



Optimal Feature Selection of Technical Indicator and Stock Prediction Using Machine Learning Technique

Nagaraj Naik^(✉) and Biju R. Mohan

Department of Information Technology, National Institute of Technology,
Karnataka, Surathkal, India
{it16fv04.nagaraj,biju}@nitk.ac.in
<http://www.nitk.c.in>

Abstract. Short-term trading is a difficult task due to fluctuating demand and supply in the stock market. These demands and supply are reflected in stock prices. The stock prices may be predicted using technical indicators. Most of the existing literature considered the limited technical indicators to measure short-term prices. We have considered 33 different combinations of technical indicators to predict the stock prices. The paper has two objectives, first is the technical indicator feature selection and identification of the relevant technical indicators by using Boruta feature selection technique. The second objective is an accurate prediction model for stocks. To predict stock prices we have proposed ANN (Artificial Neural Network) Regression prediction model and model performance is evaluated using metrics is Mean absolute error (MAE) and Root mean square error (RMSE). The experimental results are better than the existing method by decreasing the error rate in the prediction to 12%. We have used the National Stock Exchange, India (NSE) data for the experiment.

Keywords: Boruta feature selection · ANN · Stock prediction

1 Introduction

Predicting stock price for short-term trade is a challenging task due to the volatile nature of the stock market. This leads an attraction for researchers and academicians to enhance the prediction model. Interpreting stock critical information earlier gives the profitable trading [1]. Efficient Market Hypothesis stated that all stock related information impact on stock prices rather than technical trading [7].

There are two ways of stock analysis can be performed to gain huge profits [10]. First, the fundamental analysis and next is technical analysis. Fundamental analysis is about of evaluating an individual company share price by looking at company past performances. A fundamental analyst studies anything that affects the company share prices, including company earnings, the condition of

Table 1. Technical indicators and its formulas [5]

Technical indicator name	Calculation	Number of days(n)
Simple Moving Average (SMA)	$(C_t + C_{t-1} + \dots + C_{t-n+1})/n$	5,10,14,30,50,100,200
Exponential Moving Average (EMA)	$(C_t - SMA(n)_{t-1}) * (2/(n+1)) + SMA(n)_{t-1}$	5,10,14,30,50,100,200
Momentum Indicator (MOM)	$C_t - C_{n-9}$	5,10,14
Stochastic oscillator (STCK)	$100 * ((C_t - L_t(n)) / (H_t(n) - L_t(n)))$	14
Stochastic oscillator (STCD)	$(100 * ((C_t - L_t(n)) / (H_t(n) - L_t(n)))) / 3$	14
Moving Average Convergence/Divergence (MACD)	$SMA(n) - SMA(n)$	26,13,19,45,25,15
Relative Strength Index (RSI)	$100 - (100 / (1 + \text{Avg}(\text{Gain}) / \text{Avg}(\text{Loss})))$	14
Williams R (R)	$((H_t - C_t) / (H_n - L_n)) \times 100$	14,28,50,100
Accumulation/distribution index(A/D)	$H_t - C_{t-1} / H_t - L_t$ $((C_t / L) / (H / C)) / (H / L)$	14
Commodity Channel Index (CCI)	$(H + L + C/3) - SMA / (0.015 * \text{Mean deviation})$	14,50,100

company macroeconomic factors such as the overall economy, industry conditions and company management. If a particular stock is undervalued and company fundamentals are strong, then it indicates buy signal to investors. If the stock is overvalued and stock fundamentals are weak, then it indicates sell signals to investors [4].

Technical analysis is a method of measuring stock. It involves many statistical analysis of market data, such as stock open price, close price, day high price, day low price and volume of the stock. Technical analysis is different from fundamental analysis. Technical analysis does not attempt to measure a stock by looking at company financial data. Technical analysis is an attempt of finding traders or investors sentiments. By looking at the technical chart, the technical analyst may predict the trend of market sentiments [5,10]. To realize trading signals detection, several methods have been developed, among which artificial intelligence methods have drawn more and more attention by both investors and researchers [3].

The benefits involved in predicting trading signals to gain more returns have motivated researchers to design newer and more advanced tools and prediction techniques. The paper has two objectives, first is the technical indicator feature selection and identification of the relevant technical indicators by using Brouta feature selection techniques. The second objective is an accurate prediction model for stocks. The proposed ANN Regression prediction model outperform compared to existing work.

2 Related Work

Existing trading rules were not gainful for future periods when market condition changes dynamically. Chourmouziadis and Chatzoglou [4] proposed short-term

technical trading strategy by considering the daily price of the stock using fuzzy systems. An automatic way of buying and selling financial securities without the help of portfolio managers has been discussed. The combination of technical trading indicators like moving average, alpha, beta and volatility of the stock over a period of time has been proposed [2]. Nakano et al. [9] proposed a method in which non-linear financial time-series data are considered and machine learning techniques were used for predicting stock prices. Mousavi et al. [8] proposed generalized Exponential Moving Average technical indicator model to predict the stock prices. The future performance of stock indices has been studied using fuzzy time series modeling [12]. Return and risