



SCIENTIFIC OASIS

Journal of Operations Intelligence

Journal homepage: [www.jopi-journal.org](http://www.jopi-journal.org)  
eISSN: 3009-4267

# Predictive Stock Selection: A Hybrid RF-CNN XGBoost Model Integrated with Dynamic Adaptive Index and Stepwise Elimination Techniques

Farshid Abdi<sup>1</sup>, Shaghayegh Abolmakarem<sup>1</sup>, Amir Karbassi Yazdi<sup>2,\*</sup><sup>1</sup> Department of Industrial Engineering, ST.C., Islamic Azad University, Tehran, Iran<sup>2</sup> Departamento de Ingeniería Industrial y de Sistemas, Facultad de Ingeniería, Universidad de Tarapacá, Arica 10000, Chile

## ARTICLE INFO

### Article history:

Received 9 February 2025

Received in revised form 27 April 2025

Accepted 19 May 2025

Available online 24 May 2025

### Keywords:

Stock Selection, Stock prediction, Random Forest, XGBoost, Dynamic Adaptive Index.

## ABSTRACT

This paper presents a predictive model for stock selection using a hybrid approach. The proposed model, which combines advanced machine learning algorithms with a dynamic adaptive index (DAI), aims to enhance portfolio performance and risk management. In this context, Random Forest (RF) and XGBoost algorithms are utilized to select optimal features and forecast stock returns. In the subsequent step, the DAI method assesses and selects stocks. By integrating various metrics such as correlation, volatility, relative strength index (RSI), and cumulative returns, the DAI method enables a comprehensive assessment of available stocks that dynamically adapts to changing market conditions. In addition, the stepwise elimination technique optimizes the portfolio by eliminating stocks with high volatility and low returns, ensuring that only stocks with strong performance and acceptable risk are retained. The two-stage structure of the proposed model, encompassing stock return prediction and stock selection, enhances the model's overall efficiency. The presented hybrid model excels in stock selection and risk management, helping investors achieve higher returns while reducing portfolio risk.

## 1. Introduction

Financial investors are faced with the possibility of choosing a large number of stocks to build their portfolio in the stock market [1]. Thus, stock selection is recognized as a challenging subject in financial market analysis research. Techniques for stock selection aim to differentiate stocks with high-profit potential from other stocks using evaluation or scoring models. Furthermore, stock selection can be a preliminary step in developing an optimal portfolio to identify target stocks with higher value. The use of forecasting methods can enhance understanding of future stock behavior and facilitate the selection process [2]. In fact, the ability of investors to generate profits depends largely on their forecasting abilities. One of the main reasons for investors' losses is their limited

\* Corresponding author.

E-mail address: [akarbassy@academicos.uta.cl](mailto:akarbassy@academicos.uta.cl)<https://doi.org/10.31181/jopi31202545>© The Author(s) 2025 | [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)

ability to accurately predict future stock performance. Consequently, developing an effective stock selection model to enhance investors' predictive ability is a significant and relevant topic [3]. This has led to considerable interest in stock prediction among financial investors, prompting researchers to propose various models. These models include statistical models such as integrated autoregressive moving average (ARIMA) models, exponential smoothing, and generalized autoregressive conditional variance (GARCH) models etc. Due to the non-stationary or nonlinear characteristics of the time series data, statistical models are not very capable of predicting these data, and hence, in order to overcome the limitations of statistical methods, researchers have turned to machine learning algorithms, including radial basis function multilayer perceptron, support vector machines (SVM) and support vector regression (SVR) [1]. Deep learning methods such as Deep Multilayer Perceptron [4], Long Short-Term Memory (LSTM), and Convolutional Neural Networks (CNNs) have emerged as the powerful methods for forecasting time series [5]. On another note, selecting features for predicting future stock market behavior significantly enhances the stock selection process. Several studies have employed open, high, low, and closing prices, technical indicators, and fundamental variables, among others, as inputs for forecasting models [6].

Accordingly, this paper proposes a predictive stock selection model that integrates several feature selections (FSS) methods and stock return forecasting algorithms based on deep learning, using a combined approach involving the Dynamic Adaptive Index (DAI) and the stepwise elimination technique. Central to the proposed stock selection method is the prediction of future stock returns. This method comprises two main steps: predicting stock returns and selecting stocks. In the first step, stock returns for the next ten days are projected using the RF-CNN-XGBoost technique. In the second step, attention is given to optimal stock selection, employing the Dynamic Adaptive Index (DAI) method in conjunction with the stepwise elimination technique for this purpose. The contributions of the proposed approach in this paper are as follows:

- i. This paper introduces a hybrid methodology combining Random Forest and XGBoost for return prediction. Random Forest identifies important features, minimizing model complexity and enhancing performance. Subsequently, XGBoost is trained on the selected features, effectively learning nonlinear relationships within the data.
- ii. This paper presents an innovative stock selection method based on the Dynamic Adaptive Index (DAI) which adapts dynamically to fluctuating market conditions. DAI offers a thorough assessment of stocks by integrating various metrics, including correlation, volatility, relative strength index (RSI), and cumulative return. Additionally, the stepwise elimination technique discards stocks with high volatility and low returns, ensuring that only robust-performing stocks with acceptable risk remain in the portfolio. This approach appears to have been underutilized in previous research.
- iii. The method proposed in this paper features a two-step process: stock return prediction and stock selection. This two-step framework enhances the model's efficiency.

The remainder of the paper is organized as follows: Section 2 reviews previous research findings. Section 3 elaborates on the proposed hybrid approach. Section 4 presents the experimental results, and Section 5 concludes the paper.

## **2. Literature of Past Works**

Ma *et al.*, [7] introduced an enhanced stock selection model that relies on price and volume factors. The authors use a blend of wavelet transformation and multi-task attention networks in this model. This model sought to enhance prediction precision in the Chinese stock market, which poses considerable difficulties for conventional approaches due to its substantial volatility and policy ambiguities. Wavelet transform applied to break down stock returns time series into multiple sub-

series with varying frequency levels, allowing the model to effectively detect both long-term trends and short-term volatility. Alzaman [8] proposed a reinforcement learning model to improving stock selection through ranking and pairing stocks. The combination of technical indicators, past stock prices, and various unconventional factors has led to more precise stock predictions. This method has been confirmed across different market conditions (such as bullish, bearish, and economic downturns) and outperformed conventional techniques. Sarkar *et al.*, [9] examine a hybrid model for intelligent stock portfolio selection that merges forecasting and optimization into a single framework. The authors aim to enhance investment returns and mitigate risk by more accurately modelling stock market behaviour. This approach comprises two principal steps: stock price forecasting and portfolio optimization. In the first stage, deep learning models such as RNN, LSTM, CNN-LSTM are employed to accurately predict stock prices. Following the forecasting phase, the second stage involves optimization methods such as mean-variance or meta-heuristic algorithms (like particle swarm optimization) to develop a portfolio that strikes a favourable balance between return and risk. Utilizing actual market data, this paper demonstrates that the integration of prediction and optimization can enhance portfolio performance relative to traditional methods. Furthermore, empirical results indicate that this approach maintains stable performance across varying market conditions, including bullish and bearish periods. Gao *et al.*, [10] investigated a stock selection model that employs Neural Recursive Ordinary Differential Equation Networks and relation fusion. The aim of this research is to enhance the accuracy of stock prediction and selection by modelling the complex relationships between stocks and leveraging multi-source data. The authors applied recurrent differential equations to model the temporal dynamics of stocks, combined with relational fusion techniques to extract more effective features from the data. This method facilitates the examination of interactions among different market variables, leading to more accurate predictions of price trends. The proposed model was tested with actual market data, revealing its superior performance in comparison to traditional methods. Lv *et al.*, [11] introduced a hybrid system for stock selection and portfolio optimization that utilizes deep learning and an improved NSGA-III algorithm. The primary aim of this research is to present a novel investment management approach that yields higher returns and lower risks for investors. In their study, the authors employed a CNN-BiLSTM hybrid model to forecast stock price trends and outlined a multi-objective optimization problem encompassing five criteria including mean, variance, higher order moments such as skewness, kurtosis, and distance to default) to create an optimal portfolio. In this paper, the improved NSGA-III algorithm with Tent Chaotic Mapping and the fixed penalty function method have been used to solve the optimization problem. In the proposed model, appropriate stocks are first selected using the CNN-BiLSTM model. Then, by combining the traditional mean-variance model with higher-order moments, the optimal portfolio is created. Lakshmi *et al.*, [12] introduced a new model for stock portfolio selection that employs a combination of MCDM methods, including the Fuzzy Analytic Hierarchy Process (Fuzzy AHP) and the TOPSIS technique. The authors developed an accurate and flexible decision-making system for stock selection and portfolio optimization under financial uncertainty. After identifying key financial and non-financial criteria, the authors weighted them using fuzzy AHP, then the modified TOPSIS method was used to rank the stocks. Alzaman *et al.*, [13] employed a combination of deep learning methods to analyze and predict financial time series data, showing that this approach outperforms traditional methods in price forecasting and portfolio management. The results indicate that deep learning models can yield higher Sharpe ratios, reduce risk, and provide greater cumulative returns. Brito [14] presented an innovative model for stock portfolio selection that incorporates the concepts of expected utility, entropy, and variance. The proposed model utilizes the utility function to assess investors' preferences concerning risk and returns. The authors apply the concept of entropy to measure uncertainty and variability in stock

returns, which allows the model to generate portfolios with a more balanced and stable return distribution. Variance is also employed as a traditional risk measure in this model to select portfolios with lower risk and more optimal returns. Silva *et al.*, [15] presented a portfolio optimization model that employed a support vector machine (SVM) for initial stock selection. The authors utilized the SVM algorithm as a supervised learning tool to identify stocks with high return potential. This model selects a suitable set of assets for portfolio formation by analyzing the financial and technical characteristics of stocks. After preselection, the Mean-Variance model is used to assign weights to stocks and create the portfolio. Thakur *et al.*, [16] proposed stock selection and portfolio optimization model using Dempster–Shafer evidence theory. The authors aimed to minimize the uncertainty and costs associated with the development of expert systems in stock selection. This method effectively handles incomplete and uncertain information. Four key indicators, including price-to-sales ratio (P/S), price-to-earnings ratio (P/E), price-to-book value ratio (P/B), and long-term debt-to-equity ratio (LTDER), are chosen as inputs to the fuzzy system. Following the ranking of stocks based on fuzzy output, a portfolio optimization model is implemented using the ant colony optimization (ACO) algorithm. Putra *et al.*, [17] investigate an innovative approach to selecting stock portfolios from the KOMPAS-100 index that employs a B-Spline-based clustering algorithm. The main goal of this method is to improve the process of stock selection and portfolio weighting through more accurate analysis of financial data. In this approach, stocks with similar characteristics are initially grouped more accurately due to the high flexibility of B-Spline in modelling complex data. Subsequently, stocks with the best performance or growth potential are selected from each cluster, and optimal capital is allocated among assets using the mean-variance model. Experiments conducted with actual KOMPAS-100 index data demonstrate that portfolios constructed with this method yield higher returns and lower risks compared to traditional methods. Additionally, this model has successfully created greater diversity within the portfolio and provided more stable performance under varying market conditions. Chen *et al.*, [18] proposed a hybrid optimization model that applied skewness and kurtosis to assess the distribution of stock returns time series. The authors' applied clustering techniques to select assets with high diversity. Stock return forecasting was done using a radial basis function neural network and a genetic algorithm (GA) used to solve the optimization model. Yan *et al.*, [19] introduce an advanced model for stock selection that employs hypergraph neural networks and a conceptual attention mechanism. In this approach, a conceptual attention mechanism is utilised to assess key stock features within a hypergraph framework, facilitating the identification of nonlinear relationships among stocks. The hypergraph neural network is also leveraged to capture complex relationships between stocks, encompassing temporal (historical) and spatial (inter-stock) dependencies. Experimental results across different markets indicate that this framework outperforms traditional methods and other deep learning models in predicting stock returns, resulting in higher Sharpe ratios and reduced portfolio risk. Caparrini *et al.*, [20] investigated the impact of various factors such as fundamental and technical indicators on the stock selection process using ML methods. Random Forest, Support Vector Machine (SVM), and Neural Networks are employed to predict stock performance. Additionally, the study examines the influence of traditional factors like the price-to-earnings ratio (P/E) as well as newer factors such as behavioural and sentiment data on stock return predictions. This approach has resulted in the formation of portfolios generating higher returns and lower risk, demonstrating through more accurate data analysis. Asawa *et al.*, [21] applied advanced machine learning techniques in investment resource management and asset allocation optimisation. This study applied combination of deep CNN-LSTM with classical mean-variance methods to improve the accuracy of prediction model. The results demonstrate that utilizing these methods enhances portfolio returns and mitigates risk, allowing the proposed models to outperform indices like the Nifty Index. Padhi *et al.*, [22] present a hybrid intelligent model for stock

