [3] have applied these strategies successfully in stock markets, while M. Fil and L. Kristoufek [4] and Van den Broek and Zara Sharif [5] demonstrated their

Manuscript Received: october. 11, 2023 / Revised: october. 18, 2023 / Accepted:october. 22, 2023
Corresponding Author: stkim@jbnu.ac.kr
Tel: 063-270-4788 Fax +82-63-270-4767 M +82-10-8477-0042
Author's affiliation (Professor, Department of Software Engineering, Jeonbuk National University, Korea)

effectiveness in cryptocurrencies.

Meanwhile, employing AI models to enhance pair trading strategies for researchers poses a significant challenge: How to create precise labels to train AI models, enabling them to outperform traditional pair trading methods. The mainstream economic labeling method, specifically the Triple Barrier Labeling method introduced by De Prado and Marcos Lopez [6], has received extensive recognition in studies conducted by researchers such as Kovačević et al. [7] and Khosravi et al. [8], acknowledging it as a flexible and efficient labeling technique. However, its applicability is confined to individual assets, creating a barrier to seamlessly aligning with the characteristics of pair trading strategies.

To address this issue, we introduced the P-Triple Barrier Labeling method. This innovative approach involved reformulating the Triple Barrier Labeling method and seamlessly integrating it with pair trading strategies. We achieved this integration by transforming the initial input data, previously a single price series, into spread data derived from paired assets. Diverse parameter configurations within the P-Triple Barrier Labeling method yield distinct trading signals. These parameters were further optimized using genetic algorithms, leading to the development of two sets of signals: High Risk High Profits (HRHP) and Low Risk Low Profits (LRLP). Our experimental results demonstrate that the profitability of HRHP labels has increased 15-fold compared to traditional pair trading strategies, which is a remarkable achievement. Similarly, LRLP labels also outperformed traditional pair trading strategies in simulated tests, with a maximum drawdown (MDD) lower than traditional pair trading strategies while doubling the profitability. Additionally, both strategies boast higher Sharpe ratios than their traditional pair trading counterparts. This innovative method opens new avenues for trading strategies and risk management in the financial sector, furnishing financial market participants with a more resilient and intelligent foundation for decision-making.

2. Background

2.1 Pair Trading Strategy

The pair trading strategy identifies signals by analyzing price spreads between correlated assets, capitalizing on mean reversion to capture trading opportunities. For detailed methods, refer to Figure 1.

