Fine-tuned support vector regression model for stock predictions

Ranjan Kumar Dash[1] · Tu N. Nguyen[2] · Korhan Cengiz[3] 🔟 · Aditi Sharma[4]

Abstract

In this paper, a new machine learning (ML) technique is proposed that uses the fine-tuned version of support vector regression for stock forecasting of time series data. Grid search technique is applied over training dataset to select the best kernel function and to optimize its parameters. The optimized parameters are validated through validation dataset. Thus, the tuning of this parameters to their optimized value not only increases model's overall accuracy but also requires less time and memory. Further, this also minimizes the model from being data overfitted. The proposed method is used to analysis different performance parameters of stock market like up-to-daily and up-to-monthly return, cumulative monthly return, its volatility nature and the risk associated with it. Eight different large-sized datasets are chosen from different domain, and stock is predicted for each case by using the proposed method. A comparison is carried out among the proposed method and some similar methods of same interest in terms of computed root mean square error and the mean absolute percentage error. The comparison reveals the proposed method to be more accurate in predicting the stocks for the chosen datasets. Further, the proposed method requires much less time than its counterpart methods.

Keywords Grid search · Machine learning · Root mean square error · Mean absolute percentage error · Support vector regression · Volatility

✉ Korhan Cengiz
korhancengiz@trakya.edu.tr

Ranjan Kumar Dash
drranjandash@gmail.com

Tu N. Nguyen
nguyent@pfw.edu

Aditi Sharma
aditi11121986@gmail.com

[1] College of Engineering and Technology Bhubaneswar, Bhubaneswar, Odisha, India

[2] Department of Computer Science, Purdue University Fort Wayne, Fort Wayne, IN 46805, USA

[3] Department of Electrical-Electronics Engineering, Faculty of Engineering, Trakya University, Edirne, Turkey

[4] Department of Computer Science and Engineering, Quantum University, Roorkee, Uttarakhand, India

1 Introduction

Stock market prediction is a forecasting method that relies on technical aspects of the stock price to predict its future value. The success of such a predictive system mainly depends on the availability of a huge amount of historical data so that it can be used in the pursuit of the lucrative financial markets. The data confined for this type of study are financial time series data which puts stringent constraints on the performance of these models. Furthermore, the risk associated with such models cannot be overlooked since risks are implicit due to irregular market trends, instability, noise, etc. [1]. Thus, the predictive models inherently obey the efficient market hypothesis (EMH) which states that the risk-adjusted return cannot consistently be obtained above the profitability of the whole [1, 2]. EMH assumes that the current stock price can be determined as a function of stock price history and rational expectations. Any deviation to this assumption may leave the stock price to be unpredictable. However, with the advancement of computational facilities, various ML

methods were simulated efficiently to form the stock price predictions with new technologies. The prominent algorithms that have been used by the researchers to predict stock price include an artificial neural network (ANN) and its variants, genetic algorithm, support vector machine (SVM), support vector regression (SVR), etc. The most common challenges to such predictive models are to deal with a risk while predicting the stock price with greater accuracy which in turn minimizes risks for stock market investors and results in a profitable strategy. The accuracy of these models attracts more researchers and makes them motivated to propose new predictive techniques with greater accuracy.

The proposed study in this paper is an attempt towards improve the precision of the predictive model in an time efficient manner by utilizing SVR as an predictive method. The organization of the paper is as follows: Related work is explained elaborately in Sect. 2, Sect. 3 provides a mathematical understanding about the support vector regression, a new SVR approach is proposed in Sect. 4, Sect. 5 comprises the simulation results and discussions followed by Sect. 5 which makes conclusions of the paper with future scopes.

2 Related work

Nowadays, applied machine learning has been widely studied in diverse applications [3–11]. The study carried out in [12] presents a combined method of autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) models for stock prediction. Hamzaebi et al. [13] propose two artificial neural network-based methods for multi-periodic forecasting. The first one is an iterative method that uses past observations to predict subsequent period information. This predicted value is used for the prediction of subsequent periods. These operations are repeated until stopping criterion is met. The second one is an forecast approach in which subsequent periods can be estimated all at once. The main result of this study shows that the direct scheme superiors the iterative model. The authors of paper [14] elaborate how artificial neural network (ANN)-based methods can be used as an efficient model in stock-market predictions by analyzing the following models: multi-layer perceptron (MLP), dynamic artificial NN (DAN2) and the hybrid NNs (HNN) using the generalized autoregressive conditional heteroscedasticity (GARCH). Then, they build an efficient stock index prediction scheme for the Shanghai composite index. The model uses a genetic method to train the backpropagation neural network (BPNN). The study concludes that the designed method is better in terms of function approximating capacity yielding ideal results for stock forecasting.

The work carried out in [15] a new method based on wavelet de-noising-based backpropagation (WDBP) neural network is proposed for stock prediction. A new model [16] that uses evolving partially connected neural networks (EPCNNs) for stock prediction. Their model architecture is different from traditional artificial neural networks because they have random connections among neurons, more than one hidden layer, and use evolutionary algorithms to train the artificial neural network.

Jo et al. [17] propose a method that has a filter and a modified genetic algorithm (MGA). MGA is used to set initial parameters for morphological-rank-linear (MRL) filters and to fine tune these parameters. The least mean squares (LMS) algorithm is used to further develop these parameters. The work carried out in [18] uses the least square SVM (LS-SVM) with an integration of particle swarm optimization (PSO) to predict the daily stock prices. The hyperparameters of LS-SVM are optimized by using the PSO algorithm. The authors claim that the optimized hyperparameters avoid data over-fitting and local minima issues, and thus, they improve the predictions of the precision. [19] uses hybrid intelligent model for stock prediction. The proposed model in [20] a new regression model is employed to generate training data for the recurrent neural network. The paper [21] proposes a hybrid model which is utilized to train the adaptive stock market predictions while optimizing several performance metrics.

In [22], a comparative study among different types of basis functions has been performed to find the optimal combinations of these with greater accuracy for prediction. The paper [23] combines fuzzy time series with granular calculating methods. It is shown that the numerical results obtained from this model outperform the support regression (SVR), fuzzy GARCH time series schemes as well as the hybrid fuzzy time series schemes. The paper [24] proposes a hybrid model consisting of SVM and K-nearest neighbor (KNN) method for prediction of Indian stock market indices. The study in [25] proposes back-propagation algorithm for stock price prediction. A new kernel function for support vector regression is proposed in [26] for stock market predictions. The authors of the paper [27] use proximal SVM as a predictive method for estimating twelve different stock indices. Joint prediction error is also proposed int this paper as a new performance measure. Artificial neural network is used in [28] for predicting the daily NASDAQ stock exchange rates.

The paper [29] proposes a method to predict volatility from financial data. ARIMA and MLP are combined in [30] for the stock prediction. To improve the prediction capacities of the stock price trends, a method [31] is proposed. Henrique et al. [2] use SVR to model the prediction the stock markets for daily and up-to-minute frequencies. Senapati et al. [32] propose a hybrid method for the stock

prediction. In this hybrid method, the weight updation of Adaline NN is governed by PSO. Authors of [33] present a comparative paper on the performance of several schemes on tick dataset and 15-min dataset of an Indian company. This comparison concludes that the tick dataset has preferable accuracy than the 15-min dataset.

From the discussion made so far, the following observations can be drawn:

- Stock forecasting involves time series data containing highly unpredictable errors. This is the main reason that artificial neural network and its variants are dominantly become the best choice of the researchers. However, irrespective of the type of learning algorithms used, the time to converge to optimal solution by using artificial neural networks or its variants is a major concerned.
- The fusion or combination of machine learning algorithms is found to be outperformed. However, such techniques require much more hyperparameters and thus requires more memory as well as time.
- However, support vector regression is observed to be a prominent technique for stock forecasting with a good measure of accuracy. The lack of fine tuning its parameters may lead to a time consuming method which distracts the researchers to use this technique further.

The above-mentioned problems motivates us to propose a new method based on SVR for stock forecasting. In this work, a method that efficiently manages the risk in terms of errors is presented. Grid search technique is used to optimize the parameters of support vector regression for entire dataset.

3 Support vector regression (SVR)

Suppose the training dataset D contains d number of instances each with an attribute x_i and its associated class y_i i.e. $\{(x_1, y_1), (x_2, y_2), \cdots (x_d, y_d)\}$.

