Journal of Electrical and Computer Engineering Innovations
(JECEI)
Journal homepage: http://www.jecei.sru.ac.ir

Research paper

# Application of Harris Hawks Optimization Algorithm and APSO-CLUSTERING in Predicting the Stock Market

*I. Behravan, S.M. Razavi**

*Department of Electrical Engineering, University of Birjand, Birjand, Iran.*

**Article Info**

**Article History:**
Received 01 December 2021
Reviewed 18 January 2022
Revised 10 February 2022
Accepted 11 April 2022

**Keywords:**
Tehran Stock market
Harris Hawks Optimization (HHO)
Support Vector Regression (SVR)
APSO-Clustering
Metaheuristics

*Corresponding Author's Email Address:
*smrazavi13550@gmail.com*

**Abstract**

**Background and Objectives:** Stock markets have a key role in the economic situation of the countries. Thus one of the major methods of flourishing the economy can be getting people to invest their money in the stock market. For this purpose, reducing the risk of investment can persuade people to trust the market and invest. Hence, Productive tools for predicting the future of the stock market have an undeniable effect on investors and traders' profit.

**Methods:** In this research, a two-stage method has been introduced to predict the next week's index value of the market, and the Tehran Stock Exchange Market has been selected as a case study. In the first stage of the proposed method, a novel clustering method has been used to divide the data points of the training dataset into different groups and in the second phase for each cluster's data, a hybrid regression method (HHO-SVR) has been trained to detect the patterns hidden in each group. For unknown samples, after determining their cluster, the corresponding trained regression model estimates the target value. In the hybrid regression method, HHO is hired to select the best feature subset and also to tune the parameters of SVR.

**Results:** The experimental results show the high accuracy of the proposed method in predicting the market index value of the next week. Also, the comparisons made with other metaheuristics indicate the superiority of HHO over other metaheuristics in solving such a hard and complex optimization problem. Using the historical information of the last 20 days, our method has achieved 99% accuracy in predicting the market index of the next 7 days while PSO, MVO, GSA, IPO, linear regression and fine-tuned SVR has achieved 67%, 98%, 38%, 4%, 5.6% and 98 % accuracy respectively.

**Conclusion:** in this research we have tried to forecast the market index of the next $m$ (from 1 to 7) days using the historical data of the past $n$ (from 10 to 100) days. The experiments showed that increasing the number of days ($n$), used to create the dataset, will not necessarily improve the performance of the method.

©2022 JECEI. All rights reserved.

## Introduction

Stock markets have key roles in the economic situation of the countries. In order to keep or increase capital, traders are buying and selling different companies' shares. Also, the companies can increase their fund by selling their shares in the stock market. So, investing in the stock market can result in growing the economic situation and, as a result of that, expanding the industry

section. Accordingly, persuading people to invest their money in the stock market is the first step to flourish the economy. On the other hand, it is necessary to reduce the risk of investment. Since the stock market can be affected by several internal and external factors, promising investment relies heavily on dependable prediction methods.

Thus, forecasting the stock markets has become one of the hot topics among traders and researchers in the last years. Numerous researches have been conducted in recent years based on the theory of the repetitive nature of stock market behavior. In these researches, several mathematical methods have been developed to forecast the future price of a stock or to forecast the future of a stock market index. Due to the nonlinearity, complexity, and noisy time-series data of the stock market, mathematical approaches are not reliable for prediction. but on the other hand, Machine learning methods, such as Support Vector Machine (SVM) and Artificial Neural Networks (ANN), have proven themselves as powerful and reliable forecasting methods. Metaheuristics are strong optimization algorithms that have gained much attention in the last years. These algorithms have been used frequently to solve complex optimization problems such as [1]-[6].

Many real-world problems in machine learning and artificial intelligence are hard to be tackled using conventional mathematical approaches such as conjugate gradient, sequential quadratic programming, fast steepest, and quasi-Newton methods [7], [8] due to the continuous, discrete, constrained, or unconstrained nature.

Thus these methods are not efficient in solving many large-scale real-world multimodal, non-continuous, and non-differentiable problems. Accordingly, metaheuristic algorithms have been invented and utilized to tackle these kinds of problems.

These algorithms have become very popular among researchers because of their simplicity, effectiveness, and ease of implementation process.

In this research, a two-stage prediction method is introduced to forecast the future of the Tehran Stock Exchange market index value.

In the first stage, a novel automatic clustering algorithm, called APSO-Clustering, is utilized to extract different clusters of the data points and in the second stage a hybrid regression method, combination of Harris Hawks Optimization algorithm (HHO) and Support Vector Regression (SVR), is used for each cluster's data points to detect the hidden patterns of them. Numerous experiments have been conducted to evaluate the efficiency of the method in predicting the stock market index value in the next $m$ (1 to 7) days using the historical information of the past $n$ (10 to 100) days and technical indicators.

The results of the experiments showed the effectiveness and high potential of the proposed method in forecasting the future. The main contributions of our work are:

1. Using a novel automatic clustering algorithm to divide the dataset in to smaller clusters of data to improve the performance of the model.
2. Using Harris Hawks optimization algorithm to tune the parameters of SVR and feature selection simultaneously.
3. Investigating the impact of historical information on the performance of the model.
4. Predicting the market index value for the next seven days.

There are two important issues should be mentioned as the research limitations. First, the dollar exchange rate has impact on the market but it is neglected in this research due to the lack of a structured dataset containing the daily price of dollar.

Second, some important economical indexes, such as inflation rate, have intense effect on the market index, but unfortunately, there is no structured and useful dataset containing these kinds of information for each working day. The paper is organized in the following manner:

- **Background**: a complete review on related works
- **Data preparation**: the procedure of preparing training data.
- **Proposed method:** a complete explanation about the prediction method.
- **Experimental results.**
- **Conclusion**.

## Background

In 2016 Usmani et al. studied the performance of different machine learning techniques on predicting the market performance of the Karachi Stock Exchange (KSE) [9]. They have trained the classifiers on a dataset containing different attributes such as gold and oil rates, political news, historical data of the market, etc. for a binary classification problem (positive or negative market index).

Their results indicate that machine learning techniques have a great capability in predicting the stock market. On the other hand, investigating the effect of different factors on the stock market, they proved that Petrol price is the most related factor while the foreign

exchange rate does not affect the performance of the stock market.

In 2017 Pyo et al. have investigated the predictability of machine learning techniques to predict the trends of the Korean stock market index (KOSPI 200) [10]. They have analyzed the performance of three nonparametric machine learning models: artificial neural network, support vector machine with polynomial, and radial basis function kernels. Their experiments revealed that the prediction of the KOSPI 200 using technical indicators (as inputs of the machine learning models) does not result in a good performance for market investments. On the other hand, they show that the google trend is not a suitable input factor in predicting the KOSPI 200 index prices.

In 2018 Senapati and his colleagues presented a hybrid method called, PSO-ANN, to predict the open price of a stock for 1 day ahead [11]. In this method, PSO is utilized to tune the weights of the Adaline neural network.

The time-series data of the Bombay stock market is used to evaluate the performance of the proposed method.

According to the reported results, PSO-ANN has shown better performance than ANN which indicates the efficiency of PSO in training ANN.

In 2018 Hu et al. introduced a method based on an improved sine cosine optimization algorithm (ISCA) and backpropagation neural network (BPNN) to predict the direction of stock markets (ISCA-BPNN) [12]. They have used ISCA to find the best possible values of the weights and biases of the neural network for maximum accuracy. In fact, ISCA is utilized to train the BPNN to predict the opening price of the next day with maximum accuracy. Evaluating the performance of the proposed method on "S&P 500" and "Dow Jones" datasets, they have demonstrated the superiority of the ISCA over GWO, WOA, and PSO.

In 2019 a new approach based on deep neural networks is introduced by Pang and his teammates [13]. In this project, two types of deep learning methods are used to forecast the Shanghai A-shares composite index: 1- LSTM with embedded hidden layer. 2- LSTM with automatic encoder.

Furthermore, in this research, a new concept, called "Stock Vector" is introduced. In fact, the input is not a single index or a single stock index, but multi-high dimensional historical data.

