Soft Computing
https://doi.org/10.1007/s00500-020-04957-x

METHODOLOGIES AND APPLICATION

Time series data analysis of stock price movement using machine learning techniques

Irfan Ramzan Parray[1] · Surinder Singh Khurana[1] · Munish Kumar[2] · Ali A. Altalbe[3]

Abstract

Stock market also called as equity market is the *aggregation* of the sellers and buyers. It is concerned with the domain where the shares of various public listed companies are *traded.* For predicting the growth of economy, stock market acts as an index. Due to the nonlinear nature, the prediction of the stock market becomes a difficult task. But the application of various machine learning techniques has been becoming a powerful source for the prediction. These techniques employ historical data of the stocks for the training of machine learning algorithms and help in predicting their future behavior. The three machine learning algorithms used in this paper are support vector machine, perceptron, and logistic regression, for predicting the next day trend of the stocks. For the experiment, dataset from about fifty stocks of Indian National Stock Exchange's NIFTY 50 index was taken, by collecting stock data from January 1, 2013, to December 31, 2018, and lastly by the calculation of some technical indicators. It is reported that the average accuracy for the prediction of the trend of fifty stocks obtained by support vector machine is 87.35%, perceptron is 75.88%, and logistic regression is 86.98%. Since the stock data are time series data, another dataset is prepared by reorganizing previous dataset into the supervised learning format which improves the accuracy of the prediction process which reported the results with support vector machine of 89.93%, perceptron of 76.68%, and logistic regression of 89.93%, respectively.

Keywords Stock market · Machine learning · Support vector machine · Artificial neural network · Logistic regression · Technical indicators

1 Introduction

The stock market is very important for investors and traders of a country. A stock market can attract investment that is essential for economic growth. In the past, the research on stock market prediction was pursued upon efficient market hypothesis (EMH) and random walk theory (Weng et al. 2017). These models suggest that stock price or stock price movement cannot be predicted since they are driven by new information like news rather than past/present prices. The new information refers to the qualitative dimensions that impact the stock market, which, for instance, may involve the political environment of a country. Thus, a stock market is driven by a lot of factors.

In machine learning, the computers are made to learn without being explicitly programmed. The code is not written for every specific problem; instead, data are provided to a machine learning algorithm and subsequently logic is developed based on those data. There are two types of machine learning algorithms.

Communicated by V. Loia.

✉ Munish Kumar
munishcse@gmail.com

Irfan Ramzan Parray
parrayirfan9@gmail.com

Surinder Singh Khurana
surinder.seeker@gmail.com

Ali A. Altalbe
aaltalbi@kau.edu.sa

[1] Department of Computer Science and Technology, Central University of Punjab, Bathinda, India

[2] Department of Computational Sciences, Maharaja Ranjit Singh Punjab Technical University, Bathinda, India

[3] Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, Saudi Arabia

Published online: 22 April 2020

- Supervised learning
- Unsupervised learning

In supervised learning algorithm, the dataset used for training the algorithm is labeled. An algorithm is used to learn a mapping function that maps input variables onto the output variables. The aim of this learning process is to approximate the mapping function in such a way that for given input data the mapping function should predict the output data. It is important to note that the incorrect labeling in the dataset can negatively impact the effectiveness of the machine learning model. Some commonly used supervised machine learning algorithms include artificial neural networks, support vector machines, logistic regression, Naive Bayes, and random forest. In unsupervised learning, the dataset is not labeled, hence difficult to interpret. Therefore, the algorithm uses the training data to generate a structure by looking at relations in training data itself. Some of the commonly used unsupervised learning algorithms include *k*-means clustering, *k* nearest neighbors, and auto-encoders. Supervised machine learning algorithms like a support vector machine (SVM) have been widely used to predict the stock market behavior.

The prediction in the stock market was traditionally driven by technical indicators. Technical indicators are the mathematical formulas which, when applied to historical data, give a sense of how the stock is going to behave in the future. With the emergence of technologies like machine learning, the prediction of stock market indices has improved significantly. The machine learning algorithms consider the past prices of the stocks and train the model for future predictions of stock prices or stock price movement. A lot of research has been done to predict the stock market behavior using machine learning. In the article, two machine learning algorithms, namely SVM and artificial neural network (ANN), have been considered to predict the stock market behavior since these algorithms have been quite successful.