The linear function $f(x)$ on the dataset D can be defined as:

$$f(x) = w_1 x_1 + w_2 x_2 + \cdots + w_d x_d + b \tag{1}$$

where each weight w_i is defined over real valued input space R^d i.e. $w_i \epsilon R^d$

The size of maximal margin is defined by the Euclidean norm of weights ($\|w\|$). Hence, the flatness in the case of Eq. (1) requires the minimization of the norm of weights. Here, $\|w\|$ is defined as:

$$\|w\|^2 = w_1^2 + w_2^2 + \cdots + w_d^2$$

The error for each training data $<x_i, y_i>$ can be expressed as

$$E_i(x_i) = y_i - (w_i x_i + b) \tag{2}$$

If the deviation of each error $E_i(x_i)$ is allowed to be within ϵ, Eq. (2) can be written as

$$y_i - (w_i x_i + b) \leq \epsilon$$
$$(w_i x_i + b) - y_i \leq \epsilon \tag{3}$$

Using Eqs. (2) and (3), the minimization problem for w can be

$$\text{minimize:} \ \frac{1}{2}\|w\|^2$$

$$\text{subject to:} \tag{4}$$

$$y_i - (w_i x_i + b) \leq \epsilon$$
$$(w_i x_i + b) - y_i \leq \epsilon$$

The constraints of Eq. (4) place assumptions that the function f approximates all the pairs (x_i, y_i) with a deviation of ϵ. However, the assumption does not hold for all cases which in turn requires the slack variables ξ_i, ξ_i^* to cope with the cases that violates the assumption. By using the slack variables, the optimization problem can be redefined as:

$$\text{minimize:} \ \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{d}(\xi_i + \xi_i^*)$$

$$\text{subject to}$$

$$\forall i: y_i - (w_i x_i + b) \leq \epsilon + \xi \tag{5}$$

$$\forall i: (w_i x_i + b) - y_i \leq \epsilon + \xi^*$$

$$\forall i: \xi_i \geq 0$$

$$\forall i: \xi_i^* \geq 0$$

The constant C is the penalty imposed on the observed data that does not satisfy constraints (5). It also helps to minimize data overfitting.

The linear ϵ-insensitive loss function (ι_ϵ) is defined as the loss function that assumes zero errors for the observed data satisfying (5). For the rest cases, its measure is based on the distance between observed data and the error margin i.e. ϵ. Mathematically, it can be written as:

$$\ell_\epsilon = \begin{cases} 0 & \text{if} \ \ |y - f(x)| \leq \epsilon \\ |y - f(x)| - \epsilon & \text{Otherwise} \end{cases} \tag{6}$$

The dual formula for linear SVR is constructed by using a Lagrangian function from Eq. (5) and by introducing non-negative multipliers α_i and α_i^* for each observation x_i.

$$L(\alpha) = \frac{1}{2} \sum_{i=1}^{d} \sum_{j=1}^{d} (\alpha_i - \alpha_i^*) x_i x_j$$

$$+ \epsilon \sum_{i=1}^{d} (\alpha_i + \alpha_i^*) + \sum_{i=1}^{d} y_i (\alpha_i^* - \alpha_i)$$

subject to $\qquad\qquad\qquad\qquad$ (7)

$$\sum_{i=1}^{d} (\alpha_i - \alpha_i^*) = 0$$

$$\forall i : 0 \le \alpha_i \le C$$

$$\forall i : 0 \le \alpha_i^* \le C$$

The partial derivative of L in regard to the weight vector is:

$$\delta_w L = w - \sum_{i=1}^{d} (\alpha_i - \alpha_i^*) x_i \qquad (8)$$

This partial derivative of L must be equal to 0 as per the saddle point condition. Thus, Eq. (8) becomes:

$$w - \sum_{i=1}^{d} (\alpha_i - \alpha_i^*) x_i = 0 \qquad (9)$$

Equation (9) is called SV expansion i.e. the weight vector can be defined as the linear combinations of training data x_i.

Thus Eq. (1) can be written as:

$$f(x) = \sum_{i=1}^{d} (\alpha_i - \alpha_i^*)(x_i.x) + b \qquad (10)$$

Equation (10) can be used to predict the new value x.

Let us define the following mapping:

$$\Phi : X \rightarrow f \qquad (11)$$

where f is the high dimensional feature space, ϕ is the nonlinear mapping function applied to convert the training tuple x_i to the dimension of feature space.

Since the operation involved between training tuples x_i and x_j in Eq. (10) is the dot product, thus representation in feature space is expressed as $\phi(x_i).\phi(x_j)$. However, in order to avoid performing the dot product on the transformed data tuples, a kernel function defined over original input data $K(x_i, x_j)$ can be used to replace each occurrence of dot product between transformed tuples. By doing so, all the calculations are made in the original input spaces. The admissible kernel function includes the followings:

- Linear kernel of degree n:

$$K(x_i, x_j) = x_i x_j \qquad (12)$$

- Polynomial kernel of degree n:

$$K(x_i, x_j) = (x_i.x_j + 1)^n \qquad (13)$$

- Radial basis kernel/Gaussian kernel:

$$K(x_i, x_j) = \exp{-(\|(x_i.x_j + 1)\|^2)/2\sigma^2} \qquad (14)$$

4 FTSVR: Fine-tuned support vector regression model for stock predictions

The proposed model uses SVR for stock forecasting. The mathematical description on the working support vector regression studies is presented in Sect. 2. The steps involved in the proposed model for stock forecasting are enumerated below:

1. Selection of the dataset and its pre-processing
2. Creation of training and validation datasets
3. Tune the parameters used in support vector regression model such that the model yields high accuracy in a time efficient manner
4. Training and validation of the proposed support vector regression model-FTSVR
5. Performance analysis of the proposed model-FTSVR for stock analysis

The steps involved in FTSVR, fine-tuned support vector regression model for stock predictions, are presented in Fig. 1.

4.1 Selection of dataset and its pre-processing

The SBIN[1] dataset used throughout this section includes examples to show each step of our model.

The dataset contains the following attributes:

1. *Date* The date for which stock data are observed
2. *Open* Opening price of share for a particular date
3. *High* Maximum price of the share on that date
4. *Low* Minimum price of the share on that date
5. *Close* final price of the stock for that date
6. *AdjClose* Number of shares bought or sold in the day
7. *Volume* Turnover of a company for a given date

The number of missing values for each of the attribute is shown in Table 1 (Fig. 2).

The missing values occur due to the following reasons:

- National holidays or Sunday
- The value is missing even though it is not a holiday

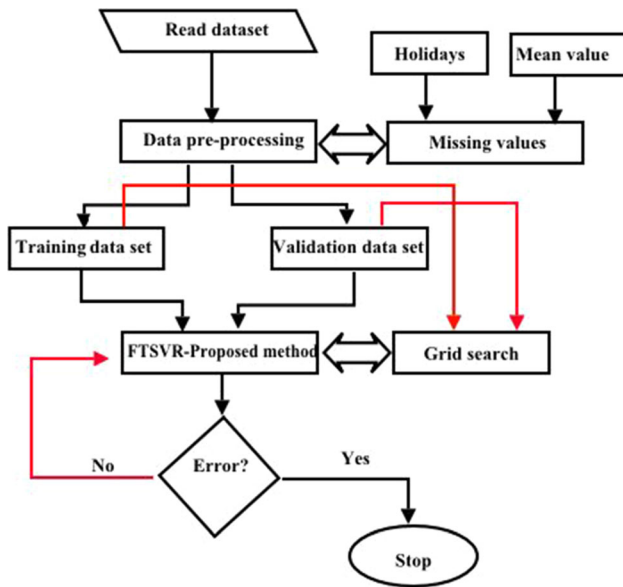The following approaches are adopted in this work to deal with missing values:

[1] https://finance.yahoo.com/quote/SBIN.NS/history/.

Fig. 1 FTSVR: Fine-tuned support vector regression model for stock predictions

Table 1 Number of missing values for each of the attribute

Open	127
High	127
Low	127
Close	127
Adj close	127
Volume	131

1 The dataset is checked against the list of Sunday or holidays for each year and the matching records are dropped from the dataset.

2 For the missing values of the refined dataset, the average value of the attribute is calculated over one month and this average value is utilized to fill the
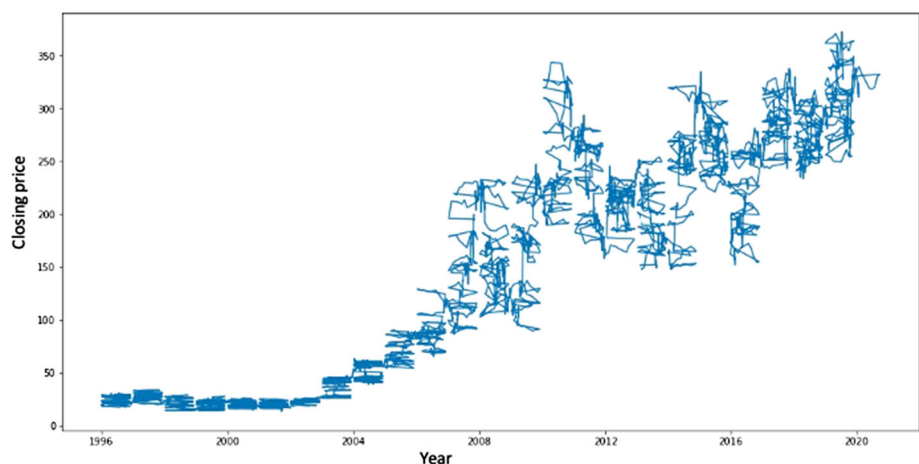
lacking values. This process is repeated for all the missing values (Fig. 3, Table 2).