Gozalpour and Teshnehlab have proposed a stock price prediction method using deep neural networks in 2019 [14]. In their method dimensional reduction algorithms (PCA and autoencoder) are used to map the data points into a new feature space.

Besides, their method is designed to predict the close price of the next day using the stock price information (open price, lowest price, highest price and volume transaction) of the past 30 days. The method is tested on three NASDAQ symbols.

Ghanbari and Arian have introduced a hybrid regression method, briefly called BOA-SVR, in 2019 [15]. In this method BOA is used for parameter tuning of SVR. Also, phase space reconstruction method is used for data preparation.

In 2020 Vijh et al. have used ANN and Random Forest for the prediction of the close price [16]. The methods are trained using 6 new features extracted from the historical close prices of different stocks including 1- stock High minus Low price. 2- stock Close minus Open price. 3- Stock price's seven days' moving average. 4- Stock price's fourteen days' moving average. 5- stock price's twenty-one day's moving average. 6- stock price's standard deviation for the past seven days. After testing the methods on different companies of the NASDAQ stock market, the results showed the superiority of ANN over Random Forest.

Ecer and his colleagues have introduced a hybrid stock index forecasting method in 2020 [17]. In their research, they have utilized evolutionary algorithms to train MLP to estimate the direction of the Borsa Istanbul (BIST) index using 9 technical indicators. Their experiments have shown that using $Tanh(x)$ as the output function of MLP results in better accuracy in compare to Gaussian function.

In a research project, conducted by Nabipour and his teammates in 2020, the performance of nine machine learning methods (Decision tree, Random forest, Adaptive Boosting (Adaboost), eXtreme Gradient Boosting (XGboost), Support Vector Classifier (SVC), Naïve Bayes, K-Nearest Neighbors (KNN), Logistic Regression and Artificial Neural Network (ANN)) and two powerful deep learning models (Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM)) are compared over the stock data of four different groups of stocks in Tehran stock exchange market [18]. They have used ten technical indicators as input variables for the learning models in two ways: continuous and binary. Their results showed the superiority of deep learning models over conventional machine learning models in both cases.

In a paper published recently, Awan et al. have used several machine learning methods for stock price prediction including linear regression, generalized linear regression, random forest, decision tree, naive Bayes

and logistic regression [19]. They have used the historical data of the 15 famous companies and also the data in news, twitter, blogs and etc. to train the predictive models. Based on their experiments, linear regression, random forest and decision tree have shown the best performance in predicting the close value of the next day.

Kofi and his colleagues have tried to address the problem of stock price prediction in a different perspective [20]. They have proposed a fusion framework, based on convolutional neural network and a Long-Short term memory, to create a structured dataset by fusion of different heterogeneous datasets. To evaluate their framework, they have created a dataset of Ghana stock exchange market using different type of datasets ad trained a CNN on the created dataset. Their prediction accuracy showed that their framework has a positive effect on the final accuracy of the prediction model.

Tuarob et al. have proposed an end-to-end framework for stock market prediction [21]. Their framework, called DAViS, has different capabilities including data collection, analyzation and visualization. This framework can process the related heterogeneous stock data.

Also, it uses ensemble learning to predict the close price of a stock for one day ahead. Their simulation results, showed an improvement in the accuracy in compare to other baseline methods.

In [22] Muhammad Ali and his colleagues, have used the resilient back-propagation neural network to predict the direction movement of the stock market index. They have tested their method on KSE-100 index, KOSPI index, Nikkei 225 index and SZSE composite index. Comparing the results to SVM, showed the superiority of the neural network over SVM.

In most of the researches conducted recently, the researchers have used deep learning models to uncover the patterns hidden in the train data since the stock markets data contain a huge amount of information. Actually, analyzing the massive time-series data of the stock markets requires powerful learning models. Although deep learning models are strong methods that can be used to build accurate prediction models, the most important issue is the high cost of using these methods. In other words, to use deep learning methods, a powerful processing machine is needed. Another group of researchers has tried hybrid learning methods for this purpose.

Usually, in these methods, a metaheuristic algorithm is utilized to train a machine learning algorithm on the data, more effectively. Despite the effectiveness of hybrid learning methods, massive and high-dimensional datasets in addition to the iterative nature of metaheuristics result in a time-consuming training phase and decreasing the efficiency of the learning method in uncovering the hidden patterns of the data. So in this research first, the huge and high-dimensional training dataset is divided into smaller datasets based on their similarity using a novel automatic clustering algorithm.

In the next step, instead of training a single machine learning method on the whole massive data, an independent regression method in trained for each group of dataset individually.

Training different regression methods on the smaller datasets, containing similar data points, has had a key role in the performance of the prediction method. On other hand in the second phase, the metaheuristic algorithm (HHO) is used to find the best feature subset in addition to the parameters of the machine learning method (SVR).

Removing redundant features and using informative ones to train the machine learning method, improve the final accuracy of the prediction model.

Briefly, the main difference between our method and other methods, is that we have tried to reduce the amount of data and dimensions in a heuristic way to improve the performance. The experimental results show the effect of clustering and feature selection.

## Data Preparation

In this research, the historical data of the past $n$ days and four technical indicators including Bollinger bands (upper band, the middle band, and the lower band) and RSI are used to predict the index of the Tehran stock market in the next $m$ days. For example, for $n = 20$ and $m = 1$, the historical data of the past 20 days (close, open, high, low and volume) besides four technical indicators are used to predict the market index in the next day. In this case, the first data point of the dataset is a vector with 104 items including the historical information of the first 20 working days and the corresponding technical indicators.

The target value is the close value of the 21$^{st}$ day. Similarly, the second object of the dataset contains the historical data and technical indicators from the second to the 21$^{st}$ working day and the target is the stock market index value in the 22$^{nd}$ working day. In fact, the main goal of this research is to predict the stock market index in the next 7 days using a two-stage prediction method. In this way, we have investigated the effect of window size ($n$) in the performance of the stock market index predicting method. The method is completely explained in the next section.

## Proposed Method

The proposed method includes two stages: 1- clustering. 2- regression. In the first stage, APSO-Clustering [23]-[25] is utilized to cluster the training dataset. This novel automatic clustering algorithm, which can detect the number of clusters in addition to the centroids, divides the whole training dataset into different clusters. In the second stage, for each detected cluster a regression method, which is a combination of Harris Hawks Optimization algorithm (HHO) and Support Vector Regression, is hired to uncover the hidden patterns of each cluster.

In this hybrid regression method, briefly called HHO-SVR, HHO is utilized for feature selection and parameter tuning of SVM. Actually, in the training phase, HHO searches the solution space to find the best subset of features and the optimal value for the SVM's parameter. To estimate the target value of a test sample, after distinguishing its cluster, the corresponding trained regression model determines the target value of the test sample.

### A. APSO-CLUSTERING

APSO-Clustering, designed based on Particle Swarm Optimization algorithm, can detect the proper number of clusters in addition to the position of centroids. This clustering method works in two phases.

Detecting the number of cluster is the main goal of the first phase while, finding the exact position of the centroids is the main goal of the second phase. Thus the main superiority of APSO-Clustering over traditional clustering methods such as K-means and fuzzy C-means is its high capability in detecting the number of clusters. This capability is more valuable when dealing with big datasets. In both phases, PSO-Clustering, a non-automatic clustering method, is used. In this clustering method, Particle Swarm Optimization algorithm is hired to find $k$ centroids, while $k$ should be predetermined by the user.

In fact, $k$ is the input of PSO-Clustering. In the first phase of APSO-Clustering, PSO-Clustering is run several times sequentially with different values of $k$ to detect the best-fitted number of clusters. Each time, the best solution found by PSO is compared with the previously found solutions to distinguish the best value of $k$. In each step of the first phase, the number of population and iteration numbers are set to 5 and 150 respectively, and also Calinski-Harabasz index is used for fitness evaluation. In the second phase, again PSO-Clustering finds out the exact position of $k$ centroids. In this phase, to explore and exploit the search space completely, the iteration number is set to 600.