1.1 Technical indicators

A technical indicator is defined as a mathematical formula which, when applied to stock's historical trading data, yields a forecasting trend as to how that stock will trade in the future. The technical indicators that are considered in this research are given below:

- *Exponential moving average (EMA)* This is a kind of weighted moving average that gives more importance or weightage to recent price data than a simple moving average does. EMA can be calculated by any number of days. This work used the 12-day EMA and 26-day EMA. The EMA of a stock is calculated by the following formula:

$$\mathrm{EMA} = \mathrm{Price}(t) \times \mathrm{EMA}(y) \times (1 - k) \tag{1}$$

where t is today, y is yesterday, and N is the number of days in EMA

$k = 2/(N + 1)$.

In this work, N is taken as 12 and 26. Hence, the two EMAs calculated are EMA 12 and EMA 26.

- *Moving average convergence divergence (MACD)* MACD is the simplest technical tool used by analysts to predict stock trends. This technique has gained popularity among traders for its reliability in predicting the broad direction or state of the market, though it does not provide exact entry or exit points as other indicators but gives the direction of the stock quite consistently. MACD itself is constructed by subtracting the long-term moving average of the stock from short-term moving average. Typically, the long term is taken as recent 26 days and the short term is taken as recent 12 days. The short-term moving average is obviously more responsive than its longer sibling. The formula for calculation of MACD is given below:

$$\mathrm{MACD} = 12\,\mathrm{day\,EMA} - 26\,\mathrm{day\,EMA} \tag{2}$$

- *Relative strength index (RSI)* It is an oscillator-type technical indicator which compares the magnitude of recent gains to recent losses to determine overbought and oversold conditions of an asset (Weng et al. 2017). RSI ranges from 0 to 100. In practice, investors sell if its value is ≥ 80 and buy if it is ≤ 20. Generally, the RSI is calculated over a 14-day period. The formula for calculation of RSI is as follows:

$$\mathrm{RSI} = 100 - \frac{100}{1 + \mathrm{RS}} \tag{3}$$

where RS = (average gain/average loss), average gain = (total gains/n), average loss = (average losses/n), n = 14.

- *Average true range (ATR)* It is a technical indicator that gives the measure of the volatility of a stock. It does not give any account of the direction of the stock price. The importance of ATR is that it measures volatility based on day-to-day closing prices as well as intraday high–low prices. The formula for calculating ATR is as follows:

$$\mathrm{ATR} = \frac{1}{n}\sum_{i=1}^{n} \mathrm{TR}_i \tag{4}$$

where TR refers to the true range and is calculated as:

$$\mathrm{TR} = \max\left[(\mathrm{high} - \mathrm{low}), \mathrm{abs}\left(\mathrm{high} - \mathrm{close}_{\mathrm{previous}}\right), \mathrm{abs}\left(\mathrm{low} - \mathrm{close}_{\mathrm{previous}}\right)\right]$$

n is the time period over which the TR is calculated (usually taken as 10 days)

- *The open* When the market is open for trading, the first price at which a trade executes is called the opening price. It is an important indicator that reflects the developments that have taken place between previous day's closing price and the current day's opening of the market. It also reflects the opening price trend.
- *The high* This represents the highest price at which the market participants were willing to transact for the given day. It gives the estimate of how good a stock is doing in a day. This indicator can be used to compare a stock with other stocks.
- *The low* This represents the lowest level at which the market participants were willing to transact for the given day.
- *The close* The close price is the most important price because it is the final price at which the market closed for a period. The close serves as an indicator for the intraday strength. If the close is higher than the open, then it is considered a positive day, else negative.
- *Previous close* For a day, previous close is simply the price at which the stocks closed yesterday.