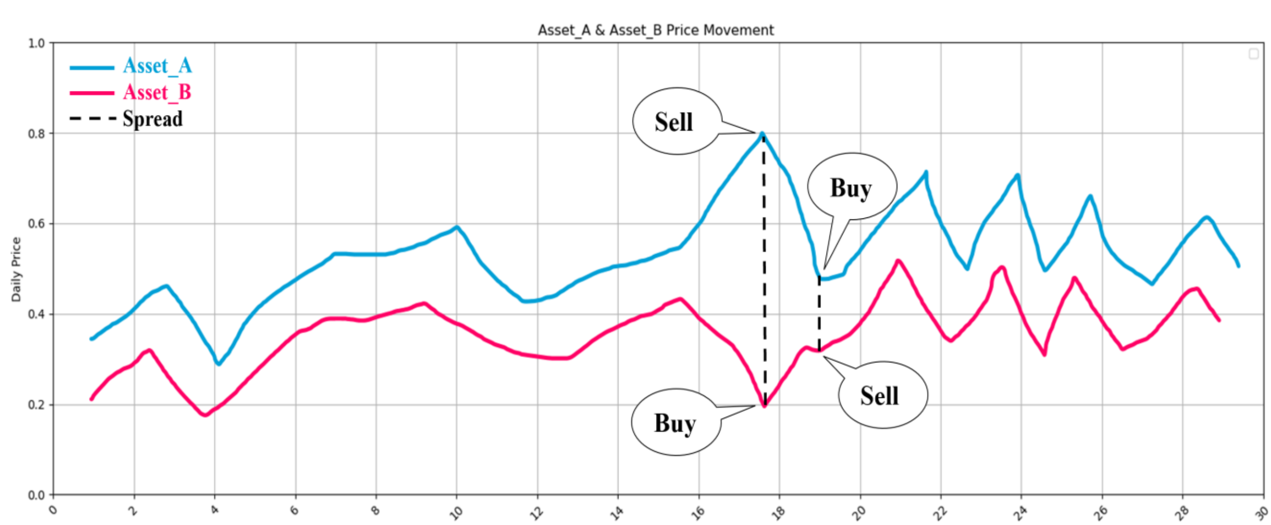

Figure 1. Pair Trading Strategy

Figure 1 illustrates the synchronized price movements of Assets_A and Asset_B, suggesting they are paired assets. Divergence begins on day 15. On day17, a substantial deviation prompts alerts to investors (a preset threshold can indicate the maximum deviation level), guiding them to make informed decisions. In this scenario, investors should sell Assets_A and buy the underperforming Asset_B. Starting from day 19, prices revert to normal, prompting alerts for investors to reverse their trade, maximizing profits [9].

2.2 Triple Barrier Labeling Method (TBM)

Trading signals are derived from predetermined barriers, including the Top Barrier (Profit-Taking Barrier), Button Barrier (Stop-Loss Barrier), and Vertical Barrier (Max Holding Period). These barriers guarantee accurate entry and exit points, effectively mitigating risks [10]. However, this method is unsuitable for pair trading as it does not handle price spread between assets. refer to Figure 2 for details.

From Figure 2, it can be observed that the triple barrier labeling method dynamically generates labels within Max Holding Period.

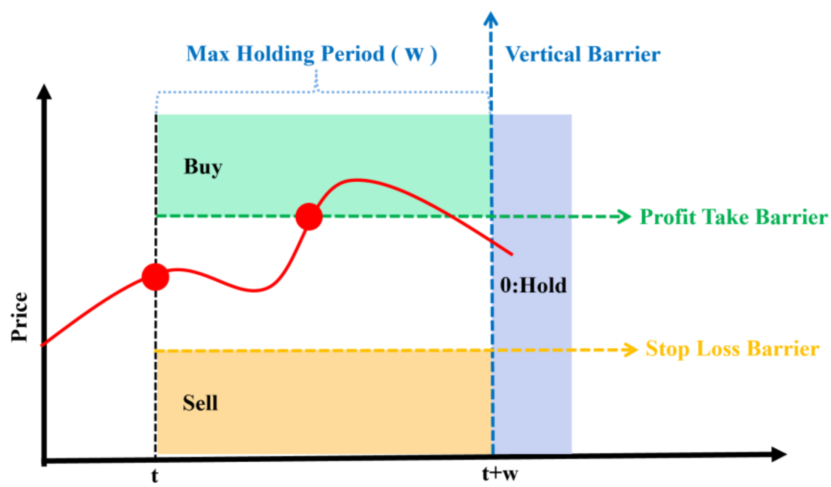

Figure 2. Triple Barrier Labeling method

For example: Within the Max Holding Period (W) of 5 days.

- If Price is hit the Profit Take Barrier first, Day_t is given Buy signal.
- If Price is hit the Stop Loss Barrier first, Day_t is given Sell signal.
- If Price is hit the Vertical Barrier (Max Holding Period), Day_t is given Maintain the state signal.

3. Method

This chapter details our research method. Initially, assets were chosen meticulously based on correlation and cointegration principles for trading pairs. Precise spread calculations were then standardized into Z-spread. Integrating Z-Spread with different parameters into the Triple Barrier Labeling method yielded multiple trading signals. These parameters were further optimized using genetic algorithms, resulting in two distinct sets of signals: HRHP and LRLP. During the simulated trading phase, experiments were conducted using these signals, allowing the computation of profits and MDD for each signal group. Refer to Figure 3 for a detailed overview.