The pseudo-code of APSO-Clustering is shown in Fig. 1. In the next sections, HHO and HHO-SVR are described respectively.

```
Stage 1:
Inputs: number of sequences, length of sequence, number of population
For i=1 to number of sequences do:
        1.k = Generate a random integer number
        2. beginning population = Generate a random population
        3.Current state.fitness = inf
        4. Current state.k=Ø
        for j=1 to length of sequence do:
            if (j==1) do:
                current state.k=k
            end
            best fitness = PSO-Clustering(k , beginning population)
            if best fitness < current state.fitness
                current state.fitness=best fitness
                curent state.k=k
            end
            k = current state.k ± ε
        end
Bestk[i] = current state.k
End
Output: find the best solution which has the best fitness amount, its corresponding k and
corresponding beginning population
Stage 2:
Final output = PSO-Clustering(best k, beginning population)
```

Fig. 1: the pseudo-code of APSO-Clustering [23].

### B. Harris Hawks Optimization algorithm (HHO)

This optimization algorithm is invented by the inspiration of Harris' Hawks hunting mechanism. The Harris' Hawk is a well-known bird of prey that survives in somewhat steady groups found in the southern half of Arizona, USA [26].

These birds are known as truly cooperative predators in the raptor realm. The main tactic of Harris' Hawks to capture prey is "Surprise pounce" which is also known as the "seven kills" strategy. In this strategy, several Hawks try to cooperatively attack from different directions and converge on a detected escaping rabbit at the same time. HHO algorithm is created by mimicking the behavior of the Harris Hawks in hunting and also the behavior of the prey (rabbit) in escaping mechanisms. In fact, in this algorithm, the Hawks are the search agents and the prey is the optimum solution supposed to be found (hunted) by the search agents (Harris Hawks). Generally, HHO consists of two phases: 1- exploration. 2- exploitation. In each phase, the search agents move in the solution space using a specified criterion [27].

**Exploration phase**: In the exploration phase, the search agents try to discover different areas of the solution space which is also a common strategy among the Harris Hawks in nature. This strategy (exploration) is modeled in HHO by the following equation:

$$X(t+1) = \begin{cases} X_{rand}(t) - r_1 \left| X_{rand}(t) - 2r_2 X(t) \right| & q \geq 0.5 \\ \left(X_{rabbit}(t) - X_m(t)\right) - r_3\left(LB + r_4(UB - LB)\right) & q < 0.5 \end{cases} \quad (1)$$

According to (1), the search agents explore the solution space using two strategies:

1- Perching based on the position of the search agents and the position of the rabbit (best search agent) for the condition of $q < 0.5$.

2- Perching on random locations inside a specified range for the condition of $q \geq 0.5$.

In this equation, $X(t+1)$ indicates the location of the search agents in the next iteration, $X_{rabbit}(t)$ is the position of the best solution found from the beginning of the optimization process, $X(t)$ is the current position of the search agents, $r_1$, $r_2$, $r_3$, $r_4$ and $q$ are random numbers inside $(0.1)$, $LB$ and $UB$ are lower and upper bounds of the variables, $X_{rand}(t)$ is a randomly selected search agent and $X_m(t)$ is the average position of the search agents.

**Exploitation phase:** In the exploitation phase, the Hawks perform the "surprise pounce" strategy to catch the rabbit. On the other hand, the rabbit also tries to escape from the dangerous situation. Thus based on the probability of escaping, which is defined by a random number $(r)$ and the energy of the rabbit, shown in (2), four strategies are defined for the exploitation phase.

$$E = 2E_0\left(1 - \frac{t}{T}\right) \tag{2}$$

Soft besiege: If $r \geq 0.5$ and $|E| \geq 0.5$ the rabbit has enough energy to escape.

The following equations show the movement of search agents in this situation:

$$X(t+1) = \Delta X(t) - E|JX_{rabbit}(t) - X( \tag{3}$$

$$\Delta X(t) = X_{rabbit}(t) - X(t) \tag{4}$$

In these equations, $\Delta X(t)$ presents the difference between the position of the rabbit and the current position. $r_5$ is a random number and $J = 2(1 - r_5)$ is the random jump strength of the rabbit in the escaping procedure.

Hard besiege: In this situation ($r \geq 0.5$ , $|E| < 0.5$) the rabbit is so exhausted and it has low escaping energy. Hence the position of the search agents is updated through the following equation:

$$X(t+1) = X_{rabbit}(t) - E|\Delta X(t)| \tag{5}$$

Soft besiege with progressive rapid dives: In this case, the prey has enough energy to escape ($|E| > 0.5$) but a soft besiege in constructed by the Hawks ($r < 0.5$). In this situation to simulate the real zigzag deceptive movements of the prey (especially rabbits) and rapid dives of Hawks around the escaping prey, the Levy Flight concept [28], [29] is utilized in the HHO algorithm. Based on this strategy the position of the search agents is updated using the following equations:

$$X(t+1) = \begin{cases} Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X(t)| & if\ F(Y) < F(X(t)) \\ Z = Y + S \times LF(D) & if\ F(Z) < F(X(t)) \end{cases} \tag{6}$$

According to this equation, it is supposed that the Hawks can evaluate their next possible move and then decide to choose the better one. In other words, in each iteration, the better position ($Y$ or $Z$) is selected as the next position of the search agent.

In this equation, $D$ is the dimension of the problem, $S$ is a random vector by size $1 \times D$ and $LF$ is the Levy Flight function.

Hard besiege with progressive rapid dives: When $|E| < 0.5$ and $r < 0.5$ the prey is exhausted and also a hard besiege is constructed by the Hawks. The following equations indicate how search agents update their position in this circumstance:

$$X(t+1) = \begin{cases} Y = X_{rabbit}(t) - E|JX_{rabbit}(t) - X_m(t)| & if\ F(Y) < F(X(t)) \\ Z = Y + S \times LF(D) & if\ F(Z) < F(X(t)) \end{cases} \tag{7}$$

The pseudocode of HHO is shown in Fig. 2.

**Inputs**: The population size N and maximum number of iterations T
**Outputs**: The location of rabbit and its fitness value
Initialize the random population Xi(i = 1, 2, . . . , N)
**while** (stopping condition is not met) **do**
  Calculate the fitness values of hawks
  Set *Xrabbit* as the location of rabbit (best location)
  **for** (each hawk (Xi)) **do**
    Update the initial energy E0 and jump strength J
    Update the E using Eq. (2)
    **if** (|E|≥ 1) **then**
      Update the location vector using Eq. (1)
    **if** (|E|< 1) **then**
      **if** (r ≥0.5 and |E|≥ 0.5 ) **then**
        Update the location vector using Eq. (4)
      **else if** (r ≥0.5 and |E|< 0.5 ) **then**
        Update the location vector using Eq. (6)
      **else if** (r < 0.5 and |E|≥ 0.5 ) **then**
        Update the location vector using Eq. (7)
      **else if** (r < 0.5 and |E|< 0.5 ) **then**
        Update the location vector using Eq. (9)
**Return** *Xrabbit*

Fig. 2: the pseudocode of HHO [27].

### C. HHO-SVR

Support vector regression (SVR) is the developed version of the support vector machine classifier (SVM) which is suitable for the regression problems [30], [31]. Considering the smallest risk minimization principle in high-dimensional feature space, this well-known regression method finds the best regression hyperplane. This non-linear method maps the data points from

vector space to high-dimensional feature space using a kernel function to facilitate the process of distinguishing different objects [32]. Several kernel functions have been introduced up to now. The Gaussian function is one of the most popular functions which has been used frequently in different researches [33], [34]. This function maps the data points into feature space using the following equation:

$$K(x_i.x_j) = exp\left(-\gamma\|x_i - x_j\|^2\right) \quad (8)$$

The amount of $\gamma$ (Gaussian kernel's parameter. See (10)) has a significant effect on the performance of SVR. In Support Vector Regression machine, the user should select a kernel function and set the kernel parameter in order to achieve better generalization performance.

One of the most important parameters is the kernel parameter which implicitly defines the structure of the high dimensional feature space where the maximal margin hyperplane is found.

Too rich a feature space would cause the system to overfit the data.

The Gaussian kernel function is the most common used kernel function. Therefore, its parameter, $\gamma$, needs to be determined before the SVR is trained. It has been proved that $\gamma$, in Gaussian kernel function, dramatically affects the generalization performance of SVR. When $\gamma$ is very small, all the training data will be regarded as support vector, and therefore they can be classified correctly.