2 Literature review

Over the years, a lot of research is conducted in the field of stock market behavior prediction. The research work that has been done in this field was reviewed before implementing this work. Support vector machine (SVM) is one of the most widely used supervised machine learning algorithms. SVM is used for both classification and regression. The regression version of SVM is called support vector regression (SVR). SVM classifies the data by demonstrating the linear separability using hyperplane. SVM constructs an optimal separating hyperplane in the high dimension feature space. Madge and Bhatt (2015) attempt to predict the stock price using SVM. It takes 34 technology stocks in its work. It aims to predict a given stock's behavior at time $t + m$ by using the SVM at time t. It takes four parameters (technical indicators) into consideration, viz. the momentum and the recent price volatility of the individual stock and the technology sector. The work concludes that the short-term predictions have very low accuracy, but in long-term prediction it reaches prediction accuracies between 55 and 60%. This prediction accuracy suggests that the EMH does not necessarily hold for the long-term forecasting. This is indicative of the fact that machine learning algorithms can use the influence of various stock market technical indicators for forecasting of stock prices or stock price movement. There are various improvements that can be done to the simple SVM algorithm in order to enhance the performance of the algorithm.

For instance, Wang and Shang (2014) have used least square support vector machine (LSSVM) for predicting the daily stock movement direction of stock market index. The work has been conducted on stocks from China Security Index 300b (CSI 300). The results are compared with artificial intelligence (AI) model, probabilistic neural network (PNN), and two discriminant analysis models. Experimental results reveal that LSSVM outperforms PNN, linear discriminant analysis (LDA), and quadratic discriminant analysis (QDA) in terms of accuracy. Since the SVM training is a time-consuming process, LSSVM is used to overcome this shortcoming. This work uses ten technical indicators. The number of technical indicators is very high. Some technical indicators give a better account of the future of the stock, and some are not that effective. Hence, a combination of best technical indicators is to be found to predict the future of a stock. Feature extraction using principal component analysis (PCA) is performed by Nahil and Lyhyaoui (2018) in order to select the best parameters (technical indicators). The work concludes that when PCA is used with SVM, it can achieve better generalization performance than without application of PCA. With the help of PCA, an optimal subset of all attributes can be achieved, and redundant and irrelevant attributes can be deleted. The historical stock data are generally nonlinear and non-stationary time series data. Since it becomes difficult to analyze such data, Fenghua et al. (2014) use singular spectrum analysis (SSA) to analyze such data. It decomposes the time series data into different components and interprets those components individually. Fenghua et al. (2014) break down stock price in terms of market fluctuation due to different reasons, a trend followed by the stock price, and the noise. These individual features guide SVM to make more accurate predictions. This work concludes that the SSA–SVM combination has better accuracy in predicting stock price behavior than simple SVM. Now, SVM gives the direction of stock price movement that can be either positive or negative. To predict the price of the stocks, support vector regression (SVR) is used. Henrique et al. (2018) use SVR to predict the stock prices for different capitalizations and different markets. The results obtained suggest that SVR has a significant predictive power.

Artificial neural network (ANN) is a computational system inspired by the structure, processing method, and learning ability of the human brain. It contains many processing elements like neurons. ANN acquires knowledge through the learning process. A learning algorithm is used for altering the network weights in order to find a set of weight matrices that can map the input onto the correct output. It uses different functions for its training called as activation function. Here again, the learning process is of two types: supervised learning and unsupervised learning. ANNs have also played a significant role in stock market prediction research. The ANN can be trained according to the historical stock data and then used for the prediction of stock prices or stock price movements. Moghaddam et al. (2016) use ANN for forecasting of stock price movement. It uses different learning models such as gradient decent with adaptive learning (GDA), Levenberg–Marquardt (LM), and one-step secant (OSS) for training and developing constructed models. This work has used a variable number of neurons in hidden layers and concluded that prediction accuracy is affected by the neuron structure in the neural network. The training process in ANN is performed using different learning techniques such as gradient decent or conjugate gradient. These learning techniques use a first-order derivative of performance index to optimize performance weights. Moghaddam et al. (2016) conclude that the LM algorithm that uses a second-order derivative of performance index performs better than gradient decent or conjugate decent algorithms. Increased number of hidden layers in the ANN given generate deep neural network (DNN). The hidden layers in the DNN do increase the complexity of the network, but it also helps in training the neural network for complex patterns to give better performance. DNN used by Lachiheb and Gouider (2018) shows considerable improvement in the accuracy of prediction. It achieves an accuracy of 71%. Shi et al. (2018) present a system that interprets text-based deep learning models visually for the prediction of stock price movements. This work focuses on financial analytics domain. It works on interpretation of text that includes news and other information regarding the stocks. Deep learning models are also used by Balaji et al. (2018) for stock price forecasting. It uses fourteen different types of learning models which use different learning techniques like long short-term memory (LSTM), extreme learning machines (ELMs), gated recurring unit (GRU), and convolution neural network (CNN). It is observed that GRU-based models give better directional accuracy (DA) than other modes.