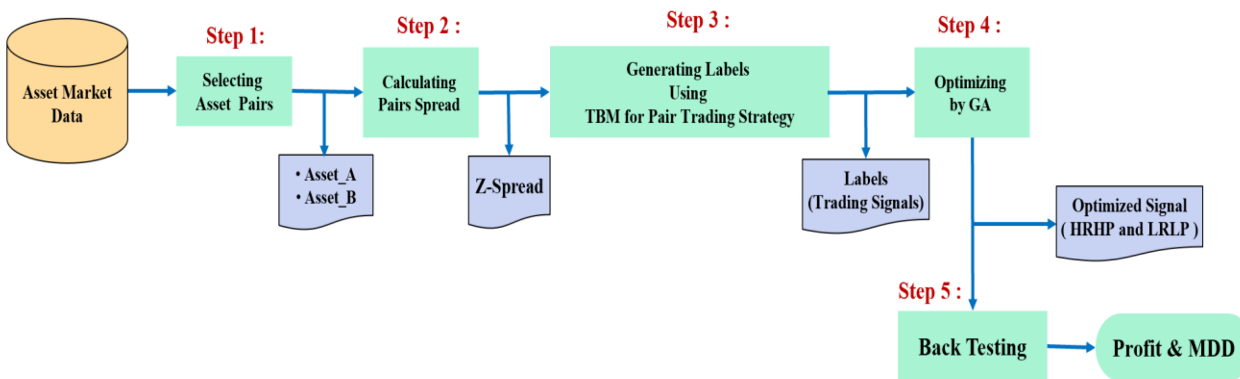

Figure 3. Overview

3.1 Selecting Asset Pairs

Selecting appropriate trading pairs is crucial for pair trading strategies. High correlation and cointegration between Asset_A and Asset_B indicate adaptable pairs.

Pearson Correlation Coefficient [11]: Measures the linear relationship strength between Asset_A and Asset_B, ranging from -1 (perfect negative) to 1 (perfect positive), with 0 indicating no correlation. Calculated as:

$$Correlation(Asset\_A, Asset\_B) = \frac{Corr(Asset\_A, Asset\_B)}{\sqrt{Var(Asset\_A) \times Var(Asset\_B)}} \quad (1)$$

Here, Corr denotes covariance, and Var represents variance.

Cointegration Test [12]**:** Estimates the linear relationship between Asset_A and Asset_B using regression:

$$\Delta Asset\_B_t = \alpha + \beta \times \Delta Asset\_A_t + \epsilon_t \quad (2)$$

The Augmented Dickey-Fuller (ADF) test checks for unit roots in the residual $\Delta\epsilon_t$. If $\Delta\epsilon_t = 0$, it suggests cointegration, indicating a long-term relationship between cryptocurrencies, essential for trading or investment strategies.

3.2 Calculating Pairs Spread

We calculate the spread for chosen cryptocurrency pairs using Osifo, Ernest, and Bhattacharyya's state-of-the-art method [13], as praised in their research. Daily price change rates between BTC and ETH form the basis for our trading signals, Calculated as:

$$\text{Spread} = \left(\frac{Asset\_A_T - Asset\_A_{T-1}}{Asset\_A_{T-1}}\right) - \left(\frac{Asset\_B_T - Asset\_B_{T-1}}{Asset\_B_{T-1}}\right) \quad (3)$$

where $Asset\_A_T$ and $Asset\_A_{T-1}$ represent the cryptocurrency prices of Asset_A on day T and day T-1 respectively, and $Asset\_B_T$ and $Asset\_B_{T-1}$ represent the cryptocurrency prices of Asset_B on day T and day T-1 respectively. As the spread alone cannot express the deviation from the historical mean, we introduce the commonly used Z-spread from statistics. The calculation method is as follows:

$$\text{Z-Spread} = \frac{\text{Spread} - \mu_{\text{spread}(Slide_{Window}=55)}}{\delta_{\text{spread}(Slide_{Window}=55)}} \quad (4)$$

where μ_{spread} represents the mean of the spread and δ_{spread} represents the standard deviation of the spread. Both are computed within a sliding window, ensuring a dynamic analysis of the data.

3.3 Generating Labels Using Triple Barrier Labeling method (TBM)for Pair Trading Strategy

Triple Barrier Labeling method (TBM) is renowned for labeling individual assets. Adapting it for pair trading necessitates redefining parameters for spread in the pair trading context. Refer to Figure 4 for details.