market forecasting that employs machine learning techniques for portfolio selection. The primary goal of this model is to process diverse financial data using hybrid techniques to deliver more accurate information for market forecasting. The proposed model leverages machine learning algorithms to identify stocks with high return potential and optimize asset allocation within the portfolio. By integrating various features such as historical price data and technical indicators, this model improves forecast accuracy and reduces errors associated with traditional methods. The employed methods consist of Deep Neural Networks and optimization algorithms for data analysis. The findings indicate that the proposed model excels in predicting market trends and managing risk compared to traditional methods, aiding investors in making more optimal decisions. Rasekhschaffe *et al.*, [23] examine the application of machine learning techniques in the stock selection process and provide a thorough review of various machine learning models and methods used in financial markets to identify and select stocks with high return potential. This study assesses several models, including regression, decision trees, random forests, neural networks, and support vector machines. It also analyzes methods for selecting effective features for predicting stock performance, encompassing fundamental, technical, and sentiment factors. Evaluating the performance of these models in real market conditions and comparing them with traditional methods reveal that machine learning can uncover complex patterns in financial data and facilitate better investment decisions. These techniques have vast applications in risk modeling, return forecasting, and portfolio management, helping to enhance investment performance. Zhao *et al.*, [24] investigate the use of Deep Reinforcement Learning (DRL) for stock portfolio selection, emphasizing asset correlation and presenting an innovative framework for investment management. This approach underlines the importance of accounting for correlations between assets and utilizes the temporal and spatial features of financial data to improve decision-making accuracy. Deep reinforcement learning is positioned as an effective tool for addressing multi-step decision-making problems that assist investors in maximizing investment returns and minimizing risk. A risk-cost-reward function is designed to include factors such as transaction costs and portfolio risk. Experiments have demonstrated that the proposed DRL framework yields higher cumulative returns compared to traditional strategies. Additionally, the Sharpe ratio and Maximum Drawdown have shown satisfactory results, signifying a favorable balance between return and risk. Extracting and amalgamating features of financial information has enhanced the model's efficiency and applicability.

According to the literature review, it can be concluded that most previous research has used static approaches for stock selection. This paper introduces a hybrid approach of Dynamic Adaptive Index (DAI) and stepwise elimination method to provide an approach that is compatible with market changes. Also, some studies focus on only one of the forecasting or selection stages. This paper fills this gap by presenting a comprehensive two-stage framework.

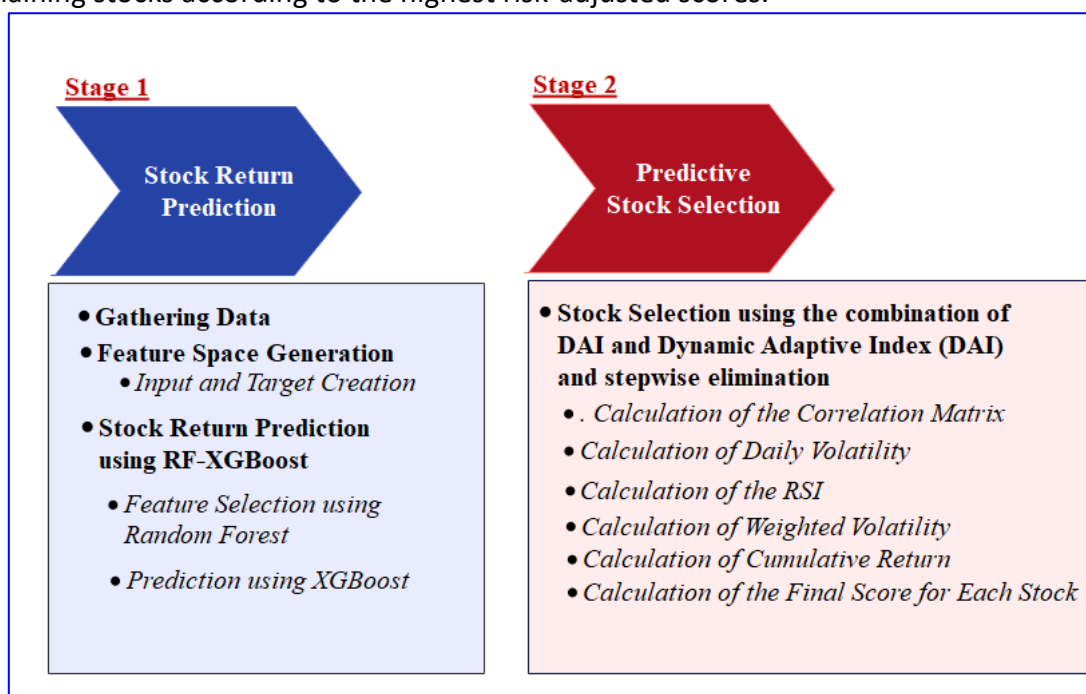
### **3. Research Methodology**

In this section the structure of proposed framework, data used in this study, data collection, Input and output creation, stock return prediction, stock selection approach are presented.

#### **3.1 Structure of Proposed framework**

The main purpose of the paper is to provide a predictive stock selection approach. The proposed method, as illustrated in Figure 1, consists of two main stages: stock return prediction and stock selection. The first stage focuses on stock return prediction. Initially, daily stock prices for 10 well-known stocks—including opening, high, low, and closing prices—were gathered for the period from January 4, 2010, to December 30, 2024, from Yahoo Finance. Subsequently, the input dataset and

target variable were created. Lagged return variables and technical indicators were regarded as input variables, with the daily return being considered the target variable. The extracted input variables served as inputs for the RF-CNN-XGBoost model to predict the daily return time series. The daily return of each of the 10 stocks is forecasted for the subsequent 10-day period. In the second step, optimal stock selection is addressed. This research employs a multi-stage hybrid approach for optimal stock selection. Initially, the forecasted return data of the 10 stocks over the 10-day period (the output of the first step) is analyzed using the Pearson correlation matrix and calculating the daily standard deviation. Then, a volatility-weighted index based on the proportion of positive returns (modeled from the concepts of the relative strength index) is developed, establishing an inverse relationship between the frequency of positive returns and the level of volatility. This index functions as a return-adjusted risk measure. The final stock score is calculated by dividing the cumulative return by the associated risk, representing the risk-adjusted return. In the screening phase, a stepwise elimination process is applied: first, the 20% of stocks with the highest volatility (identified as the riskiest) and the stocks with the lowest cumulative returns are removed. The top 5 are chosen from the remaining stocks according to the highest risk-adjusted scores.



