

Article

Enhanced Genetic-Algorithm-Driven Triple Barrier Labeling Method and Machine Learning Approach for Pair Trading Strategy in Cryptocurrency Markets

Ning Fu , Mingu Kang, Joongi Hong and Suntae Kim *

Department of Software Engineering, Jeonbuk National University, Jeonju 54896, Republic of Korea; funing230@jbnu.ac.kr (N.F.); 201914090@jbnu.ac.kr (M.K.); rlwns012@jbnu.ac.kr (J.H.)

* Correspondence: stkim@jbnu.ac.kr

Abstract: In the dynamic world of finance, the application of Artificial Intelligence (AI) in pair trading strategies is gaining significant interest among scholars. Current AI research largely concentrates on regression analyses of prices or spreads between paired assets for formulating trading strategies. However, AI models typically exhibit less precision in regression tasks compared to classification tasks, presenting a challenge in refining the accuracy of pair trading strategies. In pursuit of high-performance labels to elevate the precision of classification models, this study advanced the Triple Barrier Labeling Method for enhanced compatibility with pair trading strategies. This refinement enables the creation of diverse label sets, each tailored to distinct barrier configurations. Focusing on achieving maximal profit or minimizing the Maximum Drawdown (MDD), Genetic Algorithms (GAs) were employed for the optimization of these labels. After optimization, the labels were classified into two distinct types: High Risk and High Profit (HRHP) and Low Risk and Low Profit (LRLP). These labels then serve as the foundation for training machine learning models, which are designed to predict future trading activities in the cryptocurrency market. Our approach, employing cryptocurrency price data from 9 November 2017 to 31 August 2022 for training and 1 September 2022 to 1 December 2023 for testing, demonstrates a substantial improvement over traditional pair trading strategies. In particular, models trained with HRHP signals realized a 51.42% surge in profitability, while those trained with LRLP signals significantly mitigated risk, marked by a 73.24% reduction in the MDD. This innovative method marks a significant advancement in cryptocurrency pair trading strategies, offering traders a powerful and refined tool for optimizing their trading decisions.

Keywords: pair trading; triple barrier labeling method; cryptocurrency; genetic algorithm; AdaBoost classifier

MSC: 91G20; 68T05



Citation: Fu, N.; Kang, M.; Hong, J.; Kim, S. Enhanced Genetic-Algorithm-Driven Triple Barrier Labeling Method and Machine Learning Approach for Pair Trading Strategy in Cryptocurrency Markets. *Mathematics* **2024**, *12*, 780. <https://doi.org/10.3390/math12050780>

Academic Editors: Anatoliy Swishchuk and David Carfi

Received: 18 January 2024

Revised: 24 February 2024

Accepted: 29 February 2024

Published: 6 March 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

With the continuous advancement of Artificial Intelligence (AI) in the trading sector, an increasing number of studies are focusing on its integration with pair trading strategies to achieve higher profits amid market volatility. Pair trading strategy, grounded in mean reversion theory [1], focuses on exploiting price differences between closely related assets to achieve stable returns [2]. Currently, the primary focus of AI research in this area is on regression analyses of prices or spreads of paired assets, shaping the foundation of trading strategies. Despite AI's considerable potential in various fields, its accuracy in regression tasks is generally inferior to classification tasks, thereby constraining the improvement of the accuracy of pair trading strategies. Particularly since 2017, the rapid rise of cryptocurrency trading, driven by its high returns and inherent risks, has spurred researchers to investigate the synergistic application of AI and pair trading strategies in this domain. However, the significant volatility of the cryptocurrency market adds an extra

layer of complexity to this endeavor, making the pursuit of effective and precise strategies more challenging [3].

Considering the early stages of pair trading strategies integration in the cryptocurrency market and the scarcity of the research in this area, we offer a thorough examination of the field, dividing existing methodologies into two primary categories: traditional approaches and AI-incorporated strategies. The former involves the application of diverse statistical methods, particularly cointegration and distance approaches, for selecting trading pairs and developing hedging strategies, as detailed in [4–7]. This approach focuses on optimizing profitability and reducing risk, with its effectiveness validated by empirical market data analysis. Conversely, the latter category leverages advanced AI techniques such as machine learning and deep learning to forecast trends in cryptocurrency prices or the price spread between paired assets, thereby providing a more refined basis for making trading decisions (as referenced in [8–13]). Our analysis reveals that relying solely on historical cryptocurrency price data often leads to inadequate accuracy in identifying trading opportunities, particularly given the high volatility of the cryptocurrency market. Meanwhile, AI-enhanced methods, despite their innovation, encounter significant accuracy challenges, notably in regression analyses within the turbulent cryptocurrency market.

To address these challenges, we innovatively adapted the Triple Barrier Labeling Method proposed by De Prado and Marcos Lopez [14], thereby generating trading signals (labels) for the pair trading strategy. In this adaptation, single asset price sequences were transformed into spreads between paired assets, and the parameters of the Triple Barrier Labeling Method were adjusted. This refinement not only integrated the generated signals closely with pair trading strategies but also utilized varied barrier configurations to produce a diverse range of trading signals. Further optimization of these signals through GA results in the creation of two distinct types: High Risk and High Profit (HRHP) and Low Risk and Low Profit (LRLP). These refined signals are used to train machine learning models to accurately predict future trading behavior. We analyzed cryptocurrency data from 9 November 2017 to 31 August 2022 to train our models and used data from 1 September 2022 to 1 December 2023 for testing. Our models, trained with HRHP and LRLP labels, demonstrated significant performance improvements. The HRHP-trained model exhibited a notable 51.42% increase in profitability and lower Maximum Drawdown (MDD) compared to traditional pair trading strategies. Conversely, the LRLP-trained model excelled in profit generation while reducing the MDD by 73.24%, outperforming traditional strategies. This innovative approach, blending the pair trading strategy with AI technology, showcases a new perspective in cryptocurrency market trading strategies, enhancing both accuracy and stability.

The main contributions of this paper are summarized as follows:

- Integration of the pair trading principle and the Triple Barrier Labeling Method to generate potential trading signals.
- Utilization of a Genetic Algorithm to optimize trading signals to achieve both HRHP, as well as LRLP, resulting in diverse trading strategies.
- By choosing relatively simple machine learning classifiers over deep learning, we can mitigate overfitting and, consequently, improve prediction accuracy when confronted with the high volatility and noise inherent in cryptocurrency data.

The organization of the remaining sections in this paper is as follows. Section 2 is intended to review the background knowledge relevant to this paper and present a literature review of the related research. In Section 3, we present the implementation details of the proposed method. Section 4 outlines the experimental setup and analyzes the results obtained from the experiments. The conclusion of this paper is presented in Section 5.

2. Background and Related Work

2.1. Background

2.1.1. Pair Trading Strategy

The pair trading strategy, rooted in the mean reversion theory and introduced by Tullio Tattori in 1980 within Morgan Stanley's quantitative team [15], has become widely recognized for its exceptional performance in various trading scenarios [16]. Central to this strategy is the reliance on the stable relationship between highly correlated assets [17], which tends to maintain even amidst market volatility. This allows for a strategy that involves selling the asset that is performing better and purchasing the one that is underperforming when its prices diverge beyond a normal range.

Figure 1 exemplifies the strategy in action, showing the synchronized price movements of Asset_A and Asset_B, confirming their appropriateness as paired assets. Notably, a significant divergence observed on day 12 prompts an investment alert, guiding investors to make informed decisions based on a preset threshold for maximum price deviation. Consequently, on day 13, investors would be advised to sell the overperforming Asset_A and buy the underperforming Asset_B. As prices revert to normalcy from day 14, investors are alerted to reverse their trades, thereby maximizing profits.

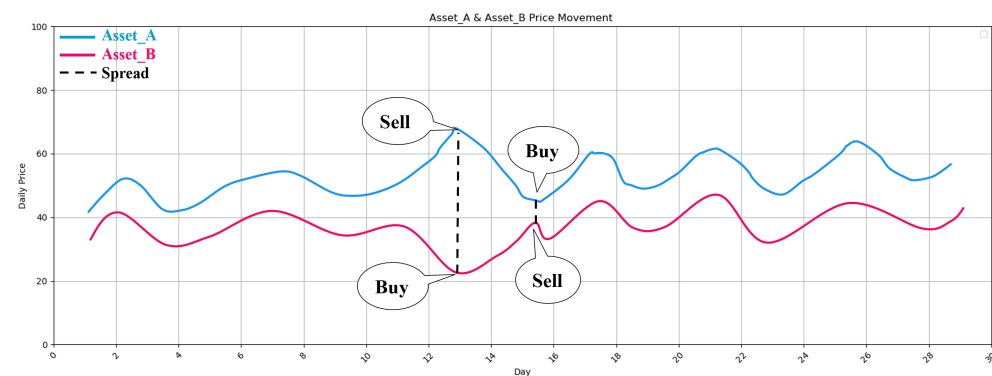


Figure 1. Pair trading strategy.

While the strategy excels in leveraging the relative value between assets, especially in volatile markets, it does face practical challenges in accurately calculating the spread and determining the optimal entry and exit points [18]. Despite these intricacies, pair trading's unique focus on the relative asset value, rather than the overall market direction, makes it a compelling choice in the dynamic landscape of trading strategies.