The proposed systems can achieve at a maximum of 71.95% accuracy for the prediction of stock price movements. It has been analyzed that the performance of a machine learning algorithm may vary from one stock market to another. Hence, it is not suitable to use the same machine learning model for different stock markets. A hybrid forecasting model can resolve these issues. Therefore, Nayak et al. (2017) develop a model that considers stock data of different stock markets. It uses the neural network whose parameters are optimized by a natural chemical reaction-inspired metaheuristic. In this model, the whole dataset is not divided into training sets and testing set; rather, it uses a sliding window technique to select the training pattern for the network. The proposed model is called artificial chemical reaction neural network (ACRNN). The artificial chemical reaction optimization (ARCO) is an evolutionary optimization technique inspired by the nature of chemical reactions (Nayak et al. 2017). Dash and Dash (2016) developed a system called computational efficient functional link artificial neural network (CEFLANN) for generating the trading decisions more effectively. The novelty of this approach is that it integrates the learning ability of CEFLANN with technical analysis rules that give rise to profitable stock trading decision making. The performance of CEFLANN is compared with *k* nearest neighbor (KNN), support vector machine (SVM), and decision tree (DT), and it is observed that CEFLANN provides superior profit percentage. Different types of ANNs give different performance regarding the prediction of stock market behavior. It is observed that combination of an input layer, hidden layer, and output layer does impact the performance of ANN. A right combination of layers is to be found to get better performance. Deng et al. (2017) have proposed an efficient multi-objective optimization model of gate assignment problem. Deng et al. (2019) have proposed an improved ant colony optimization (ICMPACO) algorithm. Their experiment results show that the proposed ICMPACO algorithm can effectively obtain the best optimization value in solving TSP, effectively solve the gate assignment problem, and obtain better assignment result, and it takes on better optimization ability and stability. Zhao et al. (2019) have proposed a new performance degradation prediction (HMEPEM) method based on high-order differential mathematical morphology gradient spectrum entropy (HOMMSE) and phase space reconstruction, and extreme learning machine (ELM) is proposed to predict performance degradation trend of rolling bearings. Zhao et al. (2020) have proposed a semi-supervised broad learning system (SS-BLS). Firstly, the features are extracted from labeled and unlabeled data by building feature nodes and enhancement nodes. Their experiment result shows that the SS-BLS can achieve higher classification accuracy for different complex data and takes on fast operation speed and strong generalization ability.

3 Research methodology

This work proposes a stock market price movement prediction technique with improved accuracy than the techniques used in the existing literature. The proposed techniques use different parameters (technical indicators) for the prediction of the stock price movement. Different machine learning classifiers are used. The flowchart of the proposed technique is described in Fig. 1.

3.1 Data collection and preparation

The data collected in this work are downloaded from nseindia.com about NIFTY 50 stocks of National Stock Exchange shown in Table 1. The stock data are downloaded in the form of csv files.

The data of 6 years have been used from January 1, 2013, up to December 31, 2018. The downloaded data only contain the open, close, high, and low indicators among those that are to be used in this research. Hence, programs are to be written to calculate the other technical indicators in order to maximize the accuracy of the prediction of stock market trend. The technical indicators that are calculated are exponential moving average (EMA), moving average convergence divergence (MACD), average true range (ATR), and relative strength index (RSI). These technical indicators have been described in the Introduction section in detail. These indicators are calculated based on the closing price of the data; the closing price gives the consolidated view of the stock price on a day.