4.2 Creation of training and validation dataset

Normally, the profit or loss is usually determined by the value of the closing price of a stock for a particular day; hence, closing price is treated as the target variable [12].

The stock prediction problem is stated as predicting the closing prices of stocks of a particular company for a specific date. This problem requires the two attributes, viz. date and close. Hence, the dataset contains only two attributes indexed by date. The first ten records of new dataset are shown in Table 3.

Out of dataset, 70% records i.e. 4163 are used for training while the rest number of records i.e. 1784 are for the purpose of testing.

4.3 Selection of kernel function and parameters tuning

The performance of machine learning algorithms largely depends on their parameters. So, it is almost essential to fine tune the parameters to design a predictive model that predicts future outcome with greater accuracy. Further, the tuning of parameters avoids the model to suffer from data overfitting which is a serious issue in machine learning. The selection of kernel function and its optimized parameters of a support vector regression is a great challenge since these are very much data oriented.

The linear model has single parameter to optimize which is C, in radial basis function (RBF) model there are two parameters: C and γ while the polynomial model has three parameters: C, γ and degree. To find the optimal

Fig. 2 SBIN dataset

Fig. 3 SBIN dataset after processing missing values
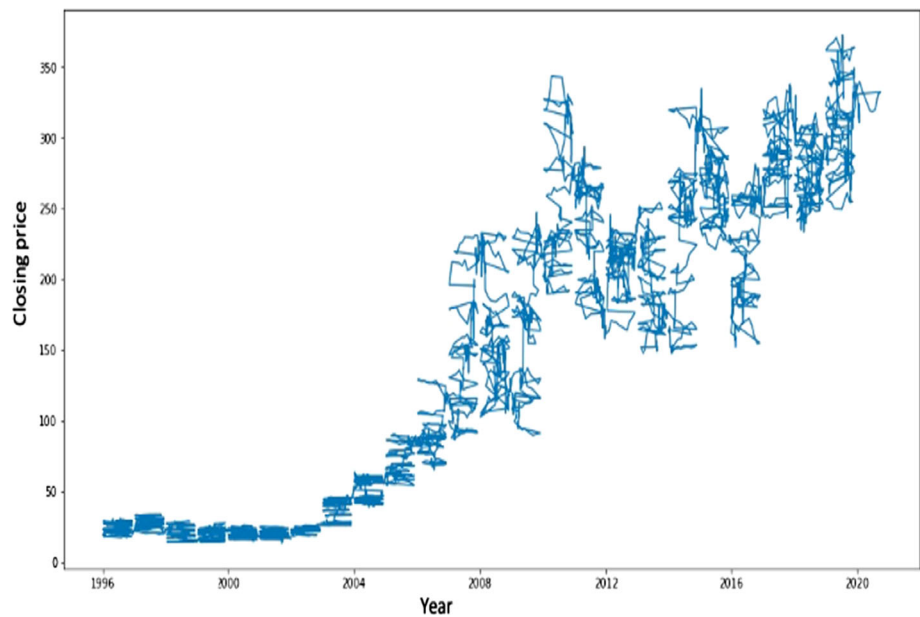
Table 2 List of holidays for the year-2019

Date	Day	
04-March-2019	Mon.	Mahashivratri
21-March-2019	Thurs.	Holi
17-Apr.-2019	Wed.	M. Jayanti
19-Apr.-2019	Fri.	Good Friday
29-Apr.-2019	Mon.	Parliamentary Elect.
01-May-2019	Wed.	Maharashtra Day
05-June-2019	Wed.	Id-Ul-Fitr (Ramzan ID)
12-Aug.-2019	Mon.	Bakri Id
15-Aug.-2019	Thurs.	Indep. Day
02-Sept.-2019	Mon.	G. Chaturthi
10-Sept.-2019	Tues.	Moharram
02-Oct.-2019	Wed.	M. G. Jayanti
08-Oct.-2019	Tues.	Dasera
21-Oct.-2019	Mon.	General Assembly Elect. in Maharashtra
28-Oct.-2019	Mon.	Diwali-Balipratipada
12-Nov.-2019	Tues.	G. Jayanti
25-Dec.-2019	Wed.	Christmas

Table 3 First ten records of new dataset

Date	Close
1996-01-01	18.8232
1996-01-02	18.2241
1996-01-03	17.7382
1996-01-04	17.6769
1996-01-05	17.5778
1996-01-08	17.0636
1996-01-09	17.9977
1996-01-10	17.1721
1996-01-11	17.8278

Table 4 Parameters and their range

Kernel function	C	γ	Degree
Linear, RBF	0.001 to 10,000	0.001 to 10,000	1 to 5
Polynomial	Step 10	Step 10	Step 1

values of different parameters, grid search technique [34] is employed.

Grid search optimizes different SVR parameters to select a combination of parameters; thus, the method can predict unknown data exactly.

Grid search technique is applied on the training dataset by inputting different kernel function and parameters (Table 4). The validation is performed over the validation dataset. The obtained optimized parameters are as follows:

{Kernel function-RBF, $C = 10$, gamma $= 0.1$} In order to ensure the results of grid search, the three SVR methods corresponding to three kernel functions (linear, RBF and polynomial) are applied on first 30 working days of the training dataset.The performance of these three models over these data is depicted in. From this figure, it can be observed that the RBF model fits to the training data more closer than other two models. The performance of linear model is the worst among them. So, RBF model can be chosen as the best model for stock predictions for this specific dataset. The parameters related to RBF models are C and γ (Fig. 4).

Fig. 4 Comparison of linear model, polynomial model and RBF model
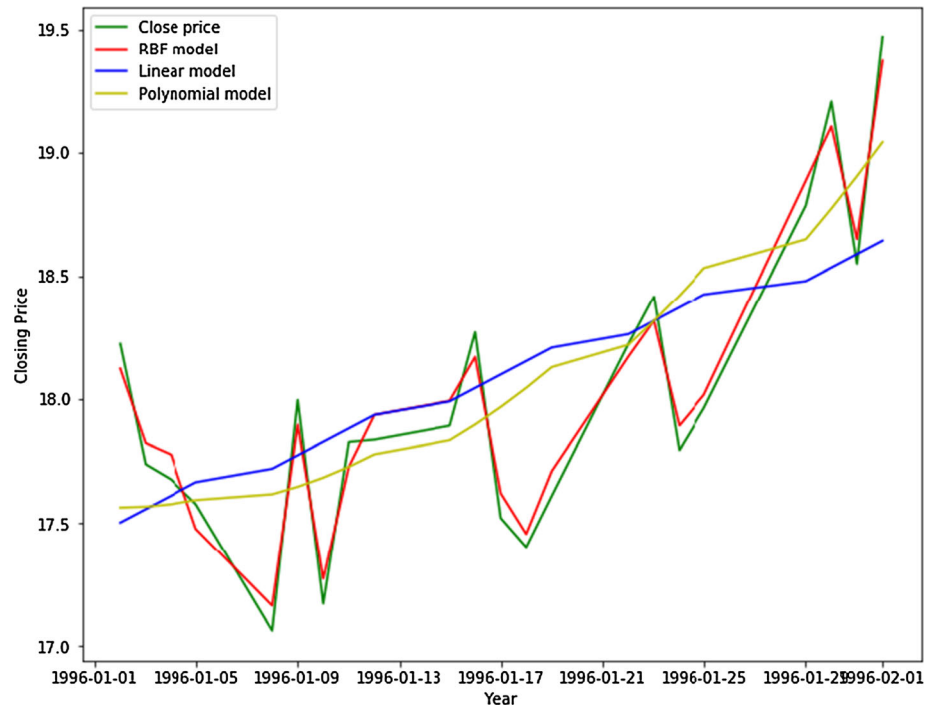
4.4 Training and validation of the support vector regression model-FTSVR

The proposed model is trained with training dataset with 4163 number of records. The actual and predicted close price is depicted in Fig. 5.

Figure 6 is a part of result obtained in Fig. 5 showing a closer look over training one year data. The error obtained during the training is shown in Fig. 7.