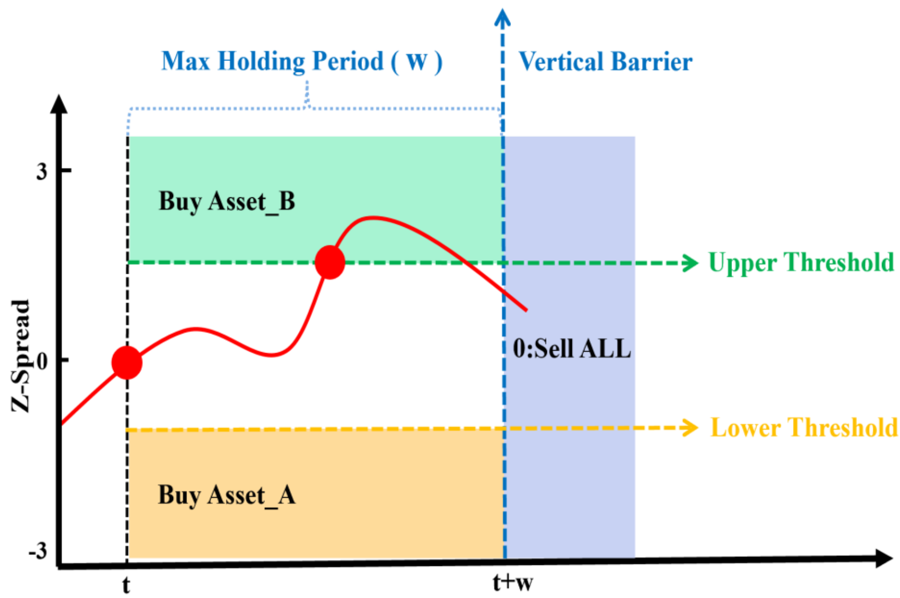

Figure 4. Triple Barrier Labeling method

From Figure 4, it is evident that within the Max Holding Period(W):

- **Upper Threshold**: The buy of Asset_B is triggered when the Z-Spread hits the Upper Threshold first.
- **Lower Threshold**: The buy of Asset_A is triggered when the Z-Spread hits the Lower Threshold first.
- **Max Holding Period**: During this period, if the Z-Spread does not exceed any threshold, sell all.

3.4 Optimizing by GA

In the P-Triple Barrier Labeling method, different hyperparameter values generate multiple trading signals, making it crucial to identify the best-performing signal with high profit and low risk. Nevertheless, balancing high returns and low risks concurrently in real-world trading situations poses a significant challenge. To address this challenge, we define two distinct signal styles: HRHP and LRLP. We optimize these parameters by genetic algorithm, aiming for signals that match both HRHP and LRLP styles. Refer to Figure 5 for details.

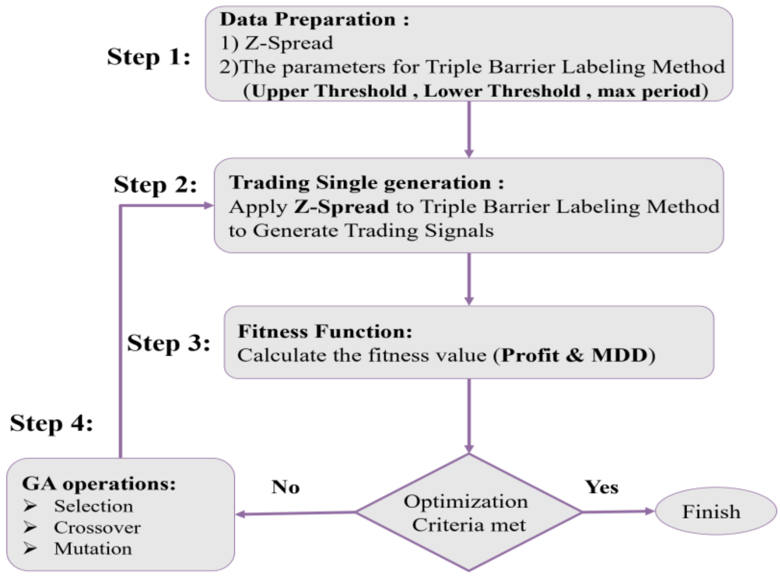

Figure 5. Optimizing the parameters of Triple Barrier Labeling using Genetic Algorithm (GA)

Figure 5 outlines the optimized process of P-Triple Barrier Labeling method using Genetic Algorithm (GA) in key steps:

Step 1: Data Preparation: Compute and standardize the spread between paired cryptocurrencies, defining upper and lower thresholds based on Z-Spread deviation.