However, for any unseen data, the SVM may not give right distinction due to "over-fitting" training. On the other hand, when $\gamma$ is very large, all the training data are regarded as one point in feature space, the SVM cannot recognize any unseen data due to "under-fitting" training. Obviously, these two extreme situations should be avoided. Deeper analysis on this important topic is provided in [35], [36].

So finding the optimal value of $\gamma$ is an important task that is done by HHO in this research. In other words, removing redundant features and building a regression model based on informative features usually results in better prediction performance.

So HHO should search the solution space to find the best feature subset and the best value of $\gamma$ simultaneosly. Thus, each search agent contains $f$ cells for feature selection ($f = number\ of\ features$), which are encoded binary (1 for selecting the feature and 0 for removing the feature), and one cell for the value of $\gamma$. For fitness evaluation Mean Squared Error (*MSE*) function is used.

## Results and Discussion

The performance of the proposed method is evaluated on the historical data of the Tehran Stock Exchange market index from 6/12/2008 to 1/11/2020. To evaluate the performance of the method after training on 70% of the dataset, the following evaluation metrics have been calculated on the rest of the data:

$$MSE = \frac{1}{n}(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2) \quad (9)$$

$$RMSE = \sqrt{\frac{1}{n}(\sum_{i=1}^{n}(y_i - \hat{y}_i)^2)} \quad (10)$$

$$MAE = \frac{1}{n}(\sum_{i=1}^{n}(y_i - \hat{y}_i)) \quad (11)$$

$$R^2 = 1 - (\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \mu)^2}) \quad (12)$$

The first three metrics (mean squared error (*MSE*), root mean squared error (*RMSE*), and mean absolute error (*MAE*)) show the error rate. In other words, these metrics show the difference between the actual and estimated target values. On the other hand, $R^2$ shows the similarity.

Bigger value of $R^2$ means better performance of the method.

In the next subsections, the values of performance metrics achieved for different amounts of $n$ (window size) and $m$ (target day) are presented. The simulation parameters and their assigned values are shown in Table 1.

Table 1: Simulation parameters

| Parameter | Value |
|---|---|
| PSO_Iter_1 | 150 |
| PSO_Pop_1 | 5 |
| PSO_Iter_2 | 600 |
| PSO_Pop_2 | 5 |
| HHO_Iter | 50 |
| HHO_Pop | 15 |

The description of these parameters are as follows:

- *PSO_Iter_1:* iteration number of PSO, in PSO_Clustering in the first stage of APSO-Clustering.
- *PSO_Pop_1:* population number of PSO_Clustering in the first stage of APSO-Clustering
- *PSO_Iter_2:* the iteration number of PSO_Clustering in the second stage of APSO-Clustering.
- *PSO_Pop_2:* population number of PSO_Clustering in the second stage of APSO-Clustering.
- *HHO_Iter:* iteration number of HHO in the second phase of the proposed method.
- *HHO_Pop:* population number of HHO in the second stage of the proposed method.

*A. Predicting one day ahead (m=1)*

In this case, the target day is tomorrow. In other words, the main goal is to predict the stock market index value of tomorrow (close value) using the last $n$ days. The evaluation metrics of the method for different values of $n$ are presented in Table 2.

According to this table, the proposed method gives the best performance when $n = 20$. This means that, to predict the index value of the next day, considering the historical information of the last 20 days is enough. The details of the results of this experiment ($n = 20$) are presented in Table 3.

This table indicates that APSO-Clustering has detected 2 clusters.

In the second step, two hybrid regression methods are trained on each cluster's data. According to this table, HHO has detected 21 and (only) 3 features for the first and second clusters respectively.

Among these features, none of the technical indicators (RSI, upper Bollinger band, lower Bollinger band, and mid-Bollinger band) is selected.

Table 2: Performance evaluation of the proposed method for m=1

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | **$1.55\times10^{-4}$** | $4.07\times10^{-4}$ | $1.907\times10^{-4}$ | $2.97\times10^{-4}$ | $4.51\times10^{-4}$ |
| RMSE | **0.0125** | 0.0202 | 0.0138 | 0.017 | 0.0212 |
| MAE | **0.0051** | 0.0069 | 0.0052 | 0.0066 | 0.0084 |
| $R^2$ | **0.9929** | 0.9816 | 0.99 | 0.986 | 0.9809 |

Table 3: Details of the achieved results for m=1 and n=20

| | Number of selected features | $\gamma$ | Selected technical indicators |
|---|---|---|---|
| HHO-SVR-1 | 21 | 1 | none |
| HHO-SVR-2 | 3 | 1 | none |

In Fig. 3, the curves of real and predicted index values are demonstrated for $n = 20$. According to Fig. 3, the predicted values are very close to the real values which shows the fabulous performance of the presented method in predicting the market index value of 1 day ahead.

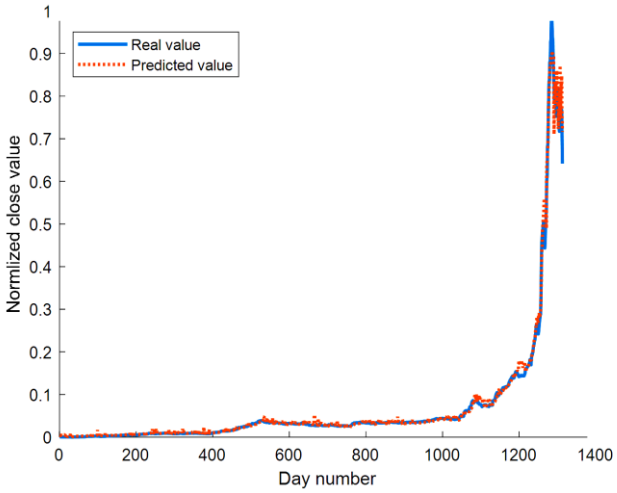


Fig. 3: Real and predicted close values for m=1 and n=20.

*B. Predicting two days ahead (m=2)*

In these experiments, the performance of the method in predicting the index value of the next two days, is evaluated for different values of $n$.

In Table 4, the values of different evaluation metrics are shown. According to Table 4, the proposed method gives the best performance when $n = 100$. Also, when $n = 60$ the $R^2$ index of the method is 99.38% which is very promising. In Table 5 more details of the results, for $n = 100$, are shown.

Table 5 indicates that SVR can predict the market situation for the next 2 days using only 4 and 2 features for the first and second clusters respectively. Furthermore, no technical indicators are needed for the prediction.

In other words, having the historical information of the past 100 days, the presented method can predict the stock market index for the next two days accurately.

Also, the best detected value for $\gamma$ for each of the trained models is 1.

In Fig. 4, the curves of real and estimated values are presented.

Table 4: Performance evaluation of the proposed method for m=2

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | $3.4\times10^{-4}$ | $5.03\times10^{-4}$ | $1.56\times10^{-4}$ | $8.43\times10^{-4}$ | **$7.82\times10^{-5}$** |
| RMSE | 0.0186 | 0.0224 | 0.0125 | 0.029 | **0.0088** |
| MAE | 0.0069 | 0.0075 | 0.0064 | 0.0071 | **0.0058** |
| $R^2$ | 0.986 | 0.9771 | 0.9938 | 0.9614 | **0.9958** |

Table 5: Details of the achieved results for m=2 and n=100

| | Number of selected features | $\gamma$ | Selected technical indicators |
|---|---|---|---|
| HHO-SVR-1 | 4 | 1 | None |
| HHO-SVR-2 | 2 | 1 | None |

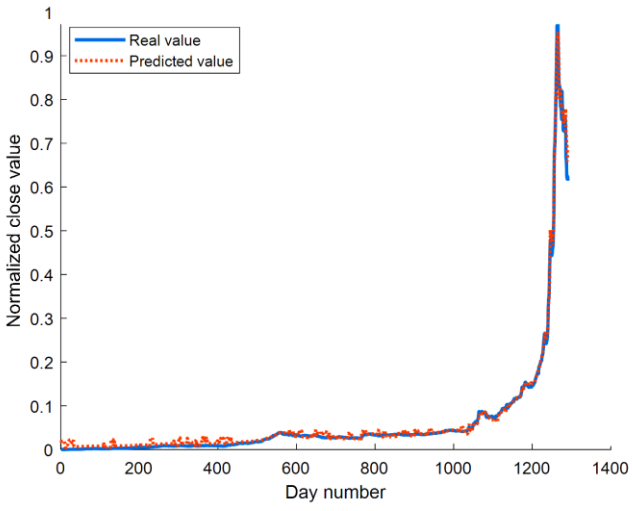


Fig. 4: Real and predicted close values for m=2 and n=100.

*C. Predicting three days ahead (m=3)*

The values of the evaluation metrics and the details of the best result are shown in Tables 6 and 7 respectively.

Table 6: Performance evaluation of the proposed method for m=3

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | $1.27\times10^{-4}$ | $6.13\times10^{-4}$ | $3.004\times10^{-4}$ | $9.18\times10^{-5}$ | **$9.16\times10^{-5}$** |
| RMSE | 0.0113 | 0.0248 | 0.0173 | 0.0096 | **0.0096** |
| MAE | **0.0045** | 0.0085 | 0.0057 | 0.0059 | 0.0068 |
| $R^2$ | 0.9938 | 0.9723 | 0.9883 | **0.9955** | 0.9952 |

Table 7: Details of the achieved results for m=3 and n=80

| | Number of selected features | $\gamma$ | Selected technical indicators |
|---|---|---|---|
| HHO-SVR-1 | 9 | 1 | None |
| HHO-SVR-2 | 178 | 5 | None |

According to Table 6, the least mean squared error is achieved when $n = 100$ while the best $R^2$ is 0.9955 for $n = 80$.

Also, when $n = 20$, the proposed method predicts the next three days fairly accurately (99.38%). Fig. 5 shows the real and estimated prices for this experiment (m=3 and n=80).

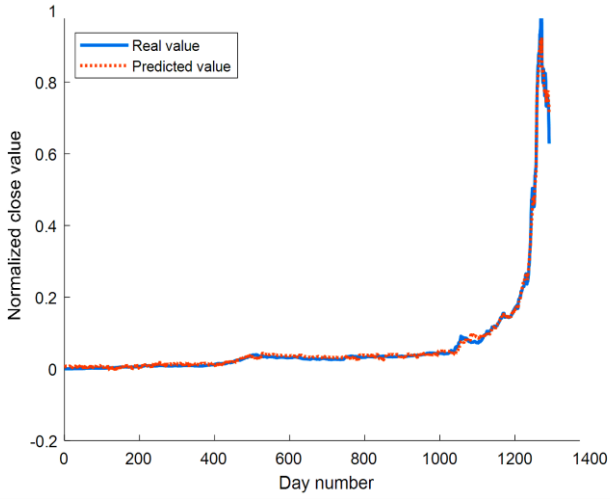


Fig. 5: Real and predicted close values for m=3 and n=80.

*D. Predicting four days ahead (m=4)*

In Table 8, the performance evaluation of the method is presented for $m = 4$.

According to this table, although the best performance ($R^2$ index) has been achieved for $n = 100$, the best performance in terms of mean absolute error (MAE) belongs to $n = 20$ and $n = 40$.

Table 8: Performance evaluation of the proposed method for m=4

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | $2.45\times10^{-4}$ | $5.12\times10^{-4}$ | 0.0013 | 0.0011 | **$1.35\times10^{-4}$** |
| RMSE | 0.0157 | 0.0226 | 0.036 | 0.0331 | **0.0116** |
| MAE | **0.0068** | 0.0068 | 0.0082 | 0.0093 | 0.0079 |
| $R^2$ | 0.9894 | 0.9766 | 0.9392 | 0.9604 | **0.994** |

In this case, the method has shown great performance for $n = 20$. This indicates that to predict the stock market index value of the next 4 days, using the historical information of the past 20 days will result in good performance although it is better to use the past 100 days. Besides that, giving the weakest performance for $n = 60$ reveals the fact that increasing the number of past days ($n$) will not always result in better performance.

It can add ambiguity to the data and thus degrade the performance of the method.

Table 9 shows that in the first phase of the method, two clusters have been detected by APSO-Clustering. For the first cluster, HHO has selected 218 features including the lower Bollinger band, while for the second cluster 276 features are selected without any of the technical indicators.

The best-detected values for $\gamma$ are 5 and 18 for the first and second clusters respectively.

Fig. 6 demonstrates the curves of real and predicted values for $n = 100$. Fig. 6 shows the high capability of the introduced method in detecting and tracking the market trend.

Table 9: Details of the achieved results for m=4 and n=100

| | Number of selected features | $\gamma$ | Selected technical indicators |
|---|---|---|---|
| HHO-SVR-1 | 218 | 5 | Lower Bollinger band |
| HHO-SVR-2 | 276 | 18 | None |

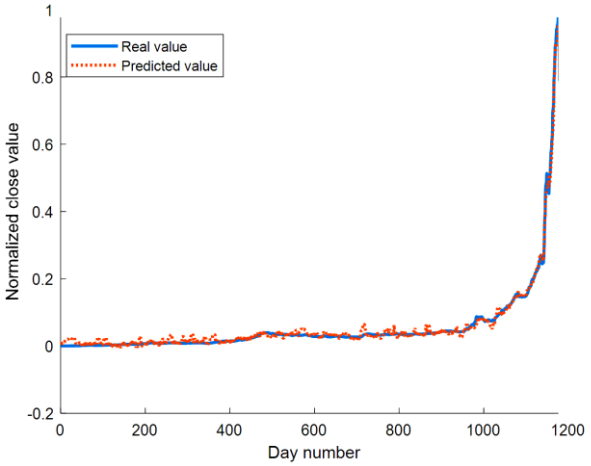


Fig. 6: Real and predicted close values for m=4 and n=100.

*E. Predicting five days ahead (m=5)*

Table 10 and Table 11 present the results provided by the introduced method for $m = 5$. These tables show that including the historical information of the last 40 days, APSO-Clustering has divided the data points of the created dataset into 2 clusters in the first phase. In the second phase, training two regression methods using 13 and 3 features of the first and second cluster's data respectively, have given the best performance. Furthermore, according to Table 10, using the historical data of the past 100 days has given the same result as using the historical data of the past 20 days which confirms that more number of days will not always result in better performance in decision making. Fig. 7 shows the real and estimated curves for $n = 40$.

Table 10: Performance evaluation of the proposed method for m=5

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | $6.31\times10^{-4}$ | **$2.1\times10^{-4}$** | 0.0018 | $7.6\times10^{-4}$ | $4.63\times10^{-4}$ |
| RMSE | 0.0251 | **0.0145** | 0.0425 | 0.0276 | 0.0215 |
| MAE | 0.008 | **0.0058** | 0.0107 | 0.0079 | 0.008 |
| $R^2$ | 0.9713 | **0.9908** | 0.9223 | 0.9702 | 0.9782 |

Table 11: Details of the achieved results for m=5 and n=40

| | Number of selected features | $\gamma$ | Selected technical indicators |
|---|---|---|---|
| HHO-SVR-1 | 13 | 1 | None |
| HHO-SVR-2 | 3 | 1 | None |

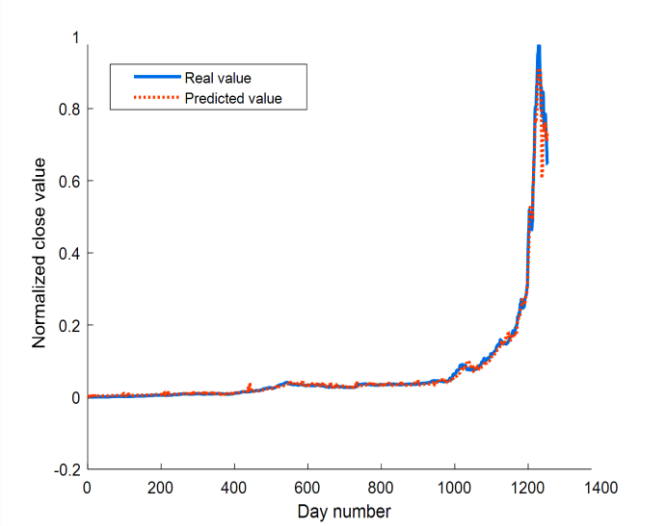


Fig. 7. Real and predicted close values for m=5 and n=40.