2.1.2. Triple Barrier Labeling Method

The Triple Barrier Labeling Method is a specialized financial technique used for labeling sample points, or observations, within time series data. This method is instrumental in enabling more in-depth data analysis and enhancing the effectiveness of model training, as discussed in [19]. It utilizes three specific barriers: the Profit Take Barrier; the Stop Loss Barrier; and the Maximum Holding Period, also known as the Vertical Barrier. The Profit Take Barrier functions as an indicator for potential buying opportunities, activated when the price of an observation exceeds this limit. Conversely, the Stop Loss Barrier serves to signal potential selling points, triggered when the price falls below its defined level. If an observation does not produce any returns during the Maximum Holding Period (the Vertical Barrier), it is given a label of 0, indicating an absence of profitable buying or selling opportunities. Further details and illustrations of this method can be found in Figure 2.

As indicated in Figure 2, The Triple Barrier Labeling Method dynamically generates labels within the Max Holding Period (w). The trading rules are as follows:

- If price hits the Profit Take Barrier first, Day_t is given the buy signal (1).
- If price hits the Stop Loss Barrier first, Day_t is given the sell signal (−1).
- If price hits the Vertical Barrier, Day_t is given the hold signal (0).

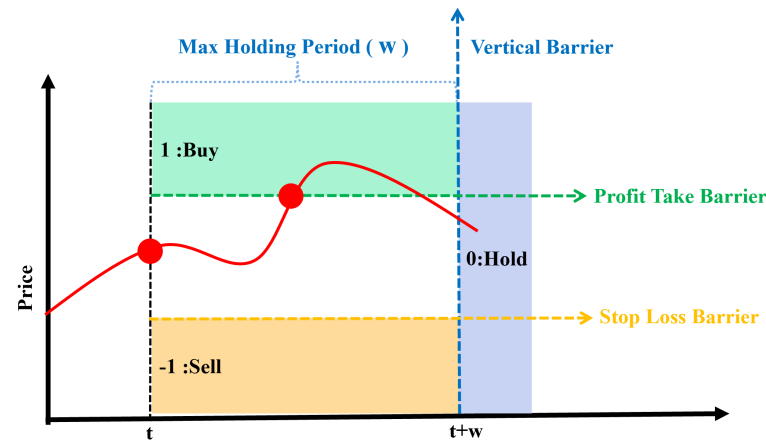


Figure 2. Triple Barrier Labeling Method.

2.2. Related Work

An increasing number of researchers are focusing on how to effectively apply pair trading strategies in the cryptocurrency market to enhance its profitability and reliability through technical analysis methods. Within these studies, several are endeavoring to directly incorporate pair trading strategies into the unique dynamics of the cryptocurrency market. For example, the research conducted by Miroslav Fil and Ladislav Kristouf [4]; Po-Chang Ko et al. [5]; Lesa, Chiara, and Hochreiter, Ronald [6]; along with Nair and Saji Thazhungal Govindan [7] employed a variety of statistical methods, including cointegration and distance approaches, to select trading pairs and devise hedging strategies. These research efforts aimed to assess and refine trading strategies, enhancing profitability while mitigating risk, with their efficacy demonstrated through real market data validation and comparison. Notably, these studies underscored the significance of evaluating strategy performance over different trading frequencies, such as intraday and hourly intervals, underscoring the necessity of selecting suitable trading strategies and frequencies to achieve profitability and effective risk management in the cryptocurrency domain. However, despite highlighting the strategies' effectiveness across various frequencies, these studies encountered limitations in accurately pinpointing specific trading opportunities.

Shifting the focus to AI-based methods, Osifo Ernest et al. [8] employed technologies such as Long Short-Term Memory networks (LSTM), Autoregressive Integrated Moving Average (ARIMA), Moving Average (MA), Cumulative Moving Average (CMA), and Artificial Neural Networks (ANN). Borges et al. [9] proposed a machine-learning-based system that uses a novel financial series resampling method based on closing price thresholds and four machine learning algorithms (Logistic Regression, Random Forest, Support Vector Classifier, and Gradient Boosting) to create high-return, low-risk trading strategies in the cryptocurrency market. Kalariya et al. [10] introduced a new trading strategy that combines mean reversion and stochastic neural networks for cryptocurrency price prediction. Long-term backtesting showed that it outperforms traditional buy-and-hold strategies in terms of stability and returns. Simão Moraes Sarmento and Nuno Horta [11] proposed a novel pair trading strategy that combines the unsupervised learning algorithm OPTICS with prediction models based on ARMA and LSTM, effectively enhancing the Sharpe ratio and reducing downturn periods in the portfolio. Vasu Kalariya and Pushpendra Parmar [12] trained random neural networks to predict cryptocurrency prices by using historical market data and social sentiment indicators, demonstrating an exceptional predictive performance. Jay Pate et al. [13] presented a random neural network model based on the Random Walk Theory for cryptocurrency price prediction, claiming its performance to be superior to traditional deterministic models. However, it is important to note that the significant volatility of the cryptocurrency market may influence the accuracy of regression methods, reducing the precision of predictions.

In machine learning, classification models often exhibit higher accuracy than regression models [20]. Therefore, our research initially generates trading signals (labels) for pair trading by using the Triple Barrier Labeling Method to produce labels. Next, we use a GA to optimize these labels and then train the machine learning model with the optimized labels. The experimental findings reveal that the model we propose excels not just in outperforming conventional pair trading and buy-and-hold strategies regarding profitability and risk management, but also shows marked superiority over the related research approaches discussed previously.

3. Methodology

The approach outlined in this study comprises two stages: label generation and Model Trading, as depicted in Figure 3. The first stage began with the selection of two cryptocurrency pairs, guided by their correlation and cointegration. We then proceeded to ascertain the spread between the paired cryptocurrencies. To ensure comparability and remove discrepancies in the scale, the spreads were standardized. Subsequently, this standardized spread (referred to as the Z-spread) was applied to an enhanced Triple Barrier Labeling Method to generate trading signals. Different barrier settings in this Triple Barrier Labeling Method result in different trading signals. Therefore, we employed a GA and used profit and the MDD as fitness values to optimize the parameters of the Triple Barrier Labeling Method. Ultimately, we obtained two types of trading signals, namely HRHP and LRLP.

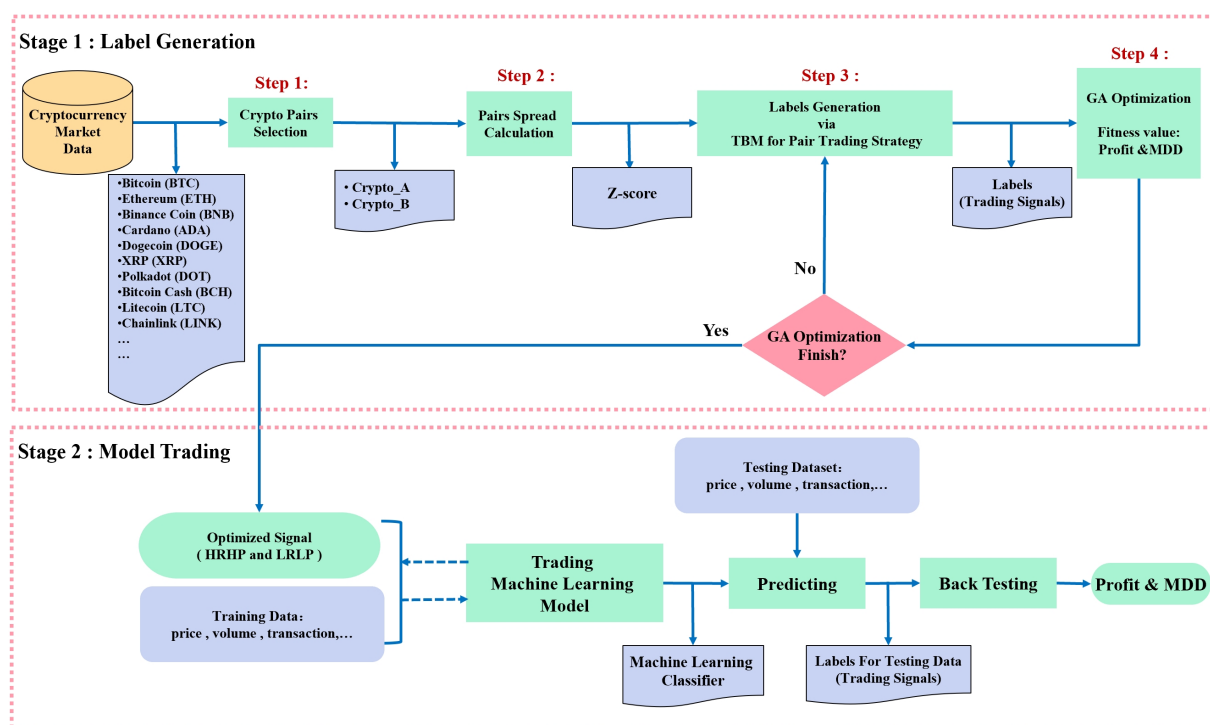


Figure 3. Signal Generation and machine learning model training framework.

In the second stage, we trained the machine learning model by using the HRHP and LRLP signals. After training, we utilized the trained model to predict trading signals for the test dataset. Then, we used these signals to simulate trading and calculate the corresponding profits and MDD.