Step 2: Trading Signal Generation: Use Z-Spread and GA-derived hyperparameters to create diverse trading signals in the Triple Barrier Labeling method.

Step 3: Fitness Function: Combine profit and MDD with varying weightings (e.g., 70% profit and 30% MDD for HRHP) to define unique trading styles.

Step 4:GA Operations: Apply the fitness function in GA's selection, crossover, and mutation processes for hyperparameters. Iterate until meeting optimization criteria, yielding optimal hyperparameter combinations.

These optimized values produce two sets of trading signals, HRHP and LRLP, offering valuable insights into different trading strategies.

3.5 Back Testing

We assessed our labeling method's performance using historical data, calculating cumulative returns (CR), maximum drawdowns (MDD), and Sharpe ratios. CR represents the cumulative profit over time, while MDD measures the largest loss from a peak to a trough. The Sharpe ratio evaluates portfolio performance after risk adjustment based on volatility and excess return [14]. This evaluation was done under real market conditions.

4. Experiment Setup and Result

4.1 Dataset

Considering the high-risk nature of the market, we specifically analyzed the cryptocurrency domain. Employing Yahoo Finance data [15], we conducted tests on various pairs, including Bitcoin (BTC) and

Ethereum (ETH). BTC and ETH displayed a strong correlation of 81.29% and an low cointegration P-value of 0.00. These superior results, outperforming other cryptocurrency pairs, prompted our selection of BTC and ETH for experimentation.

4.2 Hypermeters

In the optimization process of the P-Triple Barrier Labeling method, we utilize Genetic Algorithms (GA) to fine-tune three pivotal parameters: upper threshold, lower threshold, and maximum holding period. Typically, industry standards dictate setting the upper and lower thresholds as one standard deviation, representing a 100% deviation from the median—a common practice in pair trading. To comprehensively explore potential parameter combinations, we extend the search range for upper and lower thresholds from 0% to 300%. Given the high volatility of the cryptocurrency market, a long-term holding strategy proves suboptimal. Hence, we narrow the search range for the maximum holding period from 2 to 15 days. This approach not only considers market volatility but also preserves ample flexibility to achieve the optimal parameter combination.

4.3 Result

We utilized GA to optimize the P-Triple Barrier Labeling method with the objective of increasing CR and reducing MDD. However, it's not feasible to achieve high returns and low risks simultaneously in practical trading scenarios. Therefore, we adjusted the CR and MDD fitness combination ratio within the genetic algorithm, creating two distinct sets of labels: 70% profit and 30% MDD for HRHP, and 30% profit and 70% MDD for LRLP. These labels were utilized in back testing, and the outcomes were compared with traditional pair trading strategies as well as buy-and-hold strategies for Bitcoin and Ethereum. Please refer to Table 1 for detailed results.

Table 1. The Analysis Results of TBM Parameters Generated by GA

Strategy	Cumulative Return	Max Drawdown	Sharpe Ratio
Traditional Pair Trading Strategy [12]	39.9815	-0.1770	3.2924
Buy&Hold Bitcoin	7.1629	-0.5306	2.5537
Buy&Hold Ethereum	10.4285	-0.5002	2.6522
HRHP Label (17.30%,11.70%,2)	670.9349	-0.4247	3.5516
LRLP Label (66.20%,22.01%,2)	82.246	-0.1746	3.9392

*(Upper Threshold, Lower Threshold, Max Holding Period)

Examining Table 1 reveals that, although the HRHP labels yielded substantial profits (670.9349), they were counterbalanced by increased trading risks. Conversely, the LRLP labels offered reduced risks while maintaining profits twice (82.246) to traditional pair trading. Moreover, both of them have a higher Sharpe Ratio (3.5516 and 3.9392) than traditional pair trading strategy (3.2924), demonstrating a superior performance in risk-adjusted returns.