*F. Predicting six days ahead (m=6)*

In this case, according to Table 12, forecasting the market situation of the next 6 days using the historical data of the past 100 days has brought the best performance while the method has shown a very close performance for $n = 40$ ($R^2 = 0.9848$), $n = 60$ ($R^2 = 0.9934$) and $n = 80$ ($R^2 = 0.9946$). Table 13 shows that the second regression method has been trained on the data of the second cluster with 185 features including upper and lower Bollinger bands. In Fig. 8, the curves of the real and predicted values are shown.

Table 12: Performance evaluation of the proposed method for m=6

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | $7.98\times10^{-4}$ | $3.61\times10^{-4}$ | $1.39\times10^{-4}$ | $1.42\times10^{-4}$ | **$9.11\times10^{-5}$** |
| RMSE | 0.0282 | 0.019 | 0.0118 | 0.012 | **0.0095** |
| MAE | 0.0078 | 0.0065 | **0.006** | 0.0068 | 0.0068 |
| $R^2$ | 0.9668 | 0.9848 | 0.9934 | 0.9946 | **0.9958** |

Table 13: Details of the achieved results for m=6 and n=100

| | Number of selected features | $\gamma$ | Selected technical indicators |
|---|---|---|---|
| HHO-SVR-1 | 2 | 1 | None |
| HHO-SVR-2 | 185 | 4 | Upper and lower Bollinger bands |

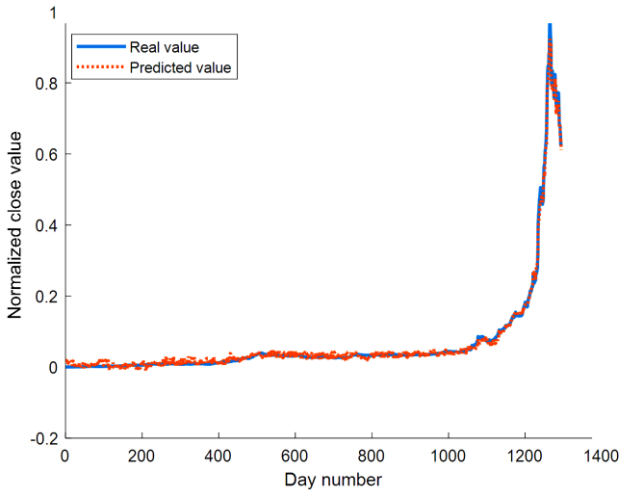


Fig. 8: Real and predicted close values for m=6 and n=100.

*G. Predicting seven days ahead (m=7)*

According to Table 14, like the previous case ($m = 6$), the best performance is achieved when $n = 100$, while for $n = 20$ and $n = 40$ the method has shown promising performance in forecasting the next 7 days.

Table 15 shows that including RSI, mid and lower Bollinger bands in the data points of the second cluster have a key role in reducing the error rate of the method. The great performance of the method is shown in Fig. 9 where the real and predicted curves are very close to each other.

Table 14: Performance evaluation of the proposed method for m=7

| N | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | $2.12\times10^{-4}$ | $2.19\times10^{-4}$ | $2.54\times10^{-4}$ | $4.25\times10^{-4}$ | **$1.28\times10^{-4}$** |
| RMSE | 0.0146 | 0.0148 | 0.160 | 0.0206 | **0.0113** |
| MAE | **0.0055** | 0.0063 | 0.0071 | 0.0067 | 0.0081 |
| $R^2$ | 0.9903 | 0.9905 | 0.9872 | 0.9808 | **0.9945** |

Table 15: Details of the achieved results for m=7 and n=100

| | Number of selected features | $\gamma$ | Selected technical indicators |
|---|---|---|---|
| HHO-SVR-1 | 2 | 1 | None |
| HHO-SVR-2 | 347 | 5 | RSI, mid and lower Bollinger band |

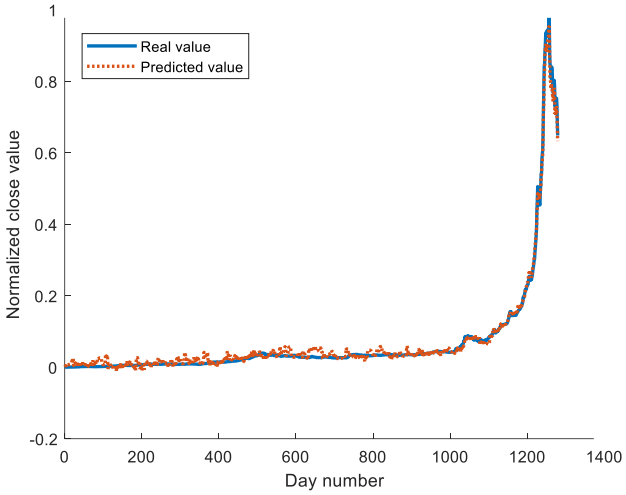


Fig. 9: Real and predicted close values for m=7 and n=100.

## Comparing With Other Metaheuristics, Self-Tuned SVR and Linear Regression

In order to reach a fair judgment about the efficiency of HHO-SVR, the second stage of the proposed method is implemented using other well-known metaheuristics (PSO, MVO [37], GSA [38], and IPO [39]). Just like HHO, the number of iteration and population for all of the metaheuristic optimization methods are set to 50 and 15 respectively. In Tables 16 to 19 the performance evaluation of PSO-SVR, MVO-SVR, GSA-SVR, and IPO-SVR for $m = 7$, are demonstrated respectively. To compare the results easily, the graphs of $R^2$ values of each method are shown in Fig. 10.

Table 16: Performance evaluation of PSO-SVR for m=7

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | 0.007 | 0.0154 | 0.0175 | 0.0017 | **1.14×10$^{-4}$** |
| RMSE | 0.0838 | 0.1243 | 0.1324 | 0.0414 | **0.0107** |
| MAE | 0.0179 | 0.0332 | 0.0405 | 0.0099 | **0.0077** |
| $R^2$ | 0.6776 | 0.3307 | 0.1220 | 0.9225 | **0.9951** |

Table 17: Performance evaluation of MVO-SVR for m=7

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | **4.13×10$^{-4}$** | 0.0083 | 0.0034 | 0.0016 | 0.0249 |
| RMSE | **0.0203** | 0.0912 | 0.0582 | 0.0401 | 0.1578 |
| MAE | **0.0074** | 0.022 | 0.0156 | 0.0094 | 0.0551 |
| $R^2$ | **0.9810** | 0.6396 | 0.8304 | 0.9273 | 0.0592 |

Table 18: Performance evaluation of GSA-SVR for m=7

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | 0.0135 | 0.0244 | 0.208 | 0.0027 | **0.0014** |
| RMSE | 0.1161 | 0.1561 | 0.1443 | 0.0516 | **0.0380** |
| MAE | 0.0366 | 0.0555 | 0.0511 | 0.0145 | **0.0125** |
| $R^2$ | 0.3812 | 0.0556 | 0.043 | 0.8798 | **0.9386** |

Table 19: Performance evaluation of IPO-SVR for m=7

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | 0.0228 | 0.0244 | 0.0184 | **0.0049** | 0.016 |
| RMSE | 0.1511 | 0.1561 | 0.1355 | **0.0698** | 0.1265 |
| MAE | 0.0547 | 0.0555 | 0.0452 | **0.0225** | 0.0843 |
| $R^2$ | 0.0487 | 0.0556 | 0.0799 | **0.7794** | 0.3188 |

Tables 13 to 18 and Fig. 10, show that HHO-SVR has a high potential in solving such a hard and complex optimization problem while the other methods failed in detecting the global optimum point successfully.

Besides that, according to Tables 13 to 18, all of the methods have shown a promising performance in predicting the index value of the next week using the historical information of the past 80 days.

In other words, using the information of the past 80 days reduces the complexity of the problem for the optimization algorithms which in turn increases the likelihood of finding the near-global optimum point by the optimization algorithm.