3.1. Label Generation

3.1.1. Crypto Pairs Selection

Choosing appropriate cryptocurrency pairs is crucial for pair trading strategies. The ideal cryptocurrency pairs consist of two cryptocurrencies that demonstrate a strong correlation,

characterized by two key correlation aspects. Firstly, these two currencies usually move in the same direction, signifying synchronicity. Secondly, they maintain long-term trends and do not experience significant drift over time, upholding a consistent linear relationship. This selection process primarily depends on the correlation and cointegration criteria between two cryptocurrencies [21]. We perform pairwise combinations of all cryptocurrencies within the cryptocurrency market and subject them to correlation and cointegration tests. Here is an explanation of these tests:

1. **Pearson Correlation Coefficient:**

It measures the linear correlation between two cryptocurrencies, Crypto_A and Crypto_B , representing any two cryptocurrencies selected for analysis. The formula for r is

$$r = \frac{\sum_{t=1}^n (\text{Crypto_A}_t - \overline{\text{Crypto_A}})(\text{Crypto_B}_t - \overline{\text{Crypto_B}})}{\sqrt{\sum_{t=1}^n (\text{Crypto_A}_t - \overline{\text{Crypto_A}})^2} \sqrt{\sum_{t=1}^n (\text{Crypto_B}_t - \overline{\text{Crypto_B}})^2}}$$

where Crypto_A_t and Crypto_B_t represent the cryptocurrency prices of Crypto_A and Crypto_B on day t , respectively; $\overline{\text{Crypto_A}}$ and $\overline{\text{Crypto_B}}$ are the mean prices of Crypto_A and Crypto_B over a specified period (from 1 to n days), respectively; and where r ranges from -1 (perfect negative correlation) to 1 (perfect positive correlation), with 0 indicating no linear correlation [22].

2. **Cointegration Test:**

For two time series, Crypto_A and Crypto_B , we estimate their linear relationship by using regression analysis:

$$\text{Crypto_B}_t = \beta_0 + \beta_1 \text{Crypto_A}_t + \varepsilon_t$$

In this formula, β_0 represents the intercept, β_1 stands for the slope coefficient, and ε_t denotes the error term [23].

We then subject the residual sequence ε_t to a unit root test, typically the Augmented Dickey–Fuller (ADF) test [24]. The ADF test checks if ε_t has a unit root (nonstationarity) and follows this general form:

$$\Delta \varepsilon_t = \alpha + \beta t + \gamma \varepsilon_{t-1} + \delta_1 \Delta \varepsilon_{t-1} + \dots + \delta_{p-1} \Delta \varepsilon_{t-p+1} + \varepsilon_t$$

In the equation, $\Delta \varepsilon_t$ represents the first-order difference in the residual sequence to make it stationary. Here, α represents the constant term; β is the coefficient for the trend component; γ captures the autoregressive component of lag 1; δ_1 to δ_{p-1} represent the coefficients for the first-order differences in the lagged residuals, ensuring the model accounts for short-term fluctuations and seasonality in the data; and ε_t is the error term, indicating the randomness or unpredictability in the model at time t .

The objective of the ADF test is to determine if γ equals zero. If $\gamma = 0$, it suggests that the residual sequence is stationary, implying cointegration between Crypto_A and Crypto_B . If $\gamma \neq 0$, it indicates nonstationarity, suggesting no cointegration between Crypto_A and Crypto_B [25]. Furthermore, from a statistical standpoint, if the p-value of the ADF test conducted on the residual sequence falls below the significance level (usually 0.05), we can reject the null hypothesis. This rejection indicates that the residual sequence is stationary, suggesting the presence of cointegration between the original time series Crypto_A and Crypto_B .

In this study, if Crypto_A and Crypto_B are cointegrated with each other and exhibit a significantly high degree of correlation, they are selected as potential cryptocurrency trading pairs.

3.1.2. Pairs Spread Calculation

When implementing pair trading strategies, calculating the spread of selected cryptocurrency pairs has always been a significant challenge. In our approach, we analyze the

spread for the selected trading pair by using the method proposed by Osifo Ernest et al. [8], which relies on the spread in daily price changes between `Crypto_A` and `Crypto_B` to calculate trading signals. The calculation method is as follows:

$$\text{Spread} = \left(\frac{\text{Crypto_A}_T - \text{Crypto_A}_{T-1}}{\text{Crypto_A}_{T-1}} \right) - \left(\frac{\text{Crypto_B}_T - \text{Crypto_B}_{T-1}}{\text{Crypto_B}_{T-1}} \right)$$

where `Crypto_AT` and `Crypto_AT-1` represent the cryptocurrency prices of `Crypto_A` on day `T` and day `T-1`, respectively, and `Crypto_BT` and `Crypto_BT-1` represent the cryptocurrency prices of `Crypto_B` on day `T` and day `T-1`, respectively.

To ensure the reliability of the trading strategy, an ADF test is conducted once the spread is obtained. Trading operations can only proceed if the spread demonstrates stationarity [26]. If the spread fails to meet the stationarity condition, a re-evaluation of the spread calculation method becomes necessary. As the spread alone cannot intuitively express the deviation from the historical mean, we introduce the commonly used Z-spread from statistics [27]. The calculation method is as follows:

$$\text{Z-spread} = \frac{\text{Spread} - \mu_{\text{spread}(\text{Slide_Window}=55)}}{\delta_{\text{spread}(\text{Slide_Window}=55)}}$$

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 (e.g., `Slide_Window = 55`), ensuring a dynamic analysis of the data.

3.1.3. Labels Generation via TBM for Pair Trading Strategy

This section details the amalgamation of the Triple Barrier Labeling Method with the pair trading strategy. As the pair trading strategy is versatile, being applicable in both spot and futures markets, tailoring the trading rules to specific requirements becomes imperative. Notably, this study primarily centers on the spot market, and our customized trading rules are delineated in Figure 4.

Figure 4 delineates the market dynamics within a specific timeframe across three subplots, with each one illustrating a distinct set of data metrics. The top subplot illustrates the cumulative returns of `Crypto_A` and `Crypto_B`, computed from their daily change rates in adjusted closing prices. The middle subplot portrays the daily change rates spread between these two cryptocurrencies. Moreover, we established upper and lower thresholds by considering the mean and standard deviation of this spread, typically adding and subtracting them from the mean to create a rational range for the fluctuation differences. The bottom subplot displays the distribution of the Z-spread, along with its corresponding upper and lower thresholds. The calculation methods for these thresholds align with those employed to determine the upper and lower limits of the spread.

By observing Figure 4, it is evident that the spread in the middle subplot and the Z-spread in the bottom subplot surpassed the upper and lower thresholds at multiple time points. We paid particular attention to the portions annotated with dashed circles in these two subplots. It was observed that the spread failed to capture fluctuations surpassing the upper and lower thresholds, whereas the Z-spread accurately detected these fluctuations. This directly confirms that utilizing the Z-spread as a trading signal significantly enhances the accuracy of pair trading strategies. Hence, this study delved deep into the foundation of using the Z-spread as a trading signal.

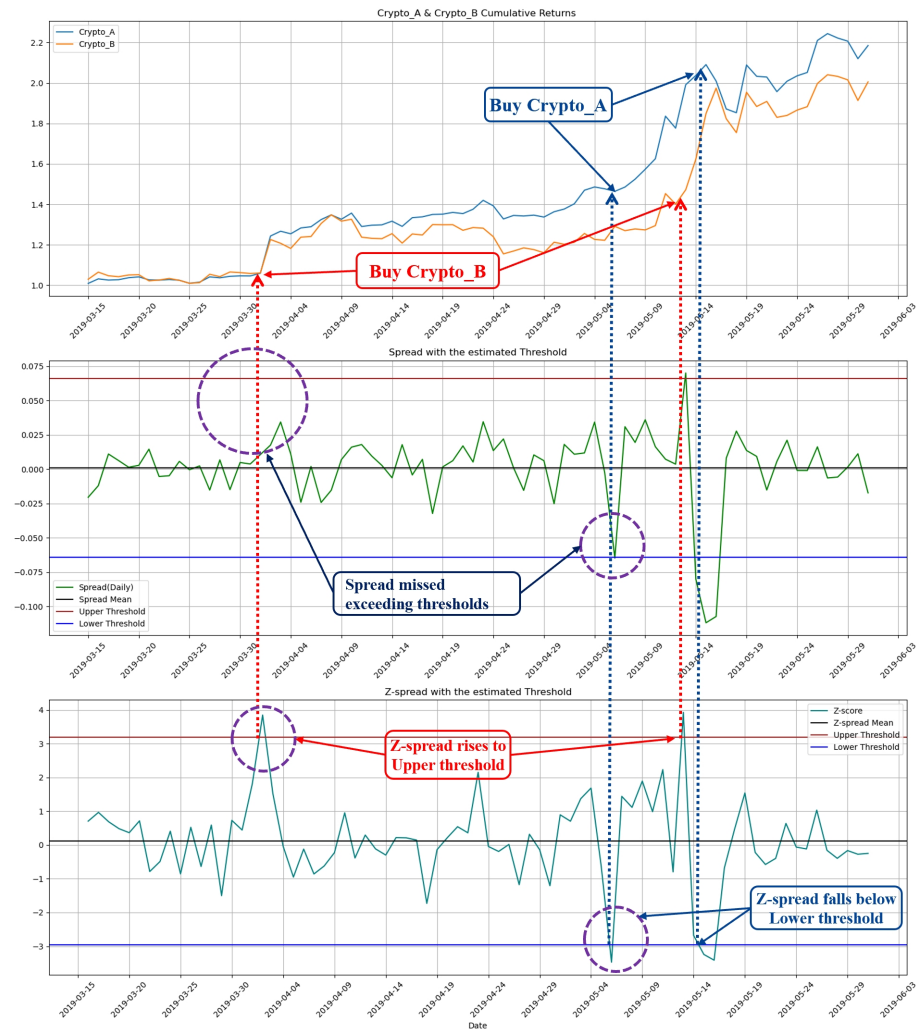


Figure 4. The principles of pair trading.