The overall accuracy of the proposed method is measured in terms of RMSE and MAPE. The MAPE is expressed as follows:

Fig. 5 Training of SVR model over training dataset
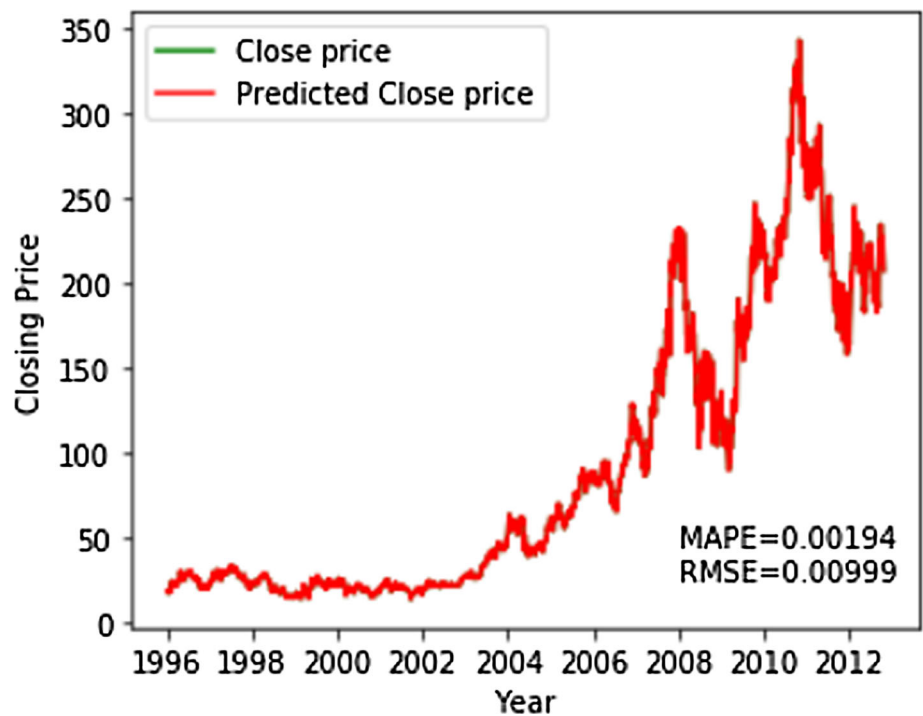
Fig. 6 Training of SVR model over training dataset (closer view)
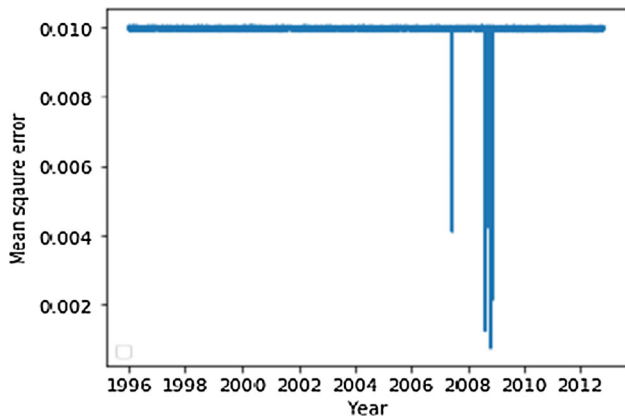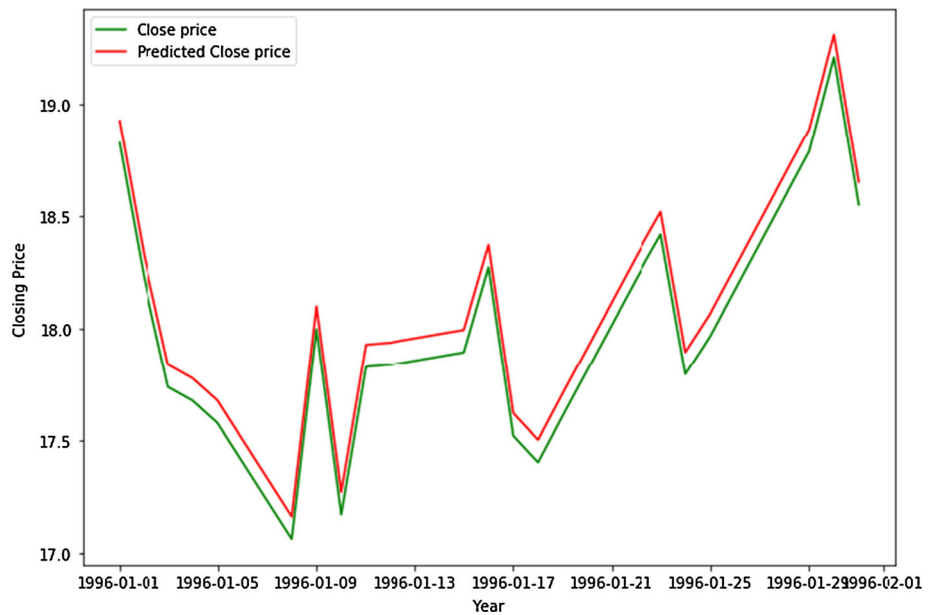
Fig. 7 Error during training of the proposed model
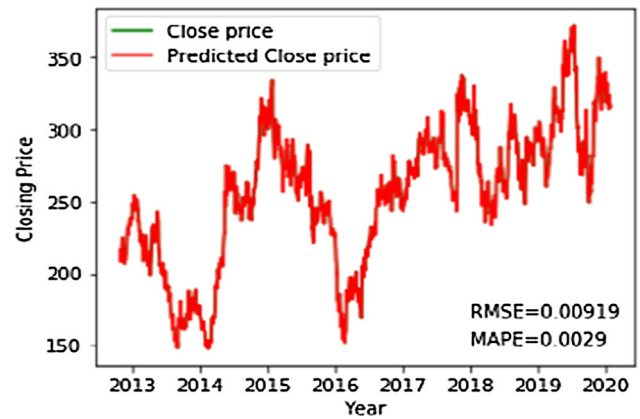
Fig. 8 Validation of SVR model over validation dataset

$$\text{MAPE} = \frac{1}{T} \sum_{i=1}^{T} \left| \frac{(t - \hat{t})}{t} \right| \qquad (15)$$

The RMSE is shown below:

$$\text{RMSE} = \sqrt{\frac{1}{T} \sum_{i=1}^{T} (t - \hat{t})^2} \mid \qquad (16)$$

where T is number of training data, t is exact closing value and $\hat{t}$ is the estimated closing price.

The computed MAPE and RMSE during training the model are 0.00188 and 0.00899, respectively.

The validation of the proposed model is performed over the validation dataset containing 1784 records. The real closing prices and predicted closing prices over this dataset are shown in Fig. 8.
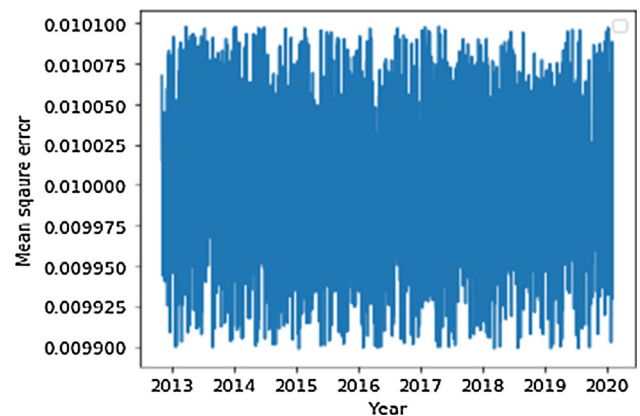
Fig. 9 Error during validation of the proposed model

The error computed on daily basis is presented in Fig. 9. The error obtained during training and validation can be compared from Figs. 7 and 8. This results show that error

during validation is always less than error during training which ensures that the proposed model is not overfitted.

The obtained MAPE and RMSE during validation of the model are 0.0029 and 0.00919, respectively.

5 Performance of the proposed model for stock analysis

The performance of the proposed model is evaluated through the following parameters:

1. Up-to-daily and up-to-monthly return prediction
2. Volatility nature of stock

5.1 Up-to-daily and Up-to-monthly return prediction

Up-to-daily or up-to-monthly return is the indicator of investors' profit or loss for one day or for a period of one month. These are the good indicators for short-term investors who can get a clear picture how the stock behaves daily or monthly. The up-to-daily (R_d) return of a stock is calculated by the following formula

$$R_d(\text{day}) = \frac{\text{Cl}(\text{day}) - \text{Cl}(\text{previous day})}{\text{Cl}((\text{previous day})} \tag{17}$$

where Cl(day) is the closing price of the day. The actual up-to-daily return and the predicted up-to-daily return are shown in Fig. 10.