**Fig. 1.** The Stages of Research Implementation

### 3.2 First stage: Stock Return Prediction

#### 3.2.1 Data Collection

In this research, we employed the Data-Base (DB) of the Financial Times Stock Exchange 100 Index (FTSE 100) to choose stocks for assessing the suggested approach. We gathered time series data for the daily closing prices of 10 stocks from the FTSE 100 spanning from January 1, 2010, to December 31, 2024, sourced from Yahoo Finance, treating them as experimental datasets

#### 3.2.2 Feature Space Generation

This study employs a comprehensive set of 32 variables derived from historical stock data to predict stock returns (Abolmakarem *et al.* [25]). These indicators can be broadly categorized as follows:

### 3.2.2.1 Logarithmic Price Ratios (v1-v18):

A foundational set of features is constructed from logarithmic ratios of closing, high, and low prices relative to open prices at various time lags (i, i-1, i-2, i-3, i-4). These ratios capture the relative price movements and volatility over different periods. Specifically:

#### Closing Price Ratios (v1-v4):

Represent the logarithmic change in closing prices over the previous four days, providing insight into recent price trends. Calculated as  $v1 = \ln\left(\frac{\text{Close Price}_i}{\text{Close Price}_{i-1}}\right)$  and similarly for v2-v4.

#### High Price Ratios (v5-v13):

Reflect the logarithmic relationship between the high price and the open price, both for the current day and with lags up to four days. These features can indicate potential upward price movements. For example,  $v5 = \ln\left(\frac{\text{High Price}_i}{\text{Open Price}_i}\right)$ .

#### Low Price Ratios (v14-v18):

Show the logarithmic ratio of low prices to open prices, signaling possible downward price movements, both on the current day and with lags up to four days. For instance,  $v14 = \ln\left(\frac{\text{Low Price}_i}{\text{Open Price}_i}\right)$ .

### 3.2.2.2 Classical Technical Indicators (v19-v32)

The feature set is augmented with a diverse range of well-established technical indicators, each designed to capture specific aspects of market dynamics.

#### Momentum and Strength:

Includes Relative Strength Index (RSI) (v19) and Momentum (v20) to gauge the speed and magnitude of price changes.

$$RSI = 100 - \frac{100}{1 + \left(\frac{\sum_{i=0}^{n-1} Up_{t-i}/n}{\sum_{i=0}^{n-1} Dw_{t-i}/n}\right)} \quad (1)$$

$$Momentum = C_t - C_{t-n} \quad (2)$$

$C_t$  is the closing price at time  $t$ ,  $Up_t$  means the upward price change,  $Dw_t$  means the downward price change at time  $t$ .

$$up_t = \max_t - \max_{t-1}, dw_t = \min_{t-1} - \min_t$$

if  $up_t > dw_t$  and  $up_t > 0$  then  $+DM = up_t$ . otherwise  $+DM = 0$ ;

if  $dw_t > up_t$  and  $dw_t > 0$  then  $-DM = dw_t$ . otherwise  $-DM = 0$ ;

$$+DI_t(n) = \frac{100 EMA_t(n,+DM)}{ATR_t(n)}, -DI_t(n) = \frac{100 EMA_t(n,-DM)}{ATR_t(n)}$$

#### Volatility:

Features True Range (TR) (v21) and Average True Range (ATR) (v22) to measure price volatility.

$$TR = \max\{\max_t, C_{t-1}\} - \min\{\min_t, C_{t-1}\} \quad (3)$$

$$ATR = \frac{1}{n} \sum_{i=1}^n TR_i \quad (4)$$

#### Trend Following:

Incorporates Parabolic SAR (v23) to identify potential trend reversals, as well as Simple Moving Average (SMA) (v24), Weighted Moving Average (WMA) (v25), and Hull Moving Average (HMA) (v26) to smooth price data and identify trends.

$$SAR_{n+1} = SAR_n + AF(EP - SAR_n) \quad (5)$$

$$SMA = \frac{C_t + C_{t-1} + \dots + C_{t-n}}{n} \quad (6)$$

$$WMA = \frac{((n) \times C_t + (n - 1) \times C_{t-1} + \dots + C_n)}{(n + (n - 1) + \dots + 1)} \quad (7)$$

$$HMA = WMA \left( 2 * WMA^{(n/2)} - WMA(n) \right) .sqrt(n) \quad (8)$$

The AF value indicates the factor of acceleration. EP represents the extreme position in a trend (the peak price achieved during an uptrend or the lowest price attained during a downtrend)

*Money Flow:*

Uses Money Flow Index (MFI) (v27) to assess buying and selling pressure.

$$MFI = 100 - [100/(1 + Money Ratio)] \quad (9)$$

Money Ratio = Positive Money Flow/Negative Money Flow, Money Flow  
 = Typical Price × Volume, Typical Price = (High + Low + Close)/3

*Overbought/Oversold Conditions:*

Employs Commodity Channel Index (CCI) (v28), William's %R (v29), %K (v30), and %D-slow (v31) to identify potential overbought or oversold conditions.

$$CCI = \frac{M_t - SM_t}{0.015D_t} \quad (10)$$

$$\text{William's \% R} = \frac{H_n - C_t}{H_n - L_n} \times 100 \quad (11)$$

$$\% K = \frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100 \quad (12)$$

$$\% D - \text{slow} = \frac{\sum_{i=0}^{n-1} \%D_{t-1}}{n} \quad (13)$$

$L_t$  is the low price at time  $t$ ,  $H_t$  is the High price at time  $t$ ,  $LL_t$  and  $HH_t$  mean lowest low and highest high in the last  $t$ ,

$$M_t: \frac{H_t + L_t + C_t}{3} . SM_t: \frac{(\sum_{i=1}^n M_{t-i+1})}{n} . D_t: \left( \sum_{i=1}^n |M_{t-i+1} - SM_t| \right)$$

*Directional Movement:*

Contains Average Directional Index (ADX) (v32) to measure the strength of a trend.

$$ADX = 100 \times \frac{EMA_t(n, abs(+DI - (-DI)))}{+DI + (-DI)} \quad (14)$$

$EMA_t(n, X)$  represents the exponential moving average of X for the previous n periods

### 3.2.3 Stock Price Prediction using RF -XGBoost

In this research, a multi-steps method based on Random Forest (RF) and XGBoost is employed to predict financial returns. In the first step, the Random Forest is used to identify influential features. Random Forest selects the variables that have the most impact on the target variable. In the second step, the XGBoost algorithm is utilized. As a robust machine learning algorithm, XGBoost effectively learns nonlinear relationships among features. This algorithm significantly improve the prediction accuracy through the use of decision trees and advanced optimization techniques. XGBoost is trained on the features selected by Random Forest and then employed to predict returns on the test data. The combination of these two methods significantly enhancing prediction accuracy. In the third step, the model is evaluated using performance metrics such as MSE, MAE, RMSE and R<sup>2</sup>. The results indicate that the combination of Random Forest and XGBoost can be used as a powerful tool for predicting financial returns.

### 3.3 Second Stage: Predictive Stock Selection

As mentioned, the second stage focuses on the optimal selection of stocks. This research employs the Dynamic Adaptive Index (DAI) method in combination with the stepwise elimination technique for stock selection. This method is explained in detail below.

#### 3.3.1 Calculation of the Correlation Matrix

The correlation matrix is computed to assess the relationships between the returns of different stocks. The correlation matrix for the predicted returns is calculated as Eq. (15).

$$corr = \begin{bmatrix} \rho_{11} & \cdots & \rho_{1n} \\ \vdots & \ddots & \vdots \\ \rho_{n1} & \cdots & \rho_{nn} \end{bmatrix} \quad (15)$$

where  $\rho_{ij}$  is the correlation between stock  $i$  and stock  $j$  and  $n$  is the number of stocks.

#### 3.3.2 Calculation of Daily Volatility (Standard Deviation)

Volatility, which is defined as the standard deviation of returns, measures the risk associated with each stock. The standard deviation for every stock is calculated using Eq. (16).

$$\sigma_i = \sqrt{\frac{1}{n} \sum_{j=1}^n (r_{ij} - \mu_i)^2} \quad (16)$$

where  $\sigma_i$  is the volatility (standard deviation) of stock  $i$ ,  $r_{ij}$  is the return of stock  $i$  for day  $j$ .  $\mu_i$  is the mean return of stock  $i$ .

#### 3.3.3 Calculation of the Relative Strength Index (RSI)

The RSI assesses the strength of a stock's price movements by comparing the magnitude of its gains to its losses over a defined period. Each stock's Relative Strength Index (RSI) is calculated as Eq. (17).

$$RSI_i = \frac{\text{Number of positive days}}{\text{Total number of days}} \times 100 \quad (17)$$

where  $RSI_i$  is the Relative Strength Index of stock  $i$ .

#### 3.3.4 Calculation of Weighted Volatility (Sigma Weighted)

Weighted volatility adjusts the standard deviation of returns by incorporating the RSI. The weighted volatility modifies the standard deviation of returns by including the RSI. The weighted volatility for each stock is determined as per Eq. (18).

$$\sigma_{w,i} = \sigma_i \times \left(1 - \frac{RSI_i}{100}\right) \quad (18)$$

where  $\sigma_{w,i}$  is the weighted volatility of stock  $i$ .

#### 3.3.5. Calculation of Cumulative Return

The cumulative return totals the returns of each stock over the specified period. The cumulative return for each stock is calculated using Eq. (19).

$$R_i = \sum_{j=1}^n r_{ij} \quad (19)$$

where  $R_i$  is the cumulative return of stock  $i$ .

### 3.3.6 Calculation of the Final Score for Each Stock

The final score for each stock is derived by dividing the cumulative return by the weighted volatility. The final score for each stock is calculated Eq. (20).

$$Score_i = \frac{R_i}{\sigma_{w,i} + \varepsilon} \quad (19)$$

where  $Score_i$  is the final score of stock  $i$ ,  $\varepsilon$  is a small value (e.g., 0.001) to prevent division by zero.

### 3.3.7 Stepwise Elimination Technique

The Stepwise Elimination Technique systematically reduces the number of stocks by removing those with undesirable characteristics:

Step 1: Eliminates the most volatile stocks to reduce overall portfolio risk.

Step 2: Removes the stock with the lowest cumulative return to ensure only high-performing stocks remain.

Step 3: Selects the top 5 stocks based on the final score, ensuring a balanced combination of performance and risk.

## 4. Empirical Results

This section presents the empirical results of the proposed method

### 4.1 Stock Price Prediction using RF-XGBoost

In this study, a hybrid approach combining Random Forest (RF) and XGBoost was employed to predict financial returns. In the initial stage, the Random Forest algorithm was applied to identify key features. By calculating the significance of each feature, those with the greatest impact on the target variable were selected. Subsequently, the XGBoost algorithm was utilized to develop the prediction model. As a powerful machine learning method, XGBoost effectively captures nonlinear relationships between features. This algorithm significantly enhanced prediction accuracy through the use of decision trees and advanced optimization techniques. XGBoost was trained on the features identified by Random Forest and then used to forecast returns on the test data. The parameter settings for each of the algorithms employed are presented in Table 1.

**Table1**

Parameter settings of applied algorithms

Algorithm	Parameter	Description	Tuned Value
Random Forest	random_state	Ensures reproducibility of results	42
	n_estimators	Number of trees in the forest	100
	criterion	Splitting criterion (mean squared error)	'squared_error'
XGBoost	n_estimators	Number of boosting trees	100
	max_depth	Maximum depth of a tree	6
	learning_rate	Step size shrinkage to prevent overfitting	0.1
	random_state	Ensures reproducibility of results	42
	objective	Loss function for regression (mean squared error)	'reg:squarederror'

The results of the feature selection procedure represented that, for most stocks, r1 (one-day return rate) has the greatest impact, with values close to 1. Following r1, features such as r5, TR, and r17 were ranked next, but their significance was considerably lower than that of r1. This indicates that although these features are somewhat effective in prediction, their impact is less than that of the r1. Technical indicators such ADX, RSI, CCI, SMA, WMA, and HMA, were identified as the least

influential variables. This means that technical indicators and longer-term averages have a diminishing effect on short-term forecasting when compared to recent price data. Overall, it can be concluded that for short-term stock price forecasting, it is very important to focus on recent price data, especially one-day returns, while technical indicators and long-term averages are less influential. The predicted performance of the proposed models, in terms of evaluation measures, applied to the 10 selected stocks, is displayed in Table 2.