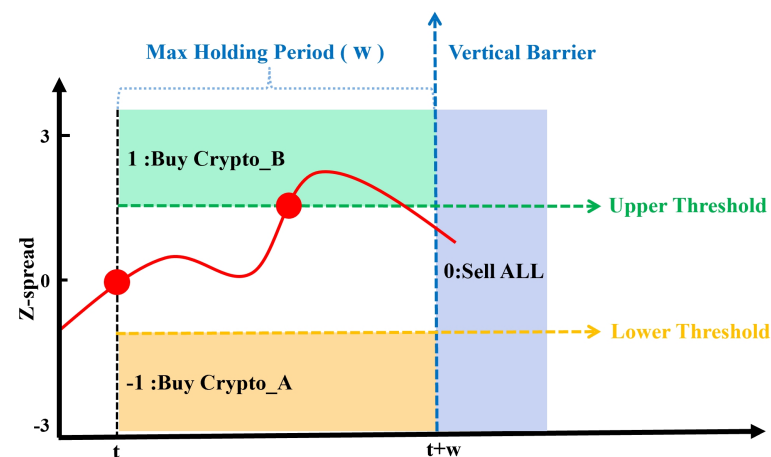
When the Z-spread rises above the Upper Threshold, it indicates a sudden expansion in the price gap between Crypto_A and Crypto_B. This suggests that Crypto_A is outperforming, potentially signaling that Crypto_B is undervalued. Therefore, buying Crypto_B at this point is advisable. Conversely, when the Z-spread falls below the lower threshold, it signifies a sudden narrowing of the price gap between Crypto_A and Crypto_B. This implies that Crypto_B is outperforming, and Crypto_A might be undervalued, thus it is advisable to buy Crypto_A at this juncture [19].

We integrate the pair trading strategy by using the Z-spread with the Triple Barrier Labeling Method to generate trading signals, a process we refer to as TBM for Z-spread. It is important to note that the Triple Barrier Labeling Method was originally designed for labeling methods by using a single asset. Therefore, incorporating the Z-spread into the Triple Barrier Labeling Method necessitates redefining its parameters. For further details, please refer to Table 1.

In Table 1, we present the four main parameters required for the Triple Barrier Labeling Method: the input sequence, upper barrier, lower barrier, and horizontal barrier. The table compares the application of the Triple Barrier Labeling Method to single cryptocurrencies and pair trading spreads. By redefining these parameters, the Triple Barrier Labeling Method can be seamlessly integrated with pair trading strategies. Figure 5 provides a visual representation of the application of the Triple Barrier Labeling Method to pair trading.

Table 1. Redefinition of parameters of TBM for Z-spread.

Parameter	TBM for Single Cryptocurrency	TBM for Z-Spread
Input Sequence	Historical price data of Individual Cryptocurrency	Historical Z-value of spread between paired cryptocurrencies.
Top Barrier	Profit-take barrier: the buy signal (1) is triggered when the price hits the profit-take barrier first.	Upper Threshold: the buy of Crypto_B (1) is triggered when the Z-spread hits the Upper Threshold first.
Bottom Barrier	Stop-loss barrier: the sell signal (−1) is triggered when the price hits the stop-loss barrier first.	Lower Threshold: the buy of Crypto_A (−1) is triggered when the Z-spread hits the lower threshold first.
Horizontal Barrier	Max Holding Period: during this period, the trading conditions are not met, and no trade occurs (0).	Max Holding Period: during this period, the movement of the Z-spread does not exceed any threshold, and all is sold (0).

**Figure 5.** Triple Barrier Labeling Method to pair trading.

From Figure 5, it is evident that within the Max Holding Period(W):

- If the Z-spread hits the Upper Threshold first (spread increases), Day_t is given the buy Crypto_B signal (1).
- If the Z-spread hits the Lower Threshold first (spread decreases), Day_t is given the buy Crypto_A signal (−1).
- If the Z-spread hits the Vertical Barrier (spread regression mean), Day_t is given the sell-all signal (0).

3.1.4. GA Optimization

The Genetic Algorithm (GA) operates through a series of systematic steps to evolve solutions toward optimality. The primary operations involved are:

- **Selection:** At iteration t , the selection operation in population P_t evaluates each chromosome c by using the fitness function $f(c)$ for suitability to the problem. It formalizes choosing the highest-scoring chromosomes to form a new population as

$$P_{t+1} = \text{Select}(P_t, f) = \{c' \mid c' \text{ is selected from } P_t \text{ based on } f(c)\}$$

where P_{t+1} is the new population generated from P_t by selecting chromosomes with higher fitness scores.

- **Crossover:** The crossover operation is applied to pairs of chromosomes (c'_1, c'_2) to produce offspring (c''_1, c''_2) :

$$(c''_1, c''_2) = \text{Cross}(c'_1, c'_2) = (c'_1 \otimes c'_2, c'_2 \otimes c'_1)$$

where $\otimes$ represents the crossover mechanism that combines genetic information from parent chromosomes.

- **Mutation:** Mutation is applied to offspring chromosomes c''_1 and c''_2 to produce mutated offspring c'''_1 and c'''_2 :

$$c'''_1 = \text{Mutate}(c''_1) = c''_1 \oplus \mu, \quad c'''_2 = \text{Mutate}(c''_2) = c''_2 \oplus \mu$$

where $\oplus$ denotes the mutation operation with μ representing the mutation effect, introducing random genetic variations to offspring chromosomes to enhance population diversity.

These operations enable the GA to explore and exploit the solution space effectively, driving the population toward an optimal solution over successive generations.

In the Triple Barrier Labeling Method, different settings of the three hyperparameters (Upper Threshold, Lower Threshold, and Max Holding Period) yield multiple sets of trading signals. Thus, it is essential to identify the best-performing signal among these sets. The best trading signal is characterized by achieving high profit with low risk. However, in practical applications, it is challenging to simultaneously achieve high profit and low risk. To tackle this issue, we can define two distinct trading signal styles: HRHP and LRLP. We fine-tune these three hyperparameters by adjusting the fitness values within the GA, aiming to derive trading signals that align with both the HRHP and LRLP styles. The detailed execution process is illustrated in Figure 6.

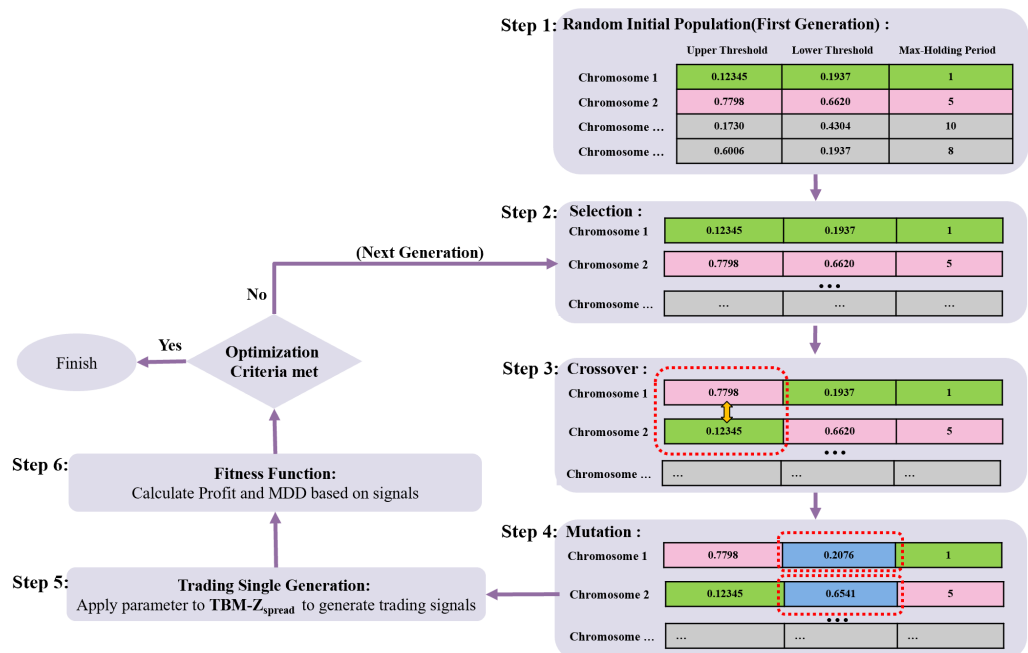


Figure 6. Optimizing the parameters of Triple Barrier Labeling using Genetic Algorithm (GA).

The process can be broken down into the following steps:

Step 1: Random Initial Population (First Generation)

A random initial population, denoted as P_0 , is generated where each chromosome $c \in P_0$ represents a potential solution with parameters like the Upper Threshold, Lower Threshold, and Max Holding Period, ensuring a diverse genetic variety across the search space.

Step 2: Selection

From the current population P_t , a subset P_{t+1}^{sel} is selected for crossover and mutation based on fitness:

$$P_{t+1}^{sel} = \text{Select}(P_t, f) = \{c' \mid c' \text{ is chosen from } P_t \text{ based on fitness } f(c)\}$$

Step 3: Crossover

Offspring P_{t+1}^{cross} are produced by applying the crossover operation Cross to pairs of chromosomes in P_{t+1}^{sel} :

$$P_{t+1}^{cross} = \{c'' \mid c'' = \text{Cross}(c'_1, c'_2) \text{ for } (c'_1, c'_2) \in P_{t+1}^{sel}\}$$

where Cross denotes the crossover function combining genetic information from parent chromosomes.

Step 4: Mutation

A mutation operation Mutate is applied to chromosomes in P_{t+1}^{cross} to produce P_{t+1}^{mut} :

$$P_{t+1}^{mut} = \{c''' \mid c''' = \text{Mutate}(c'') \text{ for } c'' \in P_{t+1}^{cross}\}$$

where Mutate introduces random variations to the genetic makeup of offspring chromosomes.

Step 5: Trading Signal Generation

The offspring population P_{t+1}^{mut} is utilized as input parameters for the Triple Barrier Method (TBM) for the Z-spread to generate trading signals S_{t+1} .