Up-to-monthly return (R_M) of a stock can be calculated by using the following equation:

$$R_M(\text{month}) = \frac{\text{Cl (Last Day of Month)}}{\text{Cl(Last Day of Previous Month)}} \tag{18}$$
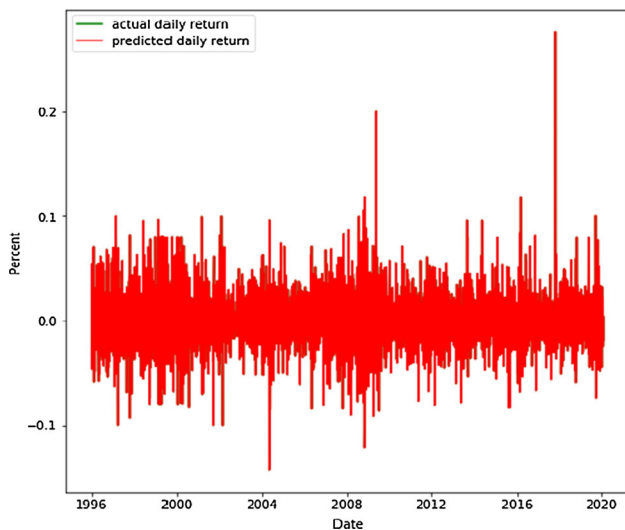
The actual up-to-monthly return and the predicted up-to-monthly return are shown in Fig. 11.

Cumulative up-to-monthly return is a measure that aggregates gain/loss amount for a certain period of time.

The cumulative up-to-monthly return (or up-to-daily return) can be calculated from up-to-monthly return(or up-to-daily return) for a certain period of time. The cumulative up-to-monthly return is depicted in Fig. 12.

Cumulative up-to-monthly return of SBIN presents the growth of Rs.1 over the entire period which indicates a return of Rs. 30 after 16 years of investment.

5.2 Prediction of volatility nature of stock

Volatility is a measure that disperses around the mean or average return of a stock. Higher degree of volatility of a stock indicates a higher probability of a declining market, while its lower value indicates to a higher probability of a rising market. The volatility nature of a stock can be predicted from the historical closing price by using the following equations

$$R_p = \log \frac{\text{Cl}_{i+1}}{\text{Cl}_i} \tag{19}$$

where R_p is the predicted return. The variance of the predicted return can be computed as

$$\text{Var} = \frac{1}{T}\sum_{i=1}^{T} R_p^2 \tag{20}$$

The predicted volatility of the stock is the standard deviation of the predicted return i.e.

$$\sigma_V = \sqrt{\frac{1}{T}\sum_{i=1}^{T} R_p^2} \tag{21}$$

The predicted volatility of SBIN is shown in Fig. 13.

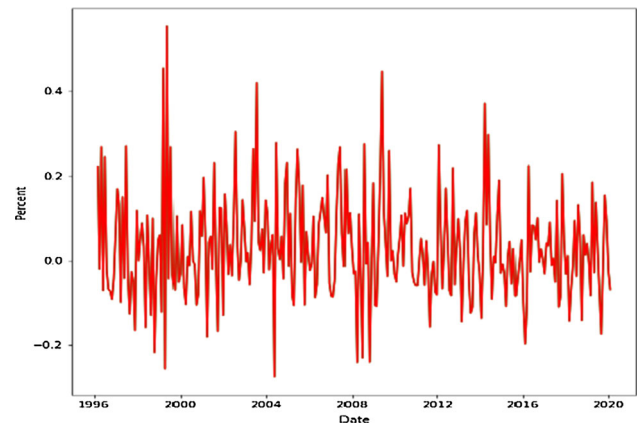Fig. 10 Up-to-daily return prediction of SBIN

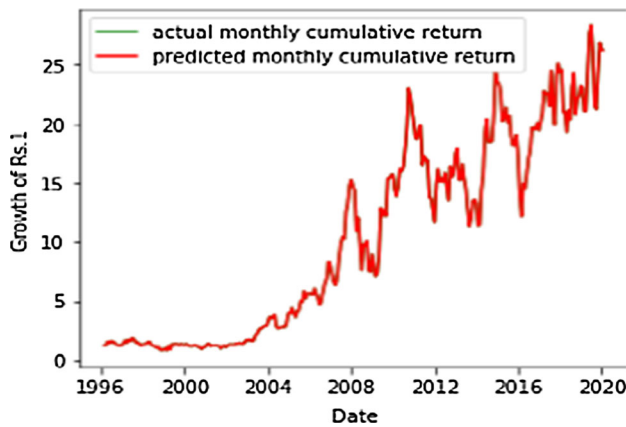Fig. 11 Up-to-monthly return prediction of SBIN

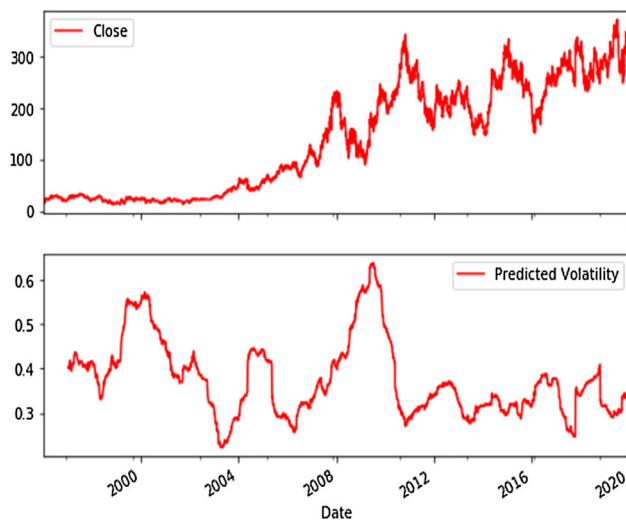Fig. 12 Cumulative up-to-monthly return prediction of SBIN

Fig. 13 Volatility return prediction of SBIN

The volatility of SBIN stock varies between 0.2 and 0.6 (refer Fig. 13). However, it is found to be almost stable after 2012. Thus, it can be suitable for the investors who prefer a long-term investment.

6 Results and discussion

This section uses data from four major sectors such as banking, steel, automobile and oil. Two datasets from each sectors containing the maximum available historical data are chosen for predicting their stock. All these datasets are NSE data drawn from (source = https://finance.yahoo.com/quote). The detail about these datasets is presented in Table 5.

The optimized parameters for each dataset are presented in Table 6. The first part of this section is devoted towards comparing the performance of this method against some existing methods while in the second part the simulated

Table 5 Datasets of eight different industries

Dataset	Duration	No. of records
SBIN Bank	01.01.1996 to 31.01.2020	5947
HDFC Bank	01.01.1996 to 31.01.2020	6077
Tata Steel	01.01.1996 to 31.01.2020	6076
Reliance Industies	01.01.1996 to 31.01.2020	6076
Maruti Suzuki	09.07.2003 to 31.01.2020	4114
Hyundai Motors	13.02.2009 to 31.01.2020	2760
Indian Oil Corp.	09.09.1995 to 31.01.2020	6179
HP Corp.	02.01.1996 to 31.01.2020	6075

Table 6 Optimized values of C and γ and radial basis function (RBF)

Dataset	Kernel function	C	γ
SBIN Bank	RBF	10	0.1
HDFC Bank	RBF	1	0.01
Tata Steel	RBF	10	0.1
Reliance Industies	RBF	1	0.1
Maruti Suzuki	RBF	10	0.1
Hyundai Motors	RBF	1	0.01
Indian Oil Corp.	RBF	1	0.001
HP Corp.	RBF	10	0.1

results for stock prediction of the above-mentioned datasets are presented and discussed. Further, the last part of this section analyses the performance of different stocks.

6.1 Comparison

The comparison is made among the existing method, SVR method [2] and hybrid NN [32] based on generated values of RMSE and MAPE for the whole dataset (see Table 7). Although SVR method [2] uses support vector regression, it is missing with optimization technique. Further, this work does not have any dedicated method to deal with missing data (SBIN has 128 number of missing closing values which is around 2% of SBIN dataset). Hybrid neural network [32] is a combined method containing the artificial neural network and particle swarm optimization.