**Table 2**

The predicted performance of the proposed models for each 10 stocks

Stock	MSE	MAE	RMSE	R <sup>2</sup> Score
AAL	7.056E-09	0.000061	0.000084	0.999962
WFC	2.1904E-08	0.000106	0.000148	0.999911
GSK	6.889E-09	0.000066	0.000083	0.999941
AZN	2.6523E-07	0.000209	0.000515	0.998891
TSCO	1.5132E-07	0.000241	0.000389	0.99965
GOLD	7.1824E-08	0.000149	0.000268	0.999758
BHP	1.9321E-08	0.000089	0.000139	0.999808
RY	5.29E-08	0.000098	0.00023	0.999467
SPXL	1.1236E-06	0.000453	0.00106	0.999081
AHT	4.5158E-07	0.000479	0.000672	0.999734

Table 3 presents the actual and predicted value of closing prices for the new 10 day.

**Table 3**

The actual and predicted value of returns for the new 10 days

		12/16/2024	12/17/2024	12/18/2024	12/19/2024	12/20/2024
AAL	Actual	-0.023641	0.012106	-0.014952	0.010322	0.014423
	Predicted	-0.023845	0.012104	-0.015002	0.010304	0.014353
AHT	Actual	0.00000	0.028015	-0.077014	-0.065469	0.061813
	Predicted	0.000009	0.028044	-0.076212	-0.064174	0.060601
AZN	Actual	-0.005257	0.014344	-0.037809	-0.003094	0.014122
	Predicted	-0.005271	0.014452	-0.039427	-0.003077	0.014
BHP	Actual	-0.010838	-0.002152	-0.031373	-0.009312	0.005313
	Predicted	-0.010828	-0.002126	-0.031755	-0.009352	0.005205
GOLD	Actual	-0.025626	-0.006728	-0.044951	-0.018053	0.015102
	Predicted	-0.025517	-0.006646	-0.045764	-0.018005	0.015047
GSK	Actual	-0.010898	0.019357	-0.015776	-0.007717	0.005085
	Predicted	-0.010927	0.019439	-0.015858	-0.007763	0.005095
RY	Actual	-0.000723	-0.005707	-0.031205	-0.001001	0.008353
	Predicted	-0.000763	-0.005679	-0.030482	-0.001015	0.008309
SPXL	Actual	0.012027	-0.013323	-0.088685	-0.002134	0.033082
	Predicted	0.01199	-0.013304	-0.085384	-0.002091	0.032886
TSCO	Actual	-0.031284	-0.004737	-0.036409	-0.004374	0.021096
	Predicted	-0.031547	-0.004694	-0.03707	-0.004418	0.020849
WFC	Actual	0.021863	-0.016255	-0.028951	0.001309	0.021641
	Predicted	0.021887	-0.016253	-0.028581	0.001389	0.021498

**Table 3**  
 Continued

		12/23/2024	12/24/2024	12/26/2024	12/27/2024	12/30/2024
AAL	Actual	0.021919	0.005797	0.000000	0.000000	0.015562
	Predicted	0.02184	0.005692	0.000012	0.000012	0.015617
AHT	Actual	0.003881	-0.001289	0.009032	-0.034527	-0.049007
	Predicted	0.00396	-0.001138	0.008874	-0.034235	-0.048242
AZN	Actual	0.019587	-0.004953	0.003318	-0.003909	-0.010414
	Predicted	0.019653	-0.004915	0.003271	-0.003855	-0.010404
BHP	Actual	0.004268	0.000202	-0.001416	-0.002837	-0.008738
	Predicted	0.004238	0.000065	-0.001441	-0.002734	-0.00877
GOLD	Actual	0.008409	0.007697	0.00191	-0.003812	-0.014031
	Predicted	0.008507	0.00768	0.001838	-0.003892	-0.014145
GSK	Actual	0.013691	-0.000881	0.002645	-0.001172	-0.012617
	Predicted	0.013832	-0.000838	0.002616	-0.001206	-0.012787
RY	Actual	0.000000	0.000166	0.002816	-0.001652	-0.002813
	Predicted	0.000013	0.000198	0.002864	-0.001677	-0.00283
SPXL	Actual	0.018225	0.031224	0.000055	-0.031536	-0.034767
	Predicted	0.018082	0.031153	0.000015	-0.032014	-0.034564
TSCO	Actual	-0.004266	0.0149	0.018352	-0.01676	-0.036107
	Predicted	-0.004255	0.014954	0.018267	-0.016798	-0.03707
WFC	Actual	0.002843	0.014885	0.002375	-0.009058	-0.009844
	Predicted	0.002741	0.014709	0.002363	-0.009159	-0.009894

#### 4.2 Stock Selection

In this section, the results of the stock selection stage will be presented. Using the predicted return data for 10 stocks over the next 10 days (output from the first stage), the stock selection process was optimized. To achieve this, the correlation matrix between the predicted returns was first calculated to examine the relationships among the stocks. Next, the daily volatility (standard deviation) for each stock was computed to assess risk. Then, by implementing the DAI (Dynamic Asset Integration) method, the RSI (Relative Strength Index) and adjusted volatility ( $\sigma_{w,i}$ ) were calculated for each stock. These indices served as criteria for evaluating the stocks' performance and risk. A combination of analytical techniques and gradual elimination was employed to select the best stocks. In the first step, the two stocks with the highest volatility were removed to lower the overall risk of the portfolio. In the second step, the stock with the lowest cumulative return (provided it was not eliminated in the first step) was also removed to filter out weaker stocks from the portfolio. Finally, in the third step, five stocks with the highest final score (based on cumulative return and adjusted volatility) were chosen from the remaining stocks. Table 4 displays the correlation matrix between the predicted returns of the 10 initial stocks.