Table 1 compares hyperparameters for HRHP and LRLP label sets. HRHP labels have lower thresholds of 17.30% and 11.70%, increasing trading volume and profits but also risks. LRLP labels have high thresholds of 66.20% and 22.01%, reducing trading volume and profits but also risks. Both sets of labels had a Max Holding Period of 2, reflective of the short holding period characteristic of cryptocurrency's high volatility. The crucial task lies in selecting appropriate thresholds that strike a balance between stability, profitability, and investors' risk preferences.

5. Conclusion

In our research, we merged pair trading strategies with the Triple Barrier Labeling method, introducing an efficient financial labeling approach, and addressing a notable gap in existing literature. By refining the Triple Barrier Labeling method, we precisely identified buying and selling opportunities for paired assets in pair trading, thereby improving the accuracy and reliability of pair trading strategies. Extensive empirical analysis with real-world financial data validated the effectiveness of this labeling method, demonstrating significant enhancements in trading strategy accuracy and risk management. In essence, our study not only integrated powerful techniques but also advanced financial decision-making and future financial analyses, offering robust support for practical applications among financial professionals.

Acknowledgement

This research was supported by "Regional Innovation Strategy(RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2023RIS-008)

References

[1] Krauss, C. (2017). STATISTICAL ARBITRAGE PAIRS TRADING STRATEGIES: REVIEW AND OUTLOOK. Journal of Economic Surveys, 31(2), 513-545. https://doi.org/10.1111/joes.12153.
[2] Chen, Huafeng, et al. "Empirical investigation of an equity pairs trading strategy." Management Science 65.1 (2019): 370-389.
[3] Mudchanatongsuk, Supakorn, James A. Primbs, and Wilfred Wong. "Optimal pairs trading: A stochastic control approach." 2008 American control conference. IEEE, 2008.
[4] M. Fil and L. Kristoufek, "Pairs Trading in Cryptocurrency Markets," in IEEE Access, vol. 8, pp. 172644-172651, 2020, doi: 10.1109/ACCESS.2020.3024619.
[5] Kovačević, Tomislav, et al. "Effect of labeling algorithms on financial performance metrics." 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO). IEEE, 2022.
[6] De Prado, Marcos Lopez. Advances in financial machine learning. John Wiley & Sons, 2018.
[7] Kovačević, Tomislav, et al. "Optimal Trend Labeling in Financial Time Series." IEEE Access (2023).
[8] Khosravi, Morteza, and Majid Mirzaee Ghazani. "Novel insights into the modeling financial time-series through machine learning methods: Evidence from the cryptocurrency market." Expert Systems with Applications 234 (2023): 121012.
[9] Van der Have, R., et al. "Pairs trading using machine learning: An empirical study." Erasmus School of Economics 12 (2017).
[10] Mendelsohn, Louis B. Trend Forecasting with Intermarket Analysis: Predicting Global Markets with Technical Analysis. Vol. 92. John Wiley & Sons, 2012.
[11] Cohen, Israel, et al. "Pearson correlation coefficient." Noise reduction in speech processing (2009): 1-4.
[12] Hubrich, Kirstin, Helmut Lütkepohl, and Pentti Saikkonen. "A review of systems cointegration tests." Econometric Reviews 20.3 (2001): 247-318.
[13] Osifo, Ernest and Bhattacharyya, Ritabrata, Cryptocurrency Trading-Pair Forecasting, Using Machine Learning and Deep Learning Technique (March 5, 2020). Available at SSRN: https://ssrn.com/abstract=3610340 or http://dx.doi.org/10.2139/ssrn.3610340
[14] Sharpe, William F. "The sharpe ratio." Streetwise–the Best of the Journal of Portfolio Management 3 (1998): 169-85.
[15] [Yahoo. Yahoo Finance Python Package. https://pypi.org/project/yfinance/, yfinance 0.2.27