The proposed model is compared against SVR method [2] and hybrid NN [32] in terms of time in second taken to predict the Closing price of eight different datasets (see Table 8). From here, it is obtained that proposed model with optimized parameter takes less than 1 second to predict the closing price for each dataset. SVR method [2] takes around 40 second on an average to predict the closing price while time taken by hybrid NN [32] is 450 on an

Table 7 Performance comparison of the proposed model, SVR model [2] and hybrid NN [32]

Dataset	Proposed model		SVR model [2]		Hybrid NN [32]	
	RMSE	MAPE	RMSE	MAPE	RMSE	MAPE
SBIN Bank	0.0012	0.004	0.034	0.062	0.09	0.095
HDFC Bank	0.0011	0.0041	0.035	0.06	0.091	0.01
Tata Steel	0.0012	0.00062	0.019	0.089	0.092	0.098
Reliance Industies	0.0099	0.0013	0.09	0.01	0.0919	0.095
Maruti Suzuki	0.001	0.00054	0.02	0.072	0.02	0.098
Hyundai Motors	0.0098	0.003	0.011	0.02	0.05	0.04
Indian Oil Corp.	0.0072	0.0023	0.10	0.08	0.071	0.02
HP Corp.	0.0012	0.0015	0.030	0.063	0.01	0.08

Table 8 Time (in second) comparison of the proposed model, SVR model [2] and hybrid NN [32]

Dataset	Proposed model	SVR model [2]	Hybrid NN [32]
SBIN Bank	<1	45	486
HDFC Bank	<1	47	490
Tata Steel	<1	45	479
Reliance Industies	<1	40	463
Maruti Suzuki	<1	34	401
Hyundai Motors	<1	27	320
Indian Oil Corp.	<1	41	445
HP Corp.	<1	51	476

average. Since hybrid NN [32] is a hybrid model, the time taken to converge to optimal solution increase with increase of data.

6.2 Stock prediction of different industries from their historical time varying data

Here, the proposed scheme is utilized to predict the close price of eight different datasets which is presented in Table 5. The real closing prices and the predicted closing prices by utilizing the method are illustrated in Figs. 14,
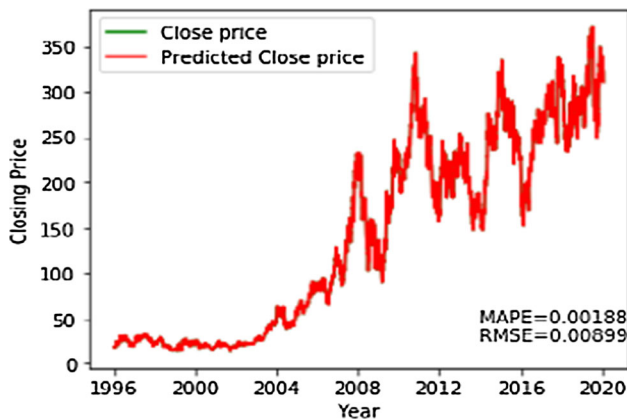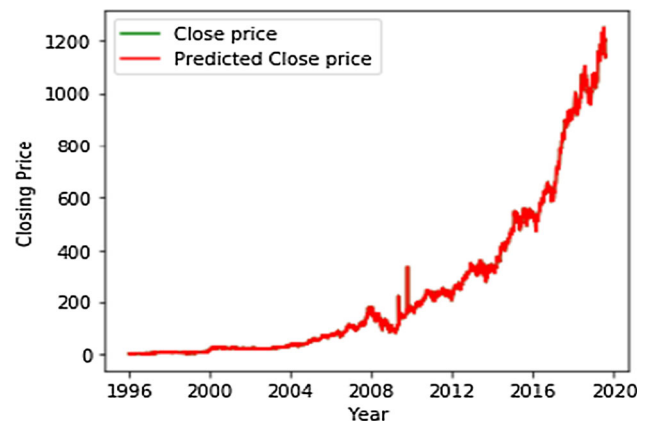

Fig. 15 Stock prediction for HDFC Bank

15, 16, 17, 18, 19, 20 and 21. The undistinguished line between real cost prices and predicted closing prices is due to the high accuracy that the proposed model yields over these dataset.

6.3 Stock analysis by using the proposed method

The proposed method is employed to predict different performance parameters of stocks of eight different industries mentioned in Table 5. The RMSE and MAPE
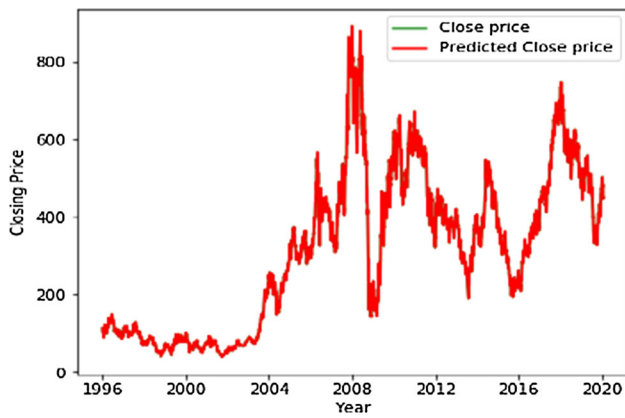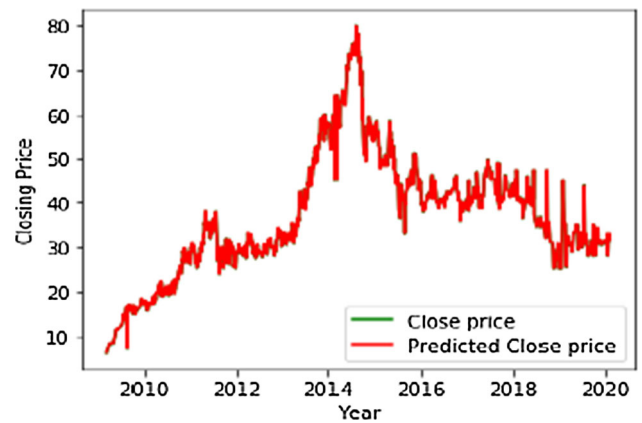

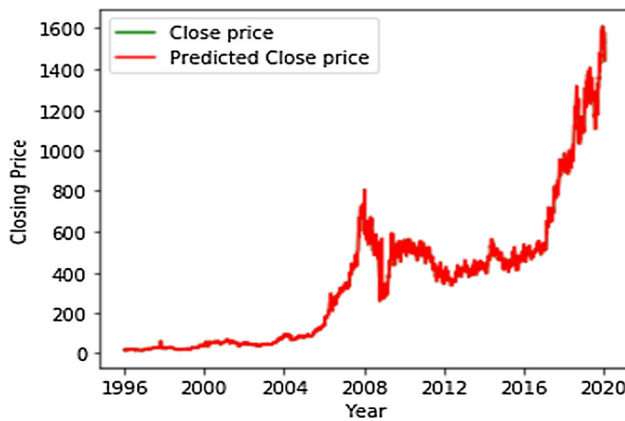**Fig. 14** Stock prediction for SBIN

Fig. 16 Stock prediction for Tata Steel

Fig. 17 Stock prediction for Reliance Industries

Fig. 18 Stock prediction for Maruti Motors

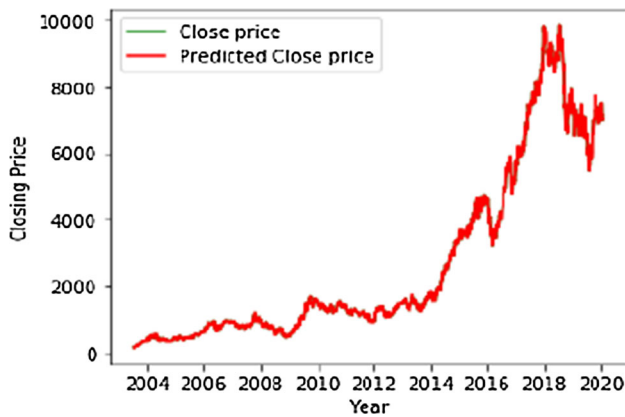**Fig. 19** Stock prediction for Hyundai Motors
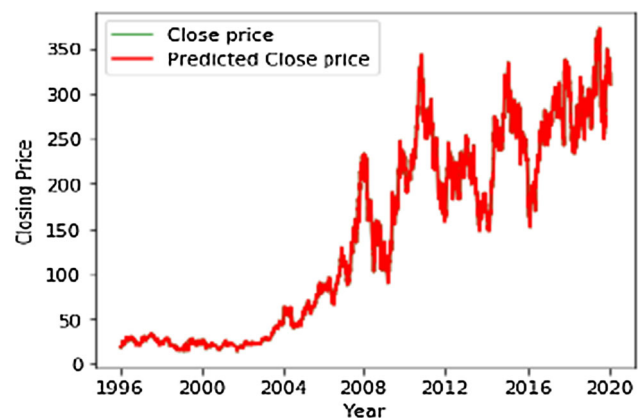

Fig. 20 Stock prediction for Indian Oil Corp

Fig. 21 Stock prediction for Hindustan Petroleum Corp

are calculated while predicting up-to-daily and up-to-monthly returns (Table 9). This table indicates a high value of predictive accuracy of the proposed method to forecast up-to-daily and up-to-monthly return of SBIN stock.