Step 6: Fitness Function

The fitness function f is defined to evaluate profit and the Maximum Drawdown (MDD) based on the generated signals:

$$f(c) = w_p \cdot \text{Profit}(c) + w_{mdd} \cdot \text{MDD}(c)$$

where w_p and w_{mdd} represent the weights assigned to the profit and MDD, respectively, whereby allocating 70% of the weight to the profit and 30% to the MDD yields HRHP signals, while allocating 30% of the weight to the profit and 70% to the MDD results in LRLP signals; these weights are used to select the optimal signals for the corresponding trading styles.

Iterate through this process continuously until it meets the optimization-stopping criteria. Eventually, determine the best hyperparameter combination. With these optimized hyperparameters, two different types of trading signals can be generated: HRHP and LRLP.

3.2. Model Training

In the training phase of the supervised learning model, we utilize the optimized trading signals obtained from Stage 1. These signals, along with their corresponding feature data (such as their close price, volume, transactions, etc.), are input into the machine learning classifiers for training. After the completion of model training, we input the test data into the trained model, generating specific trading signals (labels). Subsequently, we simulate trading activities based on these signals, enabling us to perform backtesting for the profit and MDD calculations.

It is crucial to emphasize that the accuracy of the classifier is not the sole metric for assessing its performance. In the context of time series prediction tasks, overfitting can occur, where the model demonstrates good performance during training but fails to deliver satisfactory results in real-world applications. Consequently, we evaluate the model based on the profit and the MDD obtained from simulating trades by using the predicted trading signals.

4. Experimental Setup

4.1. Research Question

- **Research Question 1 (Label Evaluation):** how can the labels generated by Triple Barrier Labeling for pair trading be evaluated in terms of profit and the MDD?
- **Research Question 2 (Model Evaluation):** how to choose a classifier for the predictive model and evaluate its performance in comparison to other models?
- **Research Question 3 (Prediction Evaluation):** what is the predictive performance of models trained by using label types obtained through genetic optimization, precisely the label types of HRHP or LRLP?

Research Question 1 focuses on evaluating the labels generated through the combination of Triple Barrier Labeling and pair trading. Labels play a critical role in training supervised learning models as they significantly impact machine learning performance. To gain a more intuitive understanding of the label performance, we selected profit and the MDD as evaluation criteria for the labels.

Research Question 2 explores the logic behind selecting a classifier. We trained a diverse set of machine learning and deep learning models by using acquired labels and pertinent training data. Following this, we conducted an exhaustive analysis to evaluate the predictive capabilities of these models.

Research Question 3 aims to evaluate the models trained by using the labels. This question holds central importance in our research. By assigning distinct weights to the profit and MDD during label generation, we obtain two sets of trading signals representing HRHP as well as LRLP. The models trained based on these two label sets exhibit different trading styles. Research Question 3 seeks to validate the performance of these two trading styles in terms of profit and the MDD, comparing them to other baseline methods through simulated trading scenarios.

4.2. Datasets Selection

Selecting a pertinent dataset is essential for the efficacy of our proposed methodology. The experimental dataset, derived from the dynamic realm of the cryptocurrency market and accessed through Yahoo Finance [28], was systematically sampled on a daily basis. Identifying synergistic trading pairs within the cryptocurrency domain is of paramount importance. Employing the methods for correlation and cointegration testing delineated in Section 3.1.1, we undertook a comprehensive pairwise analysis of the amassed cryptocurrency data. The granular outcomes of these analytical endeavors are delineated in Tables 2 and 3.

Table 2. Cointegration test results (p -value).

	BTC_DCR	ETH_DCR	XRP_DCR	BCH_DCR	USDT_DCR	LTC_DCR	EOS_DCR	BNB_DCR	XLM_DCR
BTC_DCR	-	0.0	2.5823×10^{-9}	0.0	5.3832×10^{-28}	0.0	7.0451×10^{-21}	1.1565×10^{-21}	0.0
ETH_DCR	0.0	-	1.3920×10^{-11}	0.0	3.1697×10^{-16}	0.0	3.2104×10^{-22}	1.7274×10^{-12}	6.7014×10^{-13}
XRP_DCR	1.6868×10^{-24}	2.9006×10^{-14}	-	4.3842×10^{-6}	0.0	2.4593×10^{-24}	4.2209×10^{-6}	8.4754×10^{-25}	1.2729×10^{-26}
BCH_DCR	0.0	0.0	0.0	-	6.0809×10^{-27}	0.0	1.0356×10^{-26}	0.0	0.0
USDT_DCR	4.2671×10^{-29}	4.9538×10^{-29}	4.3132×10^{-29}	4.1570×10^{-29}	-	4.2532×10^{-29}	4.0557×10^{-29}	4.5918×10^{-29}	4.2635×10^{-29}
LTC_DCR	0.0	0.0	4.9574×10^{-17}	0.0	7.3420×10^{-27}	-	0.0	0.0	0.0
EOS_DCR	9.0577×10^{-20}	2.6114×10^{-18}	1.8214×10^{-21}	7.5792×10^{-25}	0.0	1.3194×10^{-19}	-	2.6026×10^{-16}	9.9547×10^{-21}
BNB_DCR	6.1991×10^{-12}	4.5600×10^{-8}	2.5209×10^{-07}	3.4266×10^{-7}	5.3136×10^{-10}	3.3866×10^{-7}	1.9140×10^{-10}	-	1.5985×10^{-10}
XLM_DCR	3.3884×10^{-28}	9.9568×10^{-14}	0.0	0.0	0.0	2.3647×10^{-28}	0.0	0.0	-

DCR: daily change rate.

Table 2 presents the cointegration relationships between daily change rates (DCRs) in various cryptocurrencies. We employ the statistical p -value to assess the presence of cointegration between pairs of cryptocurrencies. A p -value below a specific significance level (typically 0.05) leads to rejecting the null hypothesis, signifying cointegration. Conversely, a p -value exceeding the significance level suggests a lack of cointegration. From Table 2, it

is evident that the p -values for the BTC_DCR-ETH_DCR, BTC_DCR-LTC_DCR, BTC_DCR-BCH_DCR, ETC_DCR-BCH_DCR, LTC_DCR-BCH_DCR, and ETH_DCR-LTC_DCR pairs are all 0.0, indicating cointegration among these six cryptocurrency pairs.

Table 3. Correlation Test Results.

	BTC_DCR	ETH_DCR	XRP_DCR	BCH_DCR	USDT_DCR	LTC_DCR	EOS_DCR	BNB_DCR	XLM_DCR
BTC_DCR	-	<u>0.8128</u>	0.5686	0.7351	0.0005	0.7875	0.7015	0.6343	0.6081
ETH_DCR	<u>0.8128</u>	-	0.6264	0.7649	−0.0375	0.8049	0.7577	0.6631	0.6659
XRP_DCR	0.5686	0.6264	-	0.6059	−0.0184	0.6405	0.6709	0.5110	0.7046
BCH_DCR	0.7351	0.7649	0.6059	-	−0.0179	0.8038	0.8010	0.5684	0.6326
USDT_DCR	0.0005	−0.0375	−0.0184	−0.0179	-	−0.0193	−0.0182	−0.0391	−0.0204
LTC_DCR	0.7875	0.8049	0.6405	0.8038	−0.0193	-	0.7999	0.6423	0.6497
EOS_DCR	0.7015	0.7577	0.6709	0.8010	−0.0182	0.7999	-	0.6123	0.6902
BNB_DCR	0.6343	0.6631	0.5110	0.5684	−0.0391	0.6423	0.6123	-	0.5284
XLM_DCR	0.6081	0.6659	0.7046	0.6326	−0.0204	0.6497	0.6902	0.5284	-

DCR: daily change rate.

Table 3 illustrates the correlation relationships in the DCR among different types of cryptocurrencies. The Pearson correlation coefficient is employed to assess the correlation between two cryptocurrencies, where -1 signifies a perfect negative correlation, 1 indicates a perfect positive correlation, and 0 implies no linear correlation. From Table 3, it is observed that the correlation coefficient for the BTC_DCR-ETH_DCR pair is 0.8128, higher than other cryptocurrency pairs. Considering the cointegration relationship observed in Table 2 for the BTC_DCR-ETH_DCR pair, BTC and ETH were selected as the paired cryptocurrencies for our subsequent experimental research.

In this study, we adopted a supervised learning methodology within the realm of machine learning. The dataset spanning from 9 November 2017 to 31 August 2022 was employed for label generation and model training, while the dataset covering 1 September 2022 to 1 December 2023 was employed for testing and evaluating the model's performance. Traditional pair trading strategies utilize straightforward data, frequently confined to individual price data. In Table 4, we compare the input data of our proposed method with that of traditional pair trading strategies.

Table 4. Comparison of input data for purposed method and traditional pair trading strategy.

	Pair Trading Strategy	Purposed Method
Input Data	Historical price data for paired cryptocurrencies (e.g., adj close price).	All historical features related to paired cryptocurrencies, such as high, low, close, adj close, volume, transaction, Z-spread, etc.

From Table 4, it is evident that traditional pair trading strategies rely solely on single price data, such as the adj close price of the cryptocurrency. In contrast, our proposed model leverages more comprehensive data features, including the daily low, high, close, volume, and Z-spread of the price difference calculated based on the adj close of the two cryptocurrency pairs. This rich feature set provides additional trading information, enhancing the model's learning capability and improving its predictive performance.

4.3. Hyperparameters for GA

When using the Triple Barrier Labeling Method to generate labels for pair trading, varying the values of hyperparameters can yield multiple sets of labels. To obtain optimal labels with high returns and low risk, we introduced a Genetic Algorithm (GA) to optimize the labels. The hyperparameters utilized by the GA are presented in Table 5.