3.2 Time series modeling

The time series is defined as an ordered sequence of data points of a variable at equally spaced time intervals. Considering the time series nature of the stock data, closing price of the stocks can be converted into the supervised machine learning format. If the trend of the closing price of a stock on day 't' is to be forecasted, the data of the previous N days along with other technical indicators are used to train the machine learning to predict the trend. In this work, we take N as 3, i.e., all the technical indicators are calculated for the previous three days ($t-3$, $t-2$, $t-1$) and used to forecast the trend of the day 't'.

3.3 Experimental setup and workflow

The programming language that was used in this work is Python. The machine learning algorithms that were used are SVM, logistic regression, and perceptron neural network. The machine learning models were created and evaluated using two different techniques.

The first one uses tenfold cross-validation, and the second one uses train- test split strategy. In tenfold cross-validation, the dataset is divided into ten different parts. Nine parts are used for training, and one part is used for testing. The process is repeated till the testing is done in each part. It provides an efficient way for machine learning algorithm. The training–testing split is simple; the dataset is divided into two parts, i.e., training part and testing part. 70% of the data are used for training, and 30% data are used for testing. Figure 2 shows the experimental workflow.

3.4 Algorithms used

In the reviewed literature, various algorithms have been used to predict the stock market trends. Application of machine learning algorithms to determine the behavior of the stock market has been a popular subject recently. This work used logistic regression, SVM, and perceptron model of neural network to predict the stock market trends. Logistic regression is a machine learning technique which has its roots in the field of statistics. It has proved to be a very good algorithm for binary classification problems. The function that is used at the core of this algorithm is logistic function, and that is the reason it is called the logistic regression. This is also called sigmoid function. Support vector machines (SVMs) are supervised algorithms used in machine learning. They are used in classification problems. Support vector machines also have a regression version called support vector regression (SVR). SVMs are also called the support vector networks. It classifies the data by demonstrating linear separability in higher dimensions by using hyperplanes. When used for classification of data, SVM constructs an optimal separating hyperplane in the high-dimensional feature space. In this way, it works like a maximal marginal classifier. When used for regression, SVM classifier performs linear regression in the high-dimensional vector space. Perceptron can be simply described as a single-layer neural network. It can be used as a binary classifier with supervised learning.

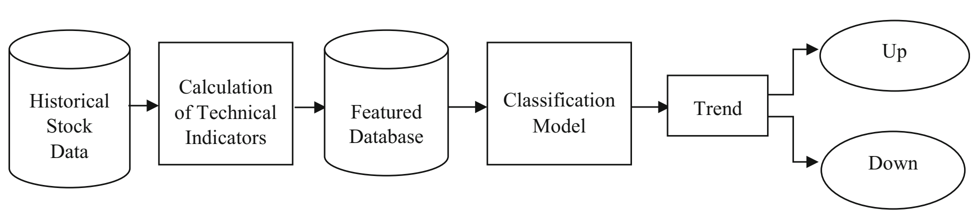

Fig. 1 Flowchart of the proposed technique

Table 1 NIFTY 50 stocks

ADANIPORTS	ASIANPAINT	AXISBANK	BAJAJ-AUTO	BAJAJFINSV
BAJFINANCE	BHARTIARTL	BIL	BPCL	CIPLA
COALINDIA	DRREDDY	EICHERMOT	GAIL	GRASIM
HCLTECH	HDFC	HDFCBANK	HINDALCO	HINDPETRO
HINDUNILVR	IBULHSGFIN	ICICIBANK	INDUSINDBK	INFRATEL
INFY	IOC	ITC	JSWSTEEL	KOTAKBANK
LT	M&M	MARUTI	NTPC	ONGC
POWERGRID	RELIANCE	SBIN	SUNPHARMA	TATAMOTORS
TATASTEEL	TCS	TECHM	TITAN	UBL
ULTRACEMCO	VEDL	WIPRO	YESBANK	ZEEL

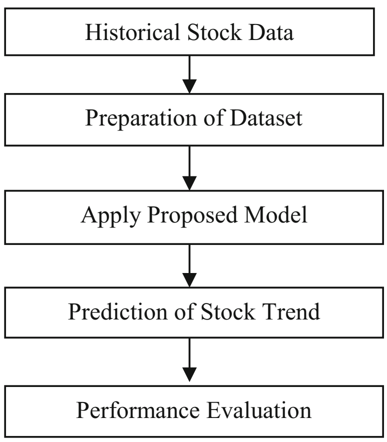

Fig. 2 Workflow

4 Results and discussion

This section mentions the results obtained from the experiments conducted during the implementation of this work. The results are followed by the discussion which involves the analysis of the obtained results. The objective is to predict the stock market trend using machine learning algorithms and then to compare the performance of these algorithms.