The up-to-monthly returns of eight different industries is predicted and compared to analyse their behaviour over the entire period (Fig. 22). However, it is quite difficult to

analyse the performance of the stocks by means of up-to-monthly returns. Hence, to get a better analysis, the commutative up-to-monthly return must be predicted and compared (Fig. 23). This figure shows a clear picture of their behaviour for the specific period of time. HDFC bank has the best return since 2000. The return of Reliance is far away from HDFC bank, and it comes in second. The third

Table 9 Performance comparison of the proposed model to predict up-to-daily and up-to-monthly return

Dataset	Up-to-daily return		Up-to-monthly return	
	RMSE	MAPE	RMSE	MAPE
SBIN Bank	0.00872	0.0098	0.0097	0.00078
HDFC Bank	0.0071	0.0041	0.0065	0.0032
Tata Steel	0.0092	0.0012	0.002	0.0089
Reliance Industies	0.0013	0.0078	0.009	0.0065
Maruti Suzuki	0.0011	0.0094	0.002	0.0022
Hyundai Motors	0.0029	0.0065	0.0067	0.0012
Indian Oil Corp.	0.0066	0.0011	0.0091	0.0081
HP Corp.	0.0010	0.0085	0.0010	0.0067

position of this list is occupied by Maruti while the performance of others are not appreciable.

The volatility of stocks of two banks, viz. SBI and HDFC, are predicted and compared (Fig. 23). From this figure, it can be observed that both the banks perform competitively till 2012. However, the stock of HDFC bank sharply decreases beyond 2012 and later varies in between 0 and 0.1. In contrast to this, SBI has a good volatility value with a variation between 0.25 and 0.45. Thus, short-term investor must invest in the stock of SBI to have profitable return (Fig. 24).

In order to study the risk associated in oil industries, the volatility of Indian Oil and HP are predicted and compared (Fig. 25). The volatility of this two stocks are mostly competitive with few exceptional growth of Indian oil. The

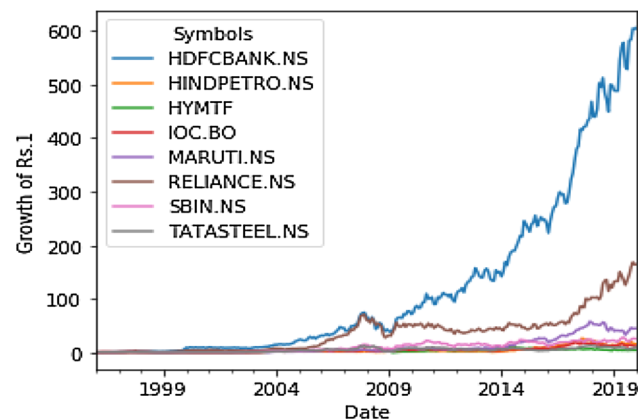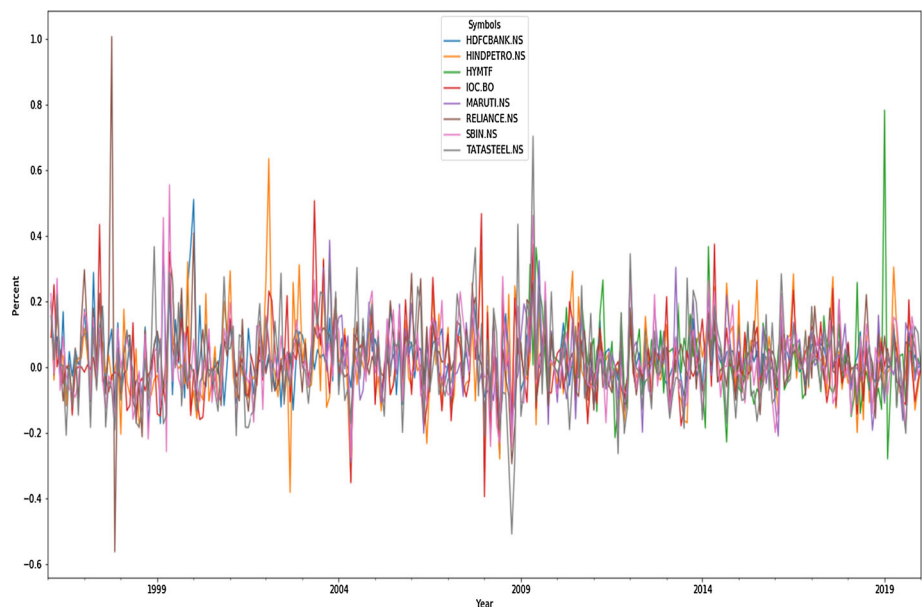**Fig. 22** Comparison of up-to-monthly return of different stocks

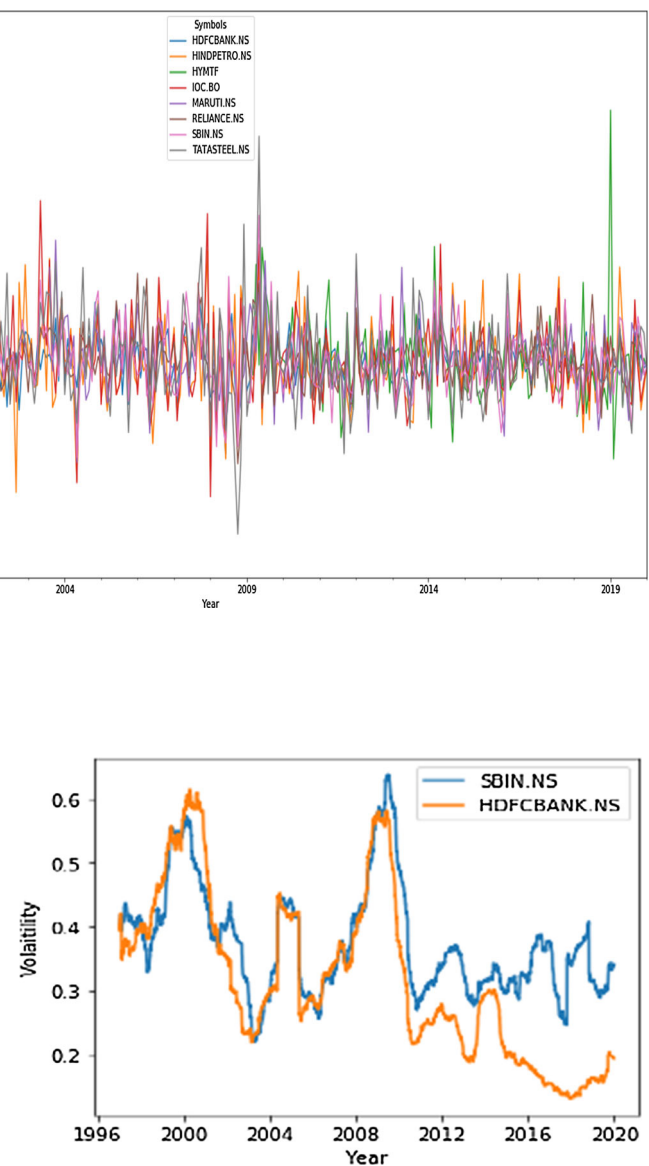**Fig. 23** Comparison of commutative up-to-monthly return of different stocks

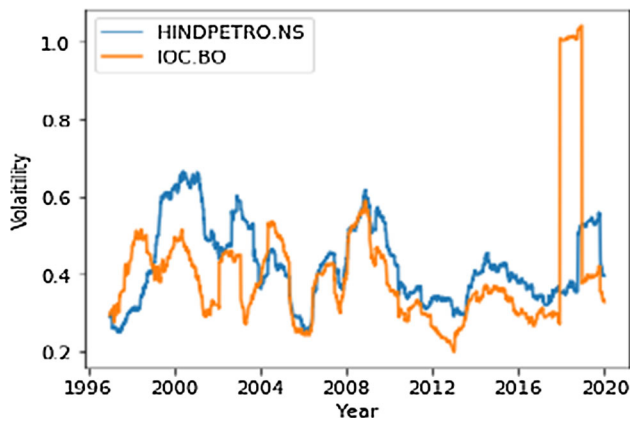**Fig. 24** Comparison of volatility of SBI and HDFC bank

Fig. 25 Comparison of volatility of Indian oil and HP

investors may choose any of these stock for their investment with almost similar risk factors.

The volatility study of automobile sector reveals Hyundai motors to be better choice for investment than Maruti Suzuki (Fig. 26).