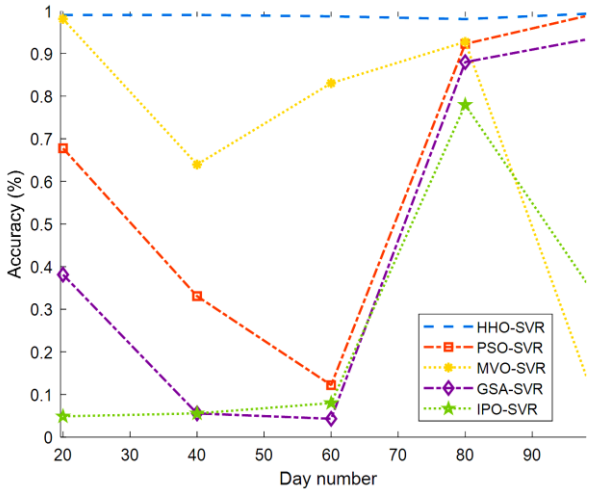


Fig. 10: $R^2$ values' graphs of different methods for m=7.

Also, to compare the performance of the whole method with other regression methods, the values of evaluation metrics of linear regression and the fine-tuned SVR, introduced in [41], for m=7, are presented in Tables 20 and 21 respectively. These two well-known regression methods are trained on the whole training data.

Table 20: Performance evaluation of linear regression SVR for m=7

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | 0.0218 | 0.0231 | 0.02 | 0.0221 | 0.0235 |
| RMSE | 0.1476 | 0.1519 | 0.1415 | 0.1486 | 0.1533 |
| MAE | 0.0746 | 0.0745 | 0.0721 | 0.0747 | 0.0753 |
| $R^2$ | 0.053 | 0.048 | 0.056 | 0.047 | 0.04 |

According to these two tables, although fine-tuned SVR has shown good performance but it cannot overcome the proposed method in predicting the stock market index value of the next 7 days.

Table 21: Performance evaluation of fine-tuned for m=7.

| n | 20 | 40 | 60 | 80 | 100 |
|---|---|---|---|---|---|
| MSE | $2.7\times10^{-4}$ | $7.9\times10^{-4}$ | 0.0025 | $5.4\times10^{-4}$ | 0.0055 |
| RMSE | 0.0165 | 0.0282 | 0.05 | 0.0233 | 0.0743 |
| MAE | 0.0055 | 0.0095 | 0.0203 | 0.0125 | 0.0282 |
| $R^2$ | 0.9876 | 0.9655 | 0.87 | 0.9755 | 0.7652 |

## Conclusion and Future Works

In this paper, an effective method for forecasting the future of the stock market is proposed which works in two stages. In the first stage, the training dataset is clustered using a novel automatic clustering method, called APSO-Clustering that can detect the proper number of clusters and the position of the centroids simultaneously.

This capability is very valuable when dealing with massive and high-dimensional datasets. In the second stage a hybrid regression method, called HHO-SVR, is trained for each cluster's data points.

In this regression method, HHO is utilized for feature selection and parameter tuning of SVR. To estimate the target value of an unknown sample, after determining its cluster, the corresponding regression method estimates the target value.

The main goal of this research was to predict the future of the Tehran Stock Exchange market. For this purpose, the historical data of the market index in addition to some technical indicators are used for data preparation. Several experiments have been conducted to evaluate the accuracy and effectiveness of the method.

In fact, in this research we have gone further in compare to our previous work [40]. While in our previous research we have analyzed the performance of the proposed method in predicting the price of the stocks in one day ahead, in this research we have tried to forecast the situation of the whole market in the next seven days which gives the traders a good opportunity to make a proper decision.

Besides, we have tried to forecast the market index of the next $m$ (from 1 to 7) days using the historical data of the past $n$ (from 10 to 100) days. The experiments show that increasing the number of days ($n$), used to create the dataset, will not necessarily improve the final accuracy of the method. Although in the last two experiments, the best performance has been achieved for $n = 100$, in most experiments the method has shown high accuracy in forecasting the future using the past 20 days.

On the other hand, in this research a new metaheuristic optimization algorithm is used for prediction which has shown a great accuracy in forecasting the market index value in the next week. This method has shown a great performance in predicting and tracking trends. Definitely, several unpredictable factors (political events, natural disasters, economic situation and etc.) affects the stock market that can produce sudden fluctuations.

These sudden jumps and rises in the index value are hard to be predicted since their causing factors are hardly predictable. Nevertheless, the most important thing is the prediction of the trends in the near future, that the methodology introduced in this paper performs it effectively.

For future works, the effect of different information (such as dollar exchange rate and inflation rate) on the performance of the proposed method, can be investigated.

Also, the method can be extended in order to process other types of data such as news and twits. Furthermore, the proposed method can be used to predict the future of other markets such as cryptocurrencies, which are very popular and interested nowadays.

Besides, other regression methods such as logistic regression and also deep learning methods can be used in the second phase, instead of HHO-SVR, to investigate their performance. Generally, our method is useful in solving different important regression problems such as electricity forecasting and etc.

## Author Contributions

The introduced method is designed by Iman Behravan, while Dr. Seyed Mohammad Razavi has interpreted the final results.

## Acknowledgements

We would like to appreciate Dr.Roberto Trasarti and Prof. Seyyed Hamid Zahiri for their helps and advices in the road of designing APSO-CLUSTERING.

## Conflict of Interest

The authors declare no potential conflict of interest regarding the publication of this work. In addition, the ethical issues including plagiarism, informed consent, misconduct, data fabrication and, or falsification, double publication and, or submission, and redundancy have been completely witnessed by the authors.

## Abbreviations

| | |
|---|---|
| *PSO* | Particle Swarm Optimization |
| *APSO-Clustering* | Automatic Particle Swarm Optimization-Clustering |
| *HHO* | Harris's Hawks Optimization |
| *MVO* | Multi-Verse Optimization |
| *GSA* | Gravitational Search Algorithm |
| *IPO* | Inclined Planes Optimization |

## References

[1] H. Chen, D.L. Fan, L. Fang, W. Huang, J. Huang, C. Cao, et al., "Particle swarm optimization algorithm with mutation operator for particle filter noise reduction in mechanical fault diagnosis," Int. J. Pattern Recognit. Artif. Intell., 34: 2058012, 2020.

[2] M. Janga Reddy, D. Nagesh Kumar, "Evolutionary algorithms, swarm intelligence methods, and their applications in water resources engineering: a state-of-the-art review," H2Open J., 3: 135-188, 2021.

[3] Q.V. Pham, D.C. Nguyen, S. Mirjalili, D.T. Hoang, D.N. Nguyen, P.N. Pathirana, et al., "Swarm intelligence for next-generation wireless networks: Recent advances and applications," arXiv preprint arXiv:2007.15221, 2020.

[4] M. Schranz, G.A. Di Caro, T. Schmickl, W. Elmenreich, F. Arvin, A. Şekercioğlu, et al., "Swarm intelligence and cyber-physical systems: concepts, challenges and future trends," Swarm Evol. Comput., 60: 100762, 2021.

[5] A. Kaveh, A.D. Eslamlou, Metaheuristic optimization algorithms in civil engineering: New applications: Springer, 2020.

[6] B. Yang, J. Wang, X. Zhang, T. Yu, W. Yao, H. Shu, et al., "Comprehensive overview of meta-heuristic algorithm applications on PV cell parameter identification," Energy Convers. Manage., 208: 112595, 2020.

[7] J. Nocedal, S. Wright, Numerical optimization: Springer Science & Business Media, 2006.

[8] G. Wu, "Across neighborhood search for numerical optimization," Inf. Sci., 329: 597-618, 2016.

[9] M. Usmani, S.H. Adil, K. Raza, S.S.A. Ali, "Stock market prediction using machine learning techniques," in Proc. 2016 3rd international conference on computer and information sciences (ICCOINS): 322-327, 2016.

[10] S. Pyo, J. Lee, M. Cha, H. Jang, "Predictability of machine learning techniques to forecast the trends of market index prices: Hypothesis testing for the Korean stock markets," PloS one, 12: e0188107, 2017.