Table 5. Search space of hyperparameters.

Parameter	Searching Space
Upper Threshold (percentage)	1–300%
Lower Threshold (percentage)	1–300%
Max Holding Period(day)	1–10
Fitness Value	Profit and MDD

In Table 5, the Upper Threshold and Lower Threshold parameters indicate the percentage by which the Z-spread deviates from the mean, relative to one standard deviation. Typically, the Z-spread deviation range for traditional pair trading strategies is established between 0.1 and 3 standard deviations [29]. Accordingly, the search space for the Upper Threshold is set from 1% to 300%. Similarly, the search space for the Lower Threshold is set from 1% to 300%. Given the high volatility of cryptocurrencies, which entails increased risk in long-term holdings, we set the search space for the Max Holding Period from 1 to 10. Additionally, this study employs a combination of profit and the MDD as the fitness value for the Genetic Algorithm. By adjusting the weight between profit and the MDD, trading signals can be obtained in HRHP or LRLP.

4.4. Market Conditions and Strategy Robustness Analysis

To ensure the robustness and adaptability of our trading strategy and the assets and indicators it relies on for practical applications, we carefully selected the period from 9 November 2017 to 31 August 2022 to generate the trading labels. During this period, the cryptocurrency market experienced several cycles of bubble bursts and price rebounds, covering a wide range of market trends including bull and bear markets, various levels of market volatility, and diverse economic cycles. The choice of this time range allowed our analysis to cover a broad market environment, ensuring that the effectiveness analysis of the selected assets and indicators could be conducted under a variety of market conditions. Therefore, the assets and indicators used in our study not only have their effectiveness confirmed through historical data analysis but also demonstrate strong adaptability to future market changes. The application of this methodology aims to ensure that our trading strategy and its supporting models remain effective and applicable in the face of market variability in the future.

4.5. Baseline Methods

This research aims to boost profits and minimize risks in cryptocurrency pair trading by using sophisticated machine learning classification models. Among the extensive body of research on pair trading strategies, the method introduced by Osifo Ernest et al. [8] is recognized as state-of-the-art, making it one of our selected baselines. Additionally, in AI-based research, Vasu Kalariya and Pushpendra Parmar [12] trained random neural networks by using historical market data and social sentiment indicators to predict cryptocurrency prices, with their outstanding outcomes also establishing a benchmark. Furthermore, we included the buy-and-hold strategies for the cryptocurrency pairs BTC and ETH as two additional baselines.

4.6. Experimental Result

Research Question 1 (Label Evaluation): how can the labels generated by Triple Barrier Labeling for pair trading be evaluated in terms of profit and the MDD?

The objective of this question is to evaluate the labels generated by combining Triple Barrier Labeling with pair trading. We employ a Genetic Algorithm (GA) with profit and the MDD as the combined fitness value to optimize these parameters. By adjusting the weights between profit and the MDD, we obtain two sets of trading signals representing distinct styles: HRHP and LRLP. Table 6 presents a comparison of the simulated trading profit and

MDD using these two sets of trading signals with the profit and MDD of traditional pair trading strategies and the buy-and-hold strategies for BTC and ETH.

Table 6. Comparison of simulated trading profit and MDD for different signal sets.

Trading Period: 9 November 2017–31 August 2022			Hyperparameter for Labeling Method		
Strategies	Profit (Times)	MDD (Percentage)	Upper Threshold (Growth of Z-spread %)	Lower Threshold (Reduction of Z-spread %)	Max Holding Period
HRHP Label	386.9837	−0.5516	17.3043	11.7092	2
LRLP Label	46.9574	−0.4337	66.2016	22.0194	2
Buy-and-Hold Bitcoin	3.1537	−0.7185			
Buy-and-Hold Ethereum	4.9373	−0.7935			
Pair Trading	20.7737	−0.5358			

Table 6 reveals that the simulated trading profits derived from the HRHP and LRLP labels, post-Genetic Algorithm optimization, surpass the buy-and-hold strategies of Bitcoin and Ethereum, along with outperforming the conventional pair trading approach. The comparative analysis demonstrates that the model trained by using HRHP labels generates a profit of 386.9837 times the initial investment amount, a substantial improvement over the profits achieved by traditional pair trading strategies. However, this is balanced by the HRHP label's accompanied MDD of −0.5516, indicative of a comparatively heightened trading risk. Conversely, the profits stemming from the LRLP label are more than double that of the traditional pair trading strategy, while its MDD stands at −0.4337, pointing to a relatively reduced trading risk.

Additionally, the table provides a comparative analysis of the hyperparameters utilized for generating the HRHP and LRLP label sets. Notably, the HRHP labels exhibit an Upper Threshold of 17.3043% and a Lower Threshold of 11.7092%. These comparatively smaller hyperparameter values facilitate the Z-spread's likelihood of exceeding the Upper Threshold and Lower Threshold, consequently inducing an upswing in the trading volume. As the trading volume amplifies, profits ascend, albeit accompanied by an escalated level of associated risk.

Conversely, the LRLP labels manifest an Upper Threshold of 66.2016% and a Lower Threshold of 22.0194%, highlighting the utilization of higher values for these hyperparameters. This leads to a proportional contraction in the trading volume, ultimately yielding diminished profits alongside a mitigated level of risk. Furthermore, both sets of labels share a Max Holding Period hyperparameter set at two, emphasizing the inherently brief holding period characteristic of the heightened volatility intrinsic to cryptocurrencies.

Research Question 2 (Model Evaluation): how to choose a classifier for the predictive model and evaluate its performance in comparison to other models?

This research question aims to compare the performance of different AI models in a classification task. We conducted experiments by using various classifiers, including both machine learning and deep learning techniques. The prediction results in accuracy and recall are presented in Table 7.

Based on the findings presented in Table 7, it is evident that the relatively intricate deep learning models did not demonstrate the substantial advantages that were initially anticipated. Specifically, the accuracy and recall rates of the Multilayer Perceptron Classifier stand at 0.6216 and 0.5473, respectively, while the composite model (CNN + LSTM + Attention) classifier exhibits similar performance trends with accuracy and recall rates of 0.6216 and 0.5417, respectively. Surprisingly, in comparison to these two models, the classifier solely utilizing LSTM showcases slightly higher accuracy and recall rates, reaching 0.6392 and 0.5500, respectively.

During experimentation, it was discovered that the Extreme Gradient Boosting Classifier outperformed alternative classification models, achieving accuracy and recall rates

of 0.7208 and 0.5597, respectively. As a result, we selected the Extreme Gradient Boosting Classifier as the foundational classifier and amalgamated it with the AdaBoost Classifier to elevate the accuracy and recall rates to 0.7903 and 0.6621, correspondingly. These empirical findings further reinforce the conclusion that machine learning models excel over deep learning models in effectively managing data characterized by high volatility and noise, a phenomenon observed within the cryptocurrency market context [30,31].

Table 7. Performance comparison of various AI models in a classification task.

AI Model	Accuracy	Recall	AUC
Multilayer Perceptron Classifier	<u>0.6216</u>	<u>0.5473</u>	0.5458
LSTM Classifier	<u>0.6392</u>	<u>0.5500</u>	0.5408
(CNN + LSTM + Attention)Classifier	<u>0.6216</u>	<u>0.5417</u>	0.5477
Ridge Classifier	0.6484	0.5550	0.5657
Support Vector Classification	0.5083	0.4350	0.4746
Extreme Gradient Boosting Classifier	<u>0.7208</u>	<u>0.5597</u>	0.6299
LightGBM Classifier	0.6518	0.5460	0.5956
Stochastic Gradient Descent Classifier	0.6061	0.4892	0.4929
Decision Tree Classifier	0.6922	0.5987	0.6167
Gaussian Process Classifier	0.6238	0.5333	0.5472
AdaBoost Classifier	<u>0.7903</u>	<u>0.6621</u>	<u>0.6409</u>
AdaBoost Classifier with Bayesian data mining	0.7732	0.6192	0.6335

Additionally, we explored a new data mining strategy by using Bayesian methods and Markov chain Monte Carlo techniques to enhance the robustness of our model. This attempt was based on the research by Giudici and Paolo [32], with the aim of augmenting the AdaBoost Classifier’s ability to handle complex datasets. Specifically, we utilized these techniques for feature selection and data transformation, aiming to improve the model’s adaptability to complex data structures and its predictive accuracy. However, as shown in Table 7, the performance of the AdaBoost Classifier employing the Bayesian data mining method was slightly inferior to the traditional AdaBoost model. This result suggests that the Bayesian data mining method might not be suitable for dealing with highly volatile cryptocurrency data.

In evaluating the performance of our classification models, we utilized not only the conventional metrics of accuracy and recall but also explored the use of other metrics, such as the Rank Graduation Accuracy (RGA). However, in our experiment, the label data consisted of time series, with each label (−1, 0, and 1) representing a trend or state shift (decrease, stable, and increase), reflecting the sequential change in events over time. As described by Raffinetti [33], The RGA metric is specifically designed to evaluate the accuracy of a system’s ability to correctly rank or recommend items, focusing on the relative positioning of these items. However, our research focus is on the precision of time series forecasting, paying special attention to the order and dependencies of events in the time series rather than the relative ranking of items. It is apparent that the RGA metric is not entirely suitable for assessing label data in time series from our experiment, suggesting the need to find more appropriate evaluation methods for such sequential data.

Therefore, in the assessment of model performance, we further compared the AUC values of different classifiers. As shown in Table 7, the AdaBoost Classifier, with an AUC value of 0.6409, outperformed other classifiers, which further confirms the credibility and reliability of using the AdaBoost Classifier as our classification model.