**Table 4**

Correlation matrix between the predicted returns of 10 initial stocks

	AAL	AHT	AZN	BHP	GOLD	GSK	RY	SPXL	TSCO	WFC
AAL	1	-0.2473	-0.3512	-0.0561	0.3059	0.0261	0.2615	-0.3603	-0.3498	-0.1423
AHT	-0.2473	1	0.0478	-0.5283	0.2826	0.3092	-0.6814	0.1284	-0.0445	-0.0903
AZN	-0.3512	0.0478	1	0.5704	-0.786	-0.01	-0.4562	-0.1237	0.8998	0.7807
BHP	-0.0561	-0.5283	0.5704	1	-0.7643	0.2809	0.2821	-0.301	0.5939	0.6833
GOLD	0.3059	0.2826	-0.786	-0.7643	1	-0.1033	0.2285	0.3567	-0.9	-0.7477
GSK	0.0261	0.3092	-0.01	0.2809	-0.1033	1	0.021	-0.5745	0.1252	0.372
RY	0.2615	-0.6814	-0.4562	0.2821	0.2285	0.021	1	0.0589	-0.5076	-0.3159
SPXL	-0.3603	0.1284	-0.1237	-0.301	0.3567	-0.5745	0.0589	1	-0.3715	-0.6058
TSCO	-0.3498	-0.0445	0.8998	0.5939	-0.9	0.1252	-0.5076	-0.3715	1	0.8741
WFC	-0.1423	-0.0903	0.7807	0.6833	-0.7477	0.372	-0.3159	-0.6058	0.8741	1

Table 5 presents daily volatility (standard deviation), RSI (Relative Strength Index), weighted volatility ( $\sigma_{w,i}$ ), and cumulative return for each stock.

**Table 5**

Standard deviation, RSI, weighted volatility, and cumulative return for each stock

Stock	AAL	AHT	AZN	BHP	GOLD
Standard Deviation	0.013553	0.040595	0.015864	0.010092	0.017423
RSI	80	50	40	30	40
$\sigma_{w,i}$	0.002711	0.020298	0.009518	0.007065	0.010454
Cumulative Return	0.041087	-0.12251	-0.01557	-0.0575	-0.0809

**Table 5**

Continued

Stock	GSK	RY	SPXL	TSCO	WFC
Standard Deviation	0.010917	0.009771	0.034173	0.020997	0.015606
RSI	40	40	50	30	60
$\sigma_{w,i}$	0.00655	0.005862	0.017086	0.014698	0.006242
Cumulative Return	-0.0084	-0.03106	-0.07323	-0.08178	0.0007

Finally, the total and final score of each share has been calculated. The results are presented in Table 6.

**Table 6**

Calculated Final Score for each stock

Stock	Cumulative Return	Volatility	Final Score
AAL	0.041087	0.002710567	11.07297125
AHT	-0.122513	0.020297552	-5.752445073
AZN	-0.015573	0.009518182	-1.480579111
BHP	-0.057498	0.007064538	-7.129732701
GOLD	-0.080897	0.01045404	-7.062748399
GSK	-0.008397	0.00655005	-1.112178117
RY	-0.031062	0.005862321	-4.526456841
SPXL	-0.073231	0.0170864	-4.048954014
TSCO	-0.081782	0.014697616	-5.20983561
WFC	0.0007	0.006242242	0.096655158

As mentioned, in the first phase of gradual elimination, two stocks with the highest volatility were removed. According to Table 6, the stocks AHT and SPXL were eliminated due to their high volatility. In the second phase, TSCO was removed because of its lowest cumulative return. Finally, among the remaining stocks, five stocks with the highest final scores, as presented in Table 7, were selected.

**Table 7**  
Selected stocks with the highest final score

Stock	Final Score
AAL	11.07297125
WFC	0.096655158
GSK	-1.112178117
AZN	-1.480579111
RY	-4.526456841

Using this method, we were able to evaluate stocks based on cumulative return, volatility, and final score and optimally select the top stocks for investment.

## 5. Conclusion Remarks and Suggestion for Future Research

This paper introduces a hybrid approach for predictive stock selection that combines advanced machine learning algorithms and dynamic optimization techniques to enhance portfolio performance. The proposed method integrates Random Forest (RF) and XGBoost algorithms for feature selection and prediction, alongside Dynamic Adaptive Index (DAI) and gradual elimination techniques for optimal stock selection. In the first phase, the RF algorithm effectively identifies key features by calculating their importance, which simplifies the model and reduces noise. XGBoost is subsequently trained on these selected features to model complex nonlinear relationships among data points. The second phase employs a hybrid DAI and stepwise elimination method that dynamically adapts to changing market conditions by integrating various metrics, such as correlation, volatility, relative strength index (RSI), and cumulative return. This comprehensive evaluation framework enables more informed stock assessments. Additionally, the stepwise elimination technique systematically removes stocks with high volatility or low returns from the portfolio, ensuring that only stocks with strong performance and acceptable risk are retained. This two-step approach optimizes portfolio composition by effectively balancing risk and return.

The current study relied exclusively on historical price data and technical indicators. Future research could enhance the model's forecasting accuracy by incorporating macroeconomic factors such as interest rates, inflation, and GDP. This study focused on a short-term time frame. Future studies could investigate the sustainability of the model's performance over medium- and long-term time frames to assess its stability across different market conditions.

## Author Contributions

Conceptualization, F.A., S.A., and A.K.Y.; methodology, F.A., S.A., and A.K.Y.; software, S.A., and A.K.Y.; validation, S.A., and A.K.Y.; formal analysis, F.A., S.A., and A.K.Y.; writing—original draft preparation, F.A., S.A., and A.K.Y.; writing—review and editing, F.A., S.A., and A.K.Y.; supervision, S.A., and A.K.Y. All authors have read and agreed to the published version of the manuscript.

## Funding

This research received no external funding.

## 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.

## Data Availability Statement

The datasets generated during and/or analyzed during the current study is available from the corresponding author on reasonable request.

## Acknowledgement

This research was not funded by any grant research.