4.1 Results based on support vector machine

The first experiment conducted was to implement the support vector machine (SVM) on the dataset. The SVM was implemented using k-fold cross-validation method with ten splits for each stock's dataset. The performance parameters are calculated by taking the mean of the ten test splits. SVM was also implemented by dividing the data into training and test splits. 70% of the data were used as training data, and 30% of the data were used for testing. The third type of implementation is done on the data after conversion into time series supervised learning format. The average of the performance parameters that were calculated for fifty stocks is shown in Table 2.

In non-time series supervised learning data, comparing k-fold cross-validation method with the train–test method of SVM implementation, it can be observed that the performance parameters do not vary with any degree of significance. This also indicates the fact that issues like overfitting and underfitting are not a cause of concern for the datasets that are being used in this work. However, the time series supervised learning data show more than 2% increase in the improvement in the accuracy for testing training split method.

4.2 Results based on perceptron model

The second method used to predict the stock market price movement is a perceptron model of neural network. It is considered as the simplest form of neural network. The performance parameters used are the same for all the machine learning algorithms used in this work. Using k-fold cross-validation, the perceptron algorithm is used to predict the movement of stock price. The data are divided into ten splits. The perceptron model is also implemented using training–testing splits. 70% of the data are used for the training, and 30% data are used for the testing. Also, the implementation is done after converting the data into time series supervised learning format. The perceptron is trained for 400 iterations. The performance achieved is given in Table 3.

4.3 Results based on logistic regression

Another algorithm considered in this work for the prediction of stock market price movement is logistic regression. The logistic regression, a classification algorithm, is also implemented by dividing the dataset in the training–testing splits. Here again, the machine learning model is implemented after conversion of the data into time series supervised learning format. The performance achieved is given in Table 4.

Table 2 Performance of the SVM algorithm

Sr. no.	Performance parameter	Non-time series supervised learning		Time series supervised learning format	
		k-fold cross-validation value (%)	Train–test split value (%)	*k*-fold cross-validation value (%)	Train–test split value (%)
1	Accuracy	87.22	87.35	75.48	89.33
2	Precision	87.69	87.57	84.93	89.33
3	Recall	87.09	86.92	70.07	89.27
4	F1 score	87.19	87.21	70.61	89.27

Table 3 Performance of the perceptron neural network

Sr. no.	Performance parameters	Non-time series supervised learning		Time series supervised learning	
		k-fold cross-validation value (%)	Train–test split value (%)	*k*-fold cross-validation value (%)	Train–test split value (%)
1.	Accuracy	76.99	75.88	75.48	76.68
2.	Precision	83.70	83.42	84.93	84.34
3.	Recall	76.36	74.80	70.07	74.91
4.	F1 score	74.67	73.17	70.61	73.61

Table 4 Performance of the logistic regression

Sr. no.	Performance parameters	Non-time series supervised learning		Time series supervised learning	
		k-fold cross-validation value (%)	Train–test split value (%)	*k*-fold cross-validation value (%)	Train–test split value (%)
1.	Accuracy	87.10	86.98	89.45	89.93
2.	Precision	87.19	86.72	89.72	89.98
3.	Recall	87.52	87.20	89.28	89.81
4.	F1 score	87.17	86.93	89.40	89.87

4.4 Comparative analysis

In the previous section, the results of the implemented classification algorithms have been mentioned. There is some difference between the accuracies obtained from the implementation of the algorithms. Overall, the *k*-fold cross-validation implementation gives a better accuracy than the training–testing method and that is expected since the *k*-fold cross-validation considers every part of the dataset for the training and then tests the model on all splits of the data. Hence, any inconsistencies in the data cease to have any significant impact on the performance. The SVM classification algorithm shows the best accuracy among the three algorithms. Figure 3 shows the graphical representation of the accuracies obtained through *k*-fold cross-validation implementation.