7 Conclusion

The proposed method in this study aims to use a new fine-tuned support vector regression model for predicting the stock price. A suitable dataset is taken to explain each step of this method. Missing attributes are well handled by using appropriate techniques. Grid search is performed on the entire training dataset to choose the kernel function and its optimized parameters. The training of the proposed method is performed over 4163 records while validation is performed by using 1800 records. RMSE and MAPE are computed for training and validation, respectively. The proposed method is compared against some existing methods of similar interest in terms of computed RMSE, MAPE and time. The results from this comparison ensure
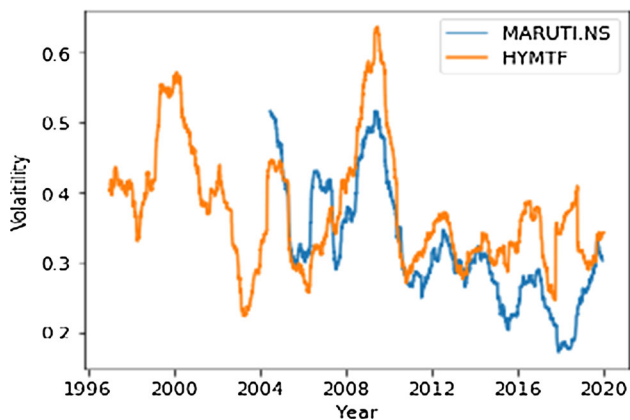
proposed method to be more accurate and time efficient method than its counterpart method. Eight different datasets from four different sectors are considered for their stock prediction. The results show the prediction values are closely fitted with the actual values. The proposed method can further be used to predict the stock of any industry with greater accuracy.

Declaration

Conflict of interest The authors declare that there is no conflict of interest.

References

1. Malkiel BG (2003) The efficient market hypothesis and its critics. J Econ Perspect 17(1):59–82. https://doi.org/10.1257/089533003321164958
2. Henrique BM, Sobreiro VA, Herbert K (2018) Stock price prediction using support vector regression on daily and up to the minute prices. J Finance Data Sci 4(3):183–201. https://doi.org/10.1016/j.jfds.2018.04.003
3. Preeti N, Tu NN, Shivani A, Jude HD (2021) Capsule and attention layer augmented convolutional methods for review categorization in Amazon dataset. Comput Electr Eng (Early Access)
4. Nguyen GL, Dumba B, Ngoc Q-D, Le H-V, Tu NN (2021) A collaborative approach to early detection of IoT Botnet. Comput Electr Eng (Early Access)
5. Nguyen CH, Pham TL, Nguyen TN, Ho CH, Nguyen TA (2021) The linguistic summarization and the interpretability, scalability of fuzzy representations of multilevel semantic structures of word-domains. Microprocess Microsyst 81:103641
6. Pham DV, Nguyen GL, Nguyen TN, Pham CV, Nguyen AV (2021) Multi-topic misinformation blocking with budget constraint on online social networks. IEEE Access 8:78879–78889
7. Le NT, Wang J, Le DH, Wang C, Nguyen TN (2021) Fingerprint enhancement based on tensor of wavelet subbands for classification. IEEE Access 8:6602–6615
8. Le NT, Wang J, Wang C, Nguyen TN (2019) Novel framework based on HOSVD for Ski goggles defect detection and classification. Sensors 19:5538
9. Le NT, Wang J, Wang C, Nguyen TN (2019) Automatic defect inspection for coated eyeglass based on symmetrized energy analysis of color channels. Symmetry 11:1518
10. Duc-Ly V, Trong-Kha N, Nguyen Tam V, Nguyen TN, Fabio M, PhuH P (2020) HIT4Mal: Hybrid image transformation for malware classification. Trans Emerg Telecommun Technol 31:e3789
11. Vu D, Nguyen T, Nguyen TV, Nguyen TN, Massacci F, Phung PH (2019) A convolutional transformation network for malware classification. In: 2019 6th NAFOSTED conference on information and computer science (NICS), pp 234–239
12. Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50(1):159–175. https://doi.org/10.1016/S0925-2312(01)00702-0
13. Hamzaebi C, Akay D, Kutay F (2009) Comparison of direct and iterative artificial neural network forecast approaches in multi-periodic time series forecasting. Expert Syst Appl 36(2):3839–3844. https://doi.org/10.1016/j.eswa.2008.02.042

Fig. 26 Stock prediction for Indian Oil Corp

14. Guresen E, Kayakutlu G, Daim TU (2011) Using artificial neural network models in stock market index prediction. Expert Syst Appl 38(8):10389–10397. https://doi.org/10.1016/j.eswa.2011.02.068

15. Wang JZ, Wang JJ, Zhang ZG, Guo SP (2011) Forecasting stock indices with back propagation neural network. Expert Syst Appl 38(11):14346–14355. https://doi.org/10.1016/j.eswa.2011.04.222

16. Chang P, Wang D, Zhou CA (2012) Novel model by evolving partially connected neural network for stock price trend forecasting. Expert Syst Appl 39(1):611–620. https://doi.org/10.1016/j.eswa.2011.07.051

17. Jo RDA, Ferrira TAE (2013) A morphological rank linear evolutionary method for stock market prediction. Inf Sci 237(1):3–17. https://doi.org/10.1016/j.ins.2009.07.007

18. Hegazy O, Soliman OS, Salam MA (2014) A Machine Learning Model for Stock Market Prediction. arXiv preprint arXiv:1402.7351

19. Bagheri A, Peyhani HM, Akbari M (2014) Financial forecasting using ANFIS network with quantum behaved particle swarm optimization. Expert Syst Appl 41(14):6235–6250. https://doi.org/10.1016/j.eswa.2014.04.003

20. Rather AM, Agarwal A, Sastry VN (2015) Recurrent neural network and a hybrid model for prediction of stock returns. Expert Syst Appl 42(6):3234–3241. https://doi.org/10.1016/j.eswa.2014.12.003

21. Majhi B, Anish CM (2015) Multi objective optimization based adaptive models with fuzzy making for stock market forecasting. Neuro Comput 167(2015):502–511. https://doi.org/10.1016/j.neucom.2015.04.044

22. Rout AK, Dash PK, Dash R, Bisoi R (2015) Forecasting financial time series using a low complexity recurrent neural network and evolutionary learning approach. J King Saud Univ Comput Inf Sci 29(4):536–552. https://doi.org/10.1016/j.jksuci.2015.06.002

23. Chen MY, Chen BT (2015) A hybrid fuzzy time series model based on granular computing for stock price forecasting. Inf Sci 294:227–241. https://doi.org/10.1016/j.ins.2014.09.038

24. Nayak RK, Mishra D, Rath AK (2015) A Naive SVM-KNN based stock market trend reversal analysis for Indian benchmark indices. Appl Soft Comput 35(1):670–680. https://doi.org/10.1016/j.asoc.2015.06.040

25. Jia H Investigation into the effectiveness of long short term memory networks for stock price prediction. arXiv preprint arXiv:1603.07893

26. Qu H, Zhang Y (2016) A new kernel of support vector regression for forecasting high-frequency stock returns. Math Probl Eng 1:1–9. https://doi.org/10.1155/2016/4907654

27. Kumar D, Meghwani SS, Thakur M (2016) Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets. J Comput Sci 17(1):1–13. https://doi.org/10.1016/j.jocs.2016.07.006

28. Moghaddam AH, Moghaddam MH, Esfandyari M (2016) Stock market index prediction using artificial neural network. J Econ Finance Adm Sci 21(41):89–93. https://doi.org/10.1016/j.jefas.2016.07.002

29. Kumar DP, Ravi V (2017) Forecasting financial time series volatility using particle swarm optimization trained quantile regression neural network. Appl Soft Comput 58(1):35–52. https://doi.org/10.1016/j.asoc.2017.04.014

30. Mehdi K, Zahra H (2017) Performance evaluation of series and parallel strategies for financial time series forecasting. Financ Innov 3(24):1–15. https://doi.org/10.1186/s40854-017-0074-9

31. Lei L (2018) Wavelet neural network prediction method of stock price trend based on rough set attribute reduction. Appl Soft Comput 62:923–932. https://doi.org/10.1016/j.asoc.2017.09.029

32. Senapati MR, Das S, Mishra S (2018) A novel model for stock prediction using hybrid neural network. J Inst Eng India Ser B 99:555–563. https://doi.org/10.1007/s40031-018-0343-7

33. Selvamuthu D, Kumar V, Mishra A (2019) Indian stock market prediction using artificial neural networks on tick data. Financ Innov. https://doi.org/10.1186/s40854-019-0131-7

34. Syarif I, Prugel-Bennett A, Wills G (2012) Data mining approaches for network intrusion detection: from dimensionality reduction to misuse and anomaly detection. J Inf Technol Rev 3(2):70–83