[11] M.R. Senapati, S. Das, S. Mishra, "A novel model for stock price prediction using hybrid neural network," J. Inst. Eng. (India): Ser. B, 99: 555-563, 2018

[12] H. Hu, L. Tang, S. Zhang, H. Wang, "Predicting the direction of stock markets using optimized neural networks with Google Trends," Neurocomputing, 285: 188-195, 2018.

[13] X. Pang, Y. Zhou, P. Wang, W. Lin, V. Chang, "An innovative neural network approach for stock market prediction," J. Supercomputing, 76: 2098-2118, 2020.

[14] N. Gozalpour, M. Teshnehlab, "Forecasting stock market price using deep neural networks," in Proc. 2019 7th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS): 1-4: 2019.

[15] M. Ghanbari, H. Arian, "Forecasting stock market with support vector regression and butterfly optimization algorithm," arXiv preprint arXiv:1905.11462, 2019.

[16] M. Vijh, D. Chandola, V. A. Tikkiwal, A. Kumar, "Stock closing price prediction using machine learning techniques," Procedia Comput. Sci., 167: 599-606, 2020.

[17] F. Ecer, S. Ardabili, S.S. Band, A. Mosavi, "Training multilayer perceptron with genetic algorithms and particle swarm optimization for modeling stock price index prediction," Entropy, 22: 1239, 2020.

[18] M. Nabipour, P. Nayyeri, H. Jabani, S. Shahab, A. Mosavi, "Predicting stock market trends using machine learning and deep learning algorithms via continuous and binary data; a comparative analysis," IEEE Access, 8: 150199-150212, 2020.

[19] M.J. Awan, M.S.M. Rahim, H. Nobanee, A. Munawar, A. Yasin, A.M. Zain, "Social media and stock market prediction: A big data approach," Comput. Mater. Continua, 67: 2569-2583, 2021.

[20] I. K. Nti, A.F. Adekoya, B.A. Weyori, "A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction," J. Big Data, 8: 1-28, 2021.

[21] S. Tuarob, P. Wettayakorn, P. Phetchai, S. Traivijitkhun, S. Lim, T. Noraset, et al., "DAViS: a unified solution for data collection, analyzation, and visualization in real-time stock market prediction," Financ. Innovation, 7: 1-32, 2021.

[22] M. Ali, D.M. Khan, M. Aamir, A. Ali, Z. Ahmad, "Predicting the direction movement of financial time series using artificial neural network and support vector machine," Complexity, 2021: 1-13, 2021.

[23] I. Behravan, S.H. Zahiri, S.M. Razavi, R. Trasarti, "Clustering a big mobility dataset using an automatic swarm intelligence-based clustering method," J. Electr. Comput. Eng. Innovations, 6(2): 251-271, 2018.

[24] I. Behravan, S.H. Zahiri, S. M. Razavi, R. Trasarti, "Finding roles of players in football using automatic particle swarm optimization-clustering algorithm," J. Big data, 7: 35-56, 2019.

[25] I. Behravan, S.M. Razavi, "A novel machine learning method for estimating football players' value in the transfer market," Soft Comput., 25: 2499-2511, 2021.

[26] J.C. Bednarz, "Cooperative hunting Harris' hawks (Parabuteo unicinctus)," Science, 239: 1525-1527, 1988.

[27] A.A. Heidari, S. Mirjalili, H. Faris, I. Aljarah, M. Mafarja, H. Chen, "Harris hawks optimization: Algorithm and applications," Future Gener. Comput. Syst., 97: 849-872, 2019.

[28] S.M. Yusuf, C. Baber, "Multi-Agent searching adaptation using levy flight and inferential reasoning," Int. J. Electr. Inf. Eng., 14: 290-297, 2020.

[29] X.S. Yang, Nature-inspired metaheuristic algorithms: Luniver press, 2010.

[30] J. Cheng, D. Yu, Y. Yang, "Application of support vector regression machines to the processing of end effects of Hilbert–Huang transform," Mech.l Syst. Sig. Process., 21: 1197-1211, 2007.

[31] M. Sabzekar, S.M.H. Hasheminejad, "Robust regression using support vector regressions," Chaos, Solitons & Fractals, 144: 110738, 2021.

[32] H. Drucker, C.J. Burges, L. Kaufman, A. Smola, V. Vapnik, "Support vector regression machines," Adv. Neur. Inf. Process. Syst., 9: 155-161, 1997.

[33] I. Behravan, O. Dehghantanha, S. H. Zahiri, "An optimal SVM with feature selection using multi-objective PSO," in Proc. 2016 1st Conference on Swarm Intelligence and Evolutionary Computation (CSIEC): 76-81, 2016.

[34] B. Madhu, A.K. Paul, R. Roy, "Performance comparison of various kernels of support vector regression for predicting option price," Int. J. Discrete Math., 4: 21, 2019.

[35] Y. Tang, W. Guo, J. Gao, "Efficient model selection for support vector machine with Gaussian kernel function," in Proc. 2009 IEEE Symposium on Computational Intelligence and Data Mining: 40-45, 2009.

[36] W. Wang, Z. Xu, W. Lu, X. Zhang, "Determination of the spread parameter in the Gaussian kernel for classification and regression," Neurocomputing, 55: 643-663, 2003.

[37] S. Mirjalili, S. M. Mirjalili, A. Hatamlou, "Multi-verse optimizer: a nature-inspired algorithm for global optimization," Neur. Comput. Appl., 27: 495-513, 2016.

[38] E. Rashedi, H. Nezamabadi-Pour, S. Saryazdi, "GSA: a gravitational search algorithm," Inf. Sci., 179: 2232-2248, 2009.

[39] M.H. Mozaffari, H. Abdy, S.H. Zahiri, "IPO: an inclined planes system optimization algorithm," Comput. Inf., 35: 222-240, 2016.

[40] I. Behravan, S. Razavi, "Stock price prediction using machine learning and swarm intelligence," J. Electr. Comput. Eng. Innovations, 8(1):31-40, 2020.

[41] R.K. Dash, T.N. Nguyen, K. Cengiz, A. Sharma, "Fine-tuned support vector regression model for stock predictions," Neur. Comp. Appl., 2021: 1-15, 2021.

## Biographies



**Iman Behravan** received his B.Sc. in electronics engineering from Shahid Bahonar University of Kerman, Kerman, Iran. Also, he received his M.Sc. and Ph.D. degrees from the University of Birjand, Birjand, Iran. Now he is a post-doctoral researcher at the University of Birjand under the supervision of Professor Seyed Mohamad Razavi. His research interests include Big data analytics, pattern recognition, machine learning, swarm intelligence, and soft computing.


- Email: i.behravan@birjand.ac.ir
- ORCID ID: 0000-0003-0319-1765
- Web of science ID: J-5326-2017
- Scopus ID: 57190213695
- Home page: https://scholar.google.com/citations?user=w9GKiVcAAAAJ&hl=en




**Seyyed Mohammad Razavi** received the B.Sc. degree in Electrical Engineering from the Amirkabir University of Technology, Tehran, Iran, in 1994 and the M.Sc. degree in Electrical Engineering from the Tarbiat Modares University, Tehran, Iran, in 1996, and the Ph.D. degree in Electrical Engineering from the Tarbiat Modares University, Tehran, Iran, in 2006. Now, he is an Associate Professor in the Department of Electrical and Computer Engineering, the University of Birjand, Birjand, Iran. His research interests include Computer Vision, Pattern Recognition, and Artificial Intelligence Algorithm.


- Email: smrazavi@birjand.ac.ir
- ORCID ID: 0000-0002-3493-7614
- Web of science ID: AGS-8258-2022
- Scopus ID: 3861265
- Home page: https://scholar.google.com/citations?user=iSd4OusAAAAJ&hl=en

**Copyrights**
©2021 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, as long as the original authors and source are cited. No permission is required from the authors or the publishers.



**How to cite this paper:**
I. Behravan, S.M. Razavi, "Application of Harris Hawks optimization algorithm and APSO-CLUSTERING in predicting the stock market," J. Electr. Comput. Eng. Innovations, 10(2): 447-462, 2022.

**DOI:** 10.22061/JECEI.2022.8659.541

**URL:** https://jecei.sru.ac.ir/article_1693.html