Research Question 3 (Prediction Evaluation): what is the predictive performance of models trained by using label types obtained through genetic optimization, precisely the label types of HRHP and LRLP?

This research question aims to evaluate the predictive capabilities of models trained by using HRHP and LRLP labels. The performance of the models trained by using two types of labels is showcased in Table 8. It is important to note that conventional metrics such as accuracy and recall were not employed as evaluation criteria for model performance. Instead, we utilized trading signals predicted by an AdaBoost Classification model and assessed the models based on simulated trading profit and the MDD. Our objective in adopting this method is to mitigate the bias introduced by overfitting during the model training process and achieve more realistic prediction outcomes.

Table 8. Comparison of model performance based on HRHP and LRLP labels.

Testing Period: 1 September 2022–1 December 2023			
	Profit (Times)	MDD (Percentage)	Sharpe Ratio
HRHP-trained model	<u>2.0754</u>	−0.1611	<u>1.8491</u>
LRLP-trained model	1.7034	<u>−0.0691</u>	<u>2.0016</u>
AI regression model	<u>1.4499</u>	<u>−0.2047</u>	0.9216
Pair trading strategy	<u>1.3702</u>	<u>−0.2579</u>	0.8476
Buy-and-hold Bitcoin	1.8737	−0.2942	1.3073
Buy-and-hold Ethereum	1.2940	−0.3806	0.6517

The testing period selected for this study spans from 1 September 2022 to 1 December 2023. Upon examining Table 8, it becomes apparent that both the HRHP- and LRLP-trained models outperform other benchmark methods in terms of profit and the MDD. Specifically, the model trained with HRHP labels demonstrated high profitability with a profit of 2.0754, significantly surpassing the AI regression analysis model's 1.4499 and the traditional pair trading strategy's 1.3702 profit. Conversely, the LRLP model exhibited remarkable risk control, with its MDD at −0.0691, better than the AI regression analysis model's −0.2047 and the traditional pair trading strategy's −0.2579 MDD. The cumulative profit curves, showcasing the performance of the models trained with HRHP and LRLP labels along with other trading strategies, are respectively presented in Figures 7 and 8.

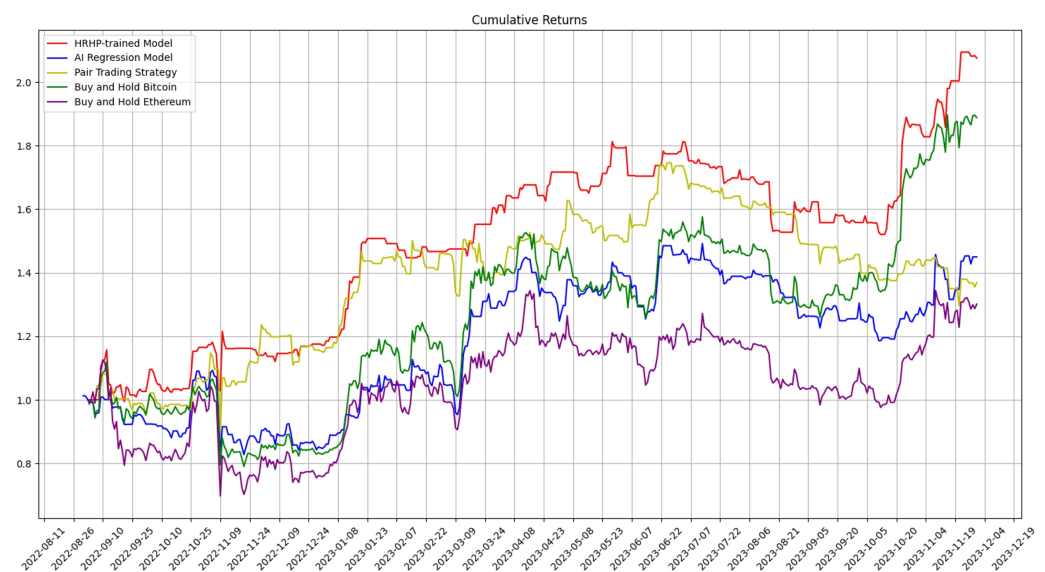


Figure 7. Cumulative return by HRHP label-trained model vs. baseline methods.

Additionally, an analysis of the Sharpe ratio indicates that the LRLP model performs exceptionally well in terms of risk-adjusted returns, with a ratio of 2.0016, surpassing the HRHP model's 1.8491. This suggests that the LRLP model might be a better choice in

unstable markets. The higher Sharpe ratio means that LRLP can deliver higher returns for the same level of risk. In contrast, while the HRHP model offers higher profits, it comes with greater risk.

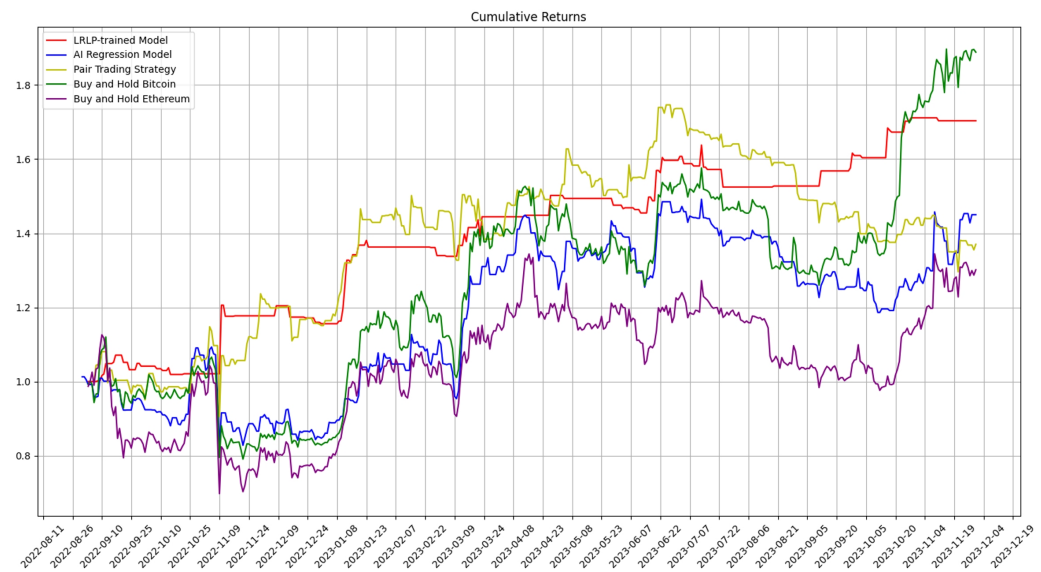


Figure 8. Cumulative return by LRLP label-trained model vs. baseline methods.

Based on the analysis of Figures 7 and 8, it is evident that the models trained using our approach outperform other benchmark methods in terms of simulated trading profit and the MDD. Upon comparing the two figures, we observe that the HRHP-trained model engages in a higher frequency of trades compared to the LRLP-trained model. As a result, the HRHP-trained model demonstrates a relatively higher profit and MDD while the LRLP-trained model showcases a more stable performance with fewer trades but a higher winning rate.

In our research, we delved into the cumulative return rates predicted by the model, focusing particularly on which cryptocurrency primarily drove gains or losses. The experiment revealed that the overall cumulative return rate for the HRHP model was 2.0754 times, whereas for ETH and BTC, it was 2.2895 times and 1.5526 times, respectively. This finding, after thorough verification, showed inconsistencies between the sum of individually calculated return rates and the overall return rate for the pair trading strategy. This discrepancy highlights the nonlinear compounding effects [34] and the compound interest effect [35] in executing hedging trading strategies, aligning with Luenberger [36] and emphasizing the crucial role of compound interest. Hence, we concluded that merely calculating the cumulative returns of ETH and BTC separately to analyze which cryptocurrency mainly influenced profits or losses is an inadequate approach.

5. Threats to Validity

5.1. Randomness Impact of Genetic Algorithms (GAs)

In our research, Genetic Algorithms (GAs) are extensively used as an optimization tool for model parameters. However, the inherent nature of the GA includes a certain level of randomness, which might pose a threat to the robustness of our research findings. This randomness primarily stems from the random selection process involved in the generation of initial populations, crossover, and mutation operations. This means that even under identical experimental setups, each run of the GA might yield slightly different results.

5.2. Mitigation Measures

To address this issue, we adopted the following mitigation measures to reduce the impact of randomness on the validity of our research findings:

- **Multiple Runs of GA:** By conducting multiple runs of the GA and averaging the results, we can reduce the random fluctuations brought by a single run, thereby obtaining more stable and reliable optimization outcomes. This method ensures that our research findings are not the result of random chance but are observable across different runs.
- **Employing Alternative Verification Methods:** Besides the GA, we explored other parameter optimization and verification methods, such as cross-validation and sensitivity analysis, to further validate our research findings. These methods offer additional perspectives for evaluating the performance and robustness of the model, enhancing the credibility of our research results.

5.3. Impact on Research Validity

Despite the challenges posed by the randomness of the GA, through the above mitigation measures, we were able to effectively reduce the impact of this randomness on our research findings. These measures not only enhance the robustness of our research but also strengthen the reliability of our conclusions. By adding a discussion on these threats and mitigation strategies to our paper, we aim to provide readers with a more complete and transparent presentation of our research process, as well as an honest discussion on the limitations of our study.