Also, it is observed that after the conversion of the data into time series supervised learning format the overall accuracy improves by more than 2% in the case of SVM and logistic regression. No significant change is observed in the accuracy of perceptron technique. The average of the Type I error and Type II error that occurred by implementing the algorithms on the data is given in Table 5. There is a significant difference in the error measurement of the non-time series supervised learning data implementation and the one converted into the time series supervised learning. After the time series conversion, the error percentage has dropped. The graphical representation of the Type I error and Type II error is shown in Fig. 4.

4.5 Discussion

The accuracies that are obtained show that logistic regression and SVM perform better than the perceptron model. The Type I error and Type II error are also high in case of the perceptron. This indicates that perceptron does

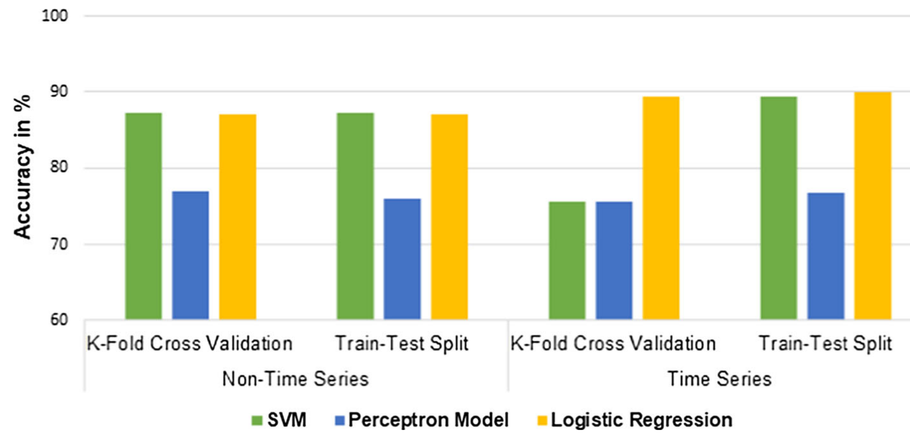

Fig. 3 Accuracy of machine learning algorithms

Table 5 Error measurement of the machine learning algorithms

	Non-time series supervised learning			Time series supervised learning		
	SVM	Perceptron	Logistic regression	SVM	Perceptron	Logistic regression
Type I error (%)	12.42	16.57	13.27	10.66	15.65	10.01
Type II error (%)	16.31	17.90	12.71	10.68	17.83	10.14

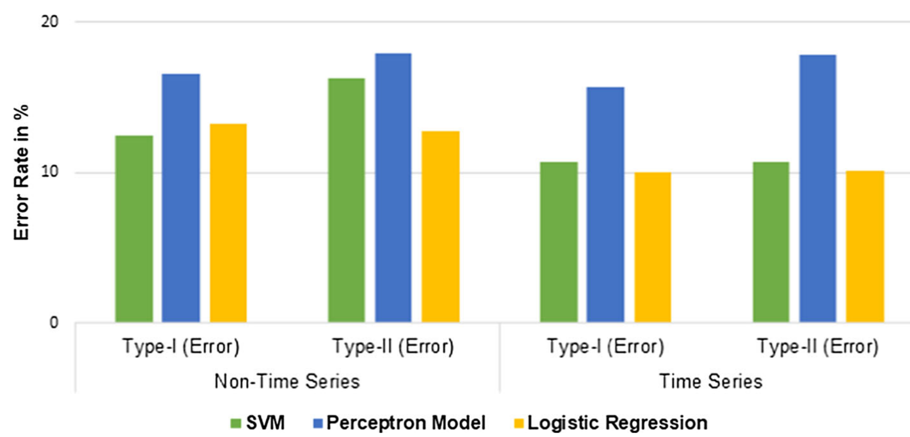

Fig. 4 Error measurement

not entirely handle the complexities of the stock data since perceptron is a single-layer neural network. The SVM algorithm performs very well with linear kernel since the problem is the binary classification problem. Also, once the algorithms are applied to the data in the time series supervised learning format, the accuracies increase and errors decrease.