## References

- [1] Rather, A. M., Sastry, V. N., & Agarwal, A. (2017). Stock market prediction and Portfolio selection models: A survey. *OPSEARCH*, 54(3), 558–579. <https://doi.org/10.1007/s12597-016-0289-y>
- [2] Yang, F., Chen, Z., Li, J., & Tang, L. (2019). A novel hybrid stock selection method with stock prediction. *Applied Soft Computing*, 80, 820–831. <https://doi.org/10.1016/j.asoc.2019.03.028>
- [3] Yuan, X., Yuan, J., Jiang, T., & Ain, Q. U. (2020). Integrated Long-Term Stock Selection Models Based on Feature Selection and Machine Learning Algorithms for China Stock Market. *IEEE Access*, 8, 22672–22685. <https://doi.org/10.1109/ACCESS.2020.2969293>
- [4] Sezer, O. B., Ozbayoglu, M., & Dogdu, E. (2017). A Deep Neural-Network Based Stock Trading System Based on Evolutionary Optimized Technical Analysis Parameters. *Procedia Computer Science*, 114, 473–480. <https://doi.org/10.1016/j.procs.2017.09.031>
- [5] Sezer, O. B., Gudelek, M. U., & Ozbayoglu, A. M. (2020). Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied Soft Computing*, 90, 106181. <https://doi.org/10.1016/j.asoc.2020.106181>
- [6] Paiva, F. D., Cardoso, R. T. N., Hanaoka, G. P., & Duarte, W. M. (2019). Decision-making for financial trading: A fusion approach of machine learning and portfolio selection. *Expert Systems with Applications*, 115, 635–655. <https://doi.org/10.1016/j.eswa.2018.08.003>
- [7] Ma, B., Xue, Y., Lu, Y., & Chen, J. (2025). Stockformer: A price–volume factor stock selection model based on wavelet transform and multi-task self-attention networks. *Expert Systems with Applications*, 273, 126803. <https://doi.org/10.1016/j.eswa.2025.126803>
- [8] Alzaman, C. (2025). Optimizing portfolio selection through stock ranking and matching: A reinforcement learning approach. *Expert Systems with Applications*, 269, 126430. <https://doi.org/10.1016/j.eswa.2025.126430>
- [9] Sarkar, P., Khanapuri, V. B., & Tiwari, M. K. (2025). Integration of prediction and optimization for smart stock portfolio selection. *European Journal of Operational Research*, 321(1), 243–256. <https://doi.org/10.1016/j.ejor.2024.08.027>
- [10] Gao, Q., Zhou, X., Huang, L., Zhang, K., Liu, S., & Zhou, F. (2024). Relational Fusion-based Stock Selection with Neural Recursive Ordinary Differential Equation Networks. *Information Fusion*, 110, 102468. <https://doi.org/10.1016/j.inffus.2024.102468>
- [11] Lv, M., Wang, J., Wang, S., Gao, J., & Guo, H. (2024). Developing a hybrid system for stock selection and portfolio optimization with many-objective optimization based on deep learning and improved NSGA-III. *Information Sciences*, 670, 120549. <https://doi.org/10.1016/j.ins.2024.120549>
- [12] Vasantha Lakshmi, K., & Udaya Kumara, K. N. (2023). A novel randomized weighted fuzzy AHP by using modified normalization with the TOPSIS for optimal stock portfolio selection model integrated with an effective sensitive analysis. *Expert Systems with Applications*, 243, 122770. <https://doi.org/10.1016/j.eswa.2023.122770>
- [13] Alzaman, C. (2024). Deep learning in stock portfolio selection and predictions. *Expert Systems with Applications*, 237, 121404. <https://doi.org/10.1016/j.eswa.2023.121404>
- [14] Brito, I. (2023). A portfolio stock selection model based on expected utility, entropy and variance. *Expert Systems with Applications*, 213, 118896. <https://doi.org/10.1016/j.eswa.2022.118896>
- [15] Silva, N. F., de Andrade, L. P., da Silva, W. S., de Melo, M. K., & Tonelli, A. O. (2024). Portfolio optimization based on the pre-selection of stocks by the Support Vector Machine model. *Finance Research Letters*, 61, 105014. <https://doi.org/10.1016/j.frl.2024.105014>
- [16] Mitra Thakur, G. S., Bhattacharyya, R., & Sarkar (Mondal), S. (2018). Stock portfolio selection using Dempster–Shafer evidence theory. *Journal of King Saud University - Computer and Information Sciences*, 30\*(2), 223–235. <https://doi.org/10.1016/j.jksuci.2016.07.001>
- [17] Putra, Y. E., Saepudin, D., & Aditsania, A. (2021). Portfolio Selection of KOMPAS-100 Stocks Index Using B-Spline Based Clustering. *Procedia Computer Science*, 179, 375–382. <https://doi.org/10.1016/j.procs.2021.01.019>

- [18] Chen, B., Zhong, J., & Chen, Y. (2020). A hybrid approach for portfolio selection with higher-order moments: Empirical evidence from Shanghai Stock Exchange. *Expert Systems with Applications*, 145, 113104. <https://doi.org/10.1016/j.eswa.2019.113104>
- [19] Yan, Y., Zhang, C., Li, X., & Zhang, B. (2024). A framework for stock selection via concept-oriented attention representation in hypergraph neural network. *\*Knowledge-Based Systems*, 284\*, 111326. <https://doi.org/10.1016/j.knosys.2023.111326>
- [20] Caparrini, A., Arroyo, J., & Escayola Mansilla, J. (2024). S&P 500 stock selection using machine learning classifiers: A look into the changing role of factors. *Research in International Business and Finance*, 70, 102336. <https://doi.org/10.1016/j.ribaf.2024.102336>
- [21] Asawa, Y. S. (2022). Modern Machine Learning Solutions for Portfolio Selection. *IEEE Engineering Management Review*, 50(1), 94–111. <https://doi.org/10.1109/EMR.2021.3131158>
- [22] Padhi, D. K., Padhy, N., Bhoi, A. K., Shafi, J., & Yesuf, S. H. (2022). An Intelligent Fusion Model with Portfolio Selection and Machine Learning for Stock Market Prediction. *Computational Intelligence and Neuroscience*, 2022, 1–18. <https://doi.org/10.1155/2022/7588303>
- [23] Rasekhschaffe, K. C., & Jones, R. C. (2019). Machine Learning for Stock Selection. *Financial Analysts Journal*, 75(3), 70–88. <https://doi.org/10.1080/0015198X.2019.1596678>
- [24] Zhao, T., Ma, X., Li, X., & Zhang, C. (2023). Asset correlation based deep reinforcement learning for the portfolio selection. *Expert Systems with Applications*, 221, 119707. <https://doi.org/10.1016/j.eswa.2023.119707>
- [25] Abolmakarem, S., Abdi, F., Khalili-Damghani, K., & Didekhani, H. (2024). A multi-stage machine learning approach for stock price prediction: Engineered and derivative indices. *Intelligent Systems with Applications*, 24, 200449. <https://doi.org/10.1016/j.iswa.2024.200449>