6. Conclusions

This study aims to utilize a novel approach by integrating the Triple Barrier Labeling Method and machine learning approach with pair trading strategies, aiming to amplify profitability and minimize risk. Firstly, we integrate pair trading strategies with the Triple Barrier Labeling Method to generate trading signal labels. Through the use of a GA with varying weights for profit and the MDD, we obtain two sets of trading signal labels representing different trading styles: HRHP and LRLP. Subsequently, we train an AdaBoost Classifier model by using these two sets of labels and conduct predictions.

To validate the predictive performance of our model, we utilize a cryptocurrency dataset from 2017 to 2022 for training purposes, while data from 2022 to 2023 serve as the testing set. The experimental results demonstrate that the two distinct trading models trained by using our proposed method outperform traditional pair trading strategies and the buy-and-hold strategy for individual cryptocurrencies in terms of profit and MDD metrics. Thus, we can assert that this research method achieved a state-of-the-art performance.

The primary contribution of this study lies in the integration of pair trading strategies with the Triple Barrier Labeling Method. We employ the deviation in the Z-spread of the price difference for selected cryptocurrency pairs from the mean as the foundation for generating the Triple Barrier Labeling labels. This approach effectively combines the mean regression characteristics of pair trading strategies with the dynamic generation of trading signals by using the Triple Barrier Labeling Method, providing an innovative solution for capturing patterns in pair trading.

To regulate the frequency of trades, we opted for a daily sampled dataset instead of smaller time intervals such as 5 min or 1 h. However, this choice may result in some information loss, which can potentially impact the accuracy of the analysis results. To address this concern, our future work will explore the application of our proposed method to datasets with smaller time intervals. Additionally, we will undertake a broader selection of classification models, including the consideration of the Temporal Convolutional Network (TCN), WaveNet, TimesNet, and others, to further enhance profitability and reduce risk.

Author Contributions: Methodology, N.F., M.K., J.H. and S.K.; Software, N.F., M.K., J.H. and S.K.; Validation, J.H. and S.K.; Writing—original draft, N.F.; Writing—review & editing, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (RS-2023-00259099) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation).

Data Availability Statement: No new data were created or analyzed in this study. The data used in this study were obtained from the Yahoo Finance Python Package, yfinance version 0.2.27, available online: <https://pypi.org/project/yfinance/>. Data sharing is not applicable to this article as all data are publicly available.

Conflicts of Interest: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- Krauss, C. Statistical arbitrage pairs trading strategies: Review and outlook. *J. Econ. Surv.* **2017**, *31*, 513–545. [\[CrossRef\]](#)
- Jackson, A.B. Financial statement analysis: A review and current issues. *China Financ. Rev. Int.* **2022**, *12*, 1–19. [\[CrossRef\]](#)
- Fang, F.; Ventre, C.; Basios, M.; Kanthan, L.; Martinez-Rego, D.; Wu, F.; Li, L. Cryptocurrency trading: A comprehensive survey. *Financ. Innov.* **2022**, *8*, 13. [\[CrossRef\]](#)
- Fil, M.; Kristoufek, L. Pairs Trading in Cryptocurrency Markets. *IEEE Access* **2020**, *8*, 172644–172651. [\[CrossRef\]](#)
- Ko, P.C.; Lin, P.C.; Do, H.T.; Kuo, Y.H.; Mai, L.M.; Huang, Y.F. Pairs trading in cryptocurrency markets: A comparative study of statistical methods. *Invest. Anal. J.* **2023**, 1–18.
- Lesa, C.; Hochreiter, R. Cryptocurrency Pair Trading. 2023. Available online: <https://ssrn.com/abstract=4433530> (accessed on 28 February 2024).
- Nair, S.T.G. Pairs trading in cryptocurrency market: A long-short story. *Invest. Manag. Financ. Innov.* **2021**, *18*, 127–141.
- Osifo, E.; Bhattacharyya, R. Cryptocurrency Trading-Pair Forecasting, Using Machine Learning and Deep Learning Technique. 5 March 2020. Available online: <https://ssrn.com/abstract=3610340> (accessed on 28 February 2024).
- Borges, T.A.; Neves, R.F. Ensemble of machine learning algorithms for cryptocurrency investment with different data resampling methods. *Appl. Soft Comput.* **2020**, *90*, 106187. [\[CrossRef\]](#)
- Kalariya, V.; Parmar, P.; Jay, P.; Tanwar, S.; Raboaca, M.S.; Alqahtani, F.; Tolba, A.; Neagu, B.C. Stochastic neural networks-based algorithmic trading for the cryptocurrency market. *Mathematics* **2022**, *10*, 1456. [\[CrossRef\]](#)
- Sarmiento, S.M.; Horta, N. Enhancing a Pairs Trading strategy with the application of Machine Learning. *Expert Syst. Appl.* **2020**, *158*, 113490. [\[CrossRef\]](#)
- Khedr, A.M.; Arif, I.; El-Bannany, M.; Alhashmi, S.M.; Sreedharan, M. Cryptocurrency price prediction using traditional statistical and machine-learning techniques: A survey. *Intell. Syst. Accounting, Financ. Manag.* **2021**, *28*, 3–34. [\[CrossRef\]](#)
- Jay, P.; Kalariya, V.; Parmar, P.; Tanwar, S.; Kumar, N.; Alazab, M. Stochastic Neural Networks for Cryptocurrency Price Prediction. *IEEE Access* **2020**, *8*, 82804–82818. [\[CrossRef\]](#)
- De Prado, M.L. *Advances in Financial Machine Learning*; John Wiley & Sons: Hoboken, NJ, USA, 2018.
- Elliott, R.J.; Van Der Hoek, J.; Malcolm, W.P. Pairs trading. *Quant. Financ.* **2005**, *5*, 271–276. [\[CrossRef\]](#)
- Vidyamurthy, G. *Pairs Trading: Quantitative Methods and Analysis*; John Wiley & Sons: Hoboken, NJ, USA, 2004; Volume 217.
- Bogomolov, T. Pairs trading based on statistical variability of the spread process. *Quant. Financ.* **2013**, *13*, 1411–1430. [\[CrossRef\]](#)
- Meng, T.L.; Khushi, M. Reinforcement learning in financial markets. *Data* **2019**, *4*, 110. [\[CrossRef\]](#)
- Hairudin, A.; Sifat, I.M.; Mohamad, A.; Yusof, Y. Cryptocurrencies: A survey on acceptance, governance and market dynamics. *Int. J. Financ. Econ.* **2022**, *27*, 4633–4659. [\[CrossRef\]](#)
- Papakyriakou, D.; Barbounakis, I.S. Data mining methods: A review. *Int. J. Comput. Appl.* **2022**, *183*, 5–19. [\[CrossRef\]](#)
- Leung, T.; Nguyen, H. Constructing cointegrated cryptocurrency portfolios for statistical arbitrage. *Stud. Econ. Financ.* **2019**, *36*, 581–599. [\[CrossRef\]](#)
- Naeem, M.A.; Karim, S.; Abrar, A.; Yarovaya, L.; Shah, A.A. Non-linear relationship between oil and cryptocurrencies: Evidence from returns and shocks. *Int. Rev. Financ. Anal.* **2023**, *89*, 102769. [\[CrossRef\]](#)
- Shumway, R.H.; Stoffer, D.S. Time series regression and exploratory data analysis. In *Time Series Analysis and Its Applications: With R Examples*; Springer: New York, NY, USA, 2006; pp. 48–83.
- Phillips, P.C.; Xiao, Z. A primer on unit root testing. *J. Econ. Surv.* **1998**, *12*, 423–470. [\[CrossRef\]](#)
- Ssekuma, R. A Study of Cointegration Models with Applications. Master's Thesis, University of South Africa, Pretoria, South Africa, 2011.
- Murat, A.; Tokat, E. Forecasting oil price movements with crack spread futures. *Energy Econ.* **2009**, *31*, 85–90. [\[CrossRef\]](#)
- Mare, D.S.; Moreira, F.; Rossi, R. Nonstationary Z-score measures. *Eur. J. Oper. Res.* **2017**, *260*, 348–358. [\[CrossRef\]](#)
- Yahoo. Yahoo Finance Python Package. yfinance 0.2.27. Available online: <https://pypi.org/project/yfinance/> (accessed on 28 February 2024).
- Carvalho, D.d.S. Pairs Trading: Cointegration-Based Methods: Applied to the Cryptocurrency Market. Ph.D. Thesis, Catholic University of Portugal, Lisbon, Portugal, 2021.

30. Velankar, S.; Valecha, S.; Maji, S. Bitcoin price prediction using machine learning. In Proceedings of the 2018 20th International Conference on Advanced Communication Technology (ICACT), Chuncheon, Republic of Korea, 11–14 February 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 144–147.
31. Chowdhury, R.; Rahman, M.A.; Rahman, M.S.; Mahdy, M. An approach to predict and forecast the price of constituents and index of cryptocurrency using machine learning. *Phys. Stat. Mech. Its Appl.* **2020**, *551*, 124569. [[CrossRef](#)]
32. Giudici, P. Bayesian data mining, with application to benchmarking and credit scoring. *Appl. Stoch. Model. Bus. Ind.* **2001**, *17*, 69–81. [[CrossRef](#)]
33. Raffinetti, E. A rank graduation accuracy measure to mitigate artificial intelligence risks. *Qual. Quant.* **2023**, *57*, 131–150. [[CrossRef](#)]
34. Sharpe, W.F.; Alexander, G.J.; Bailey, J.V. *Investment*; Prentice Hall Incorporated: Hoboken, NJ, USA, 1999.
35. Malkiel, B.G. *A Random Walk Down Wall Street: The Time-Tested Strategy for Successful Investing*; W. W. Norton & Company, Inc.: New York, NY, USA, 2021.
36. Luenberger, D.G. *Investment Science*; Oxford University Press: Oxford, UK, 1998.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.