5 Conclusion

This manuscript contributes to the process of predicting the trend of the stock market price movement through machine learning approach. The historical data of 6 years were employed in this work for the training and learning of machine learning models. The stock data inculcate the fifty stocks of the NSE NIFTY 50 index. For predicting the stock trend with significant accuracy, various machine learning models produced different accuracy rates. The accuracy rates can be improved if the data were converted into time series supervised learning format that too with the same algorithm. It is here submitted that on average 2% increase in the accuracy was reported, when the trend was estimated using time series supervised learning data. For the time series implementation, this work used the lag of the past three days. It is stated by the efficient market hypothesis that the stock market prediction is dependent on the recent information, so if the qualitative dimension of the stocks is added to this problem of predicting the behavior of the stocks, the accuracy rates will surely improve. For the future directions, the work can be expanded by adding more historical data and calculation of more technical indicators as features.

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflict of interest in this work.

References

Balaji AJ, Ram DSH, Nair BB (2018) Applicability of deep learning models for stock price forecasting an empirical study on BANKEX data. Proc Comput Sci 143:947–953

Dash R, Dash PK (2016) A hybrid stock trading framework integrating technical analysis with machine learning techniques. J Finance Data Sci 2:42–57

Deng W, Zhao H, Yang X, Xiong J, Sun M, Li B (2017) Study on an improved adaptive PSO algorithm for solving multi-objective gate assignment. Appl Soft Comput 59:288–302

Deng W, Xu J, Zhao H (2019) An improved ant colony optimization algorithm based on hybrid strategies for scheduling problem. IEEE Access 7:20281–20292

Fenghua W, Jihong X, Zhifang H, Xu G (2014) Stock price prediction based on SSA and SVM. Proc Comput Sci 31:625–631

Henrique BM, Sobreiro VA, Kimura H (2018) Stock price prediction using support vector regression on daily and up to the minute prices. J Finance Data Sci 4(3):2–36

Lachiheb O, Gouider MS (2018) A hierarchical deep neural network design for stock returns prediction. Proc Comput Sci 126:264–272

Madge S, Bhatt S (2015) Predicting stock price direction using support vector machines. Independent work report. Retrieved from https://pdfs.semanticscholar.org/1f75/856bba0feb216001ba551d249593a9624c01.pdf?_ga=2.178641506.711012539.1587273609-1120604514.1549653500

Moghaddam AH, Moghaddam MH, Esfandyari E (2016) Stock market index prediction using artificial neural network. J Econ Finance Adm Sci 21:89–93

Nahil A, Lyhyaoui A (2018) Short-term stock price forecasting using kernel principal component analysis and support vector machines: the case of Casablanca stock exchange. Proc Comput Sci 127:161–169

Nayak SC, Misra BB, Behera HS (2017) Artificial chemical reaction optimization of neural networks for efficient prediction of stock market indices. Ain Shams Eng J 8:371–390

Shi L, Teng Z, Wang L, Zhang Y, Binder A (2018) DeepClue: visual interpretation of text-based deep stock prediction. IEEE Trans Knowl Data Eng Adv Online Publ. https://doi.org/10.1109/tkde.2018.2854193

Wang S, Shang W (2014) Forecasting direction of china security index 300 movement with least squares support vector machine. Proc Comput Sci 3:869–874

Weng B, Ahmed MA, Megahed FM (2017) Stock market one-day ahead movement prediction using disparate data sources. Expert Syst Appl 79:153–163

Zhao H, Liu H, Xu J, Deng W (2019) Performance prediction using high-order differential mathematical morphology gradient spectrum entropy and extreme learning machine. IEEE Trans Instrum Meas. https://doi.org/10.1109/tim.2019.2948414

Zhao H, Zheng J, Deng W, Song Y (2020) Semi-supervised broad learning system based on manifold regularization and broad network. IEEE Trans Circuits Syst. https://doi.org/10.1109/tcsi.2019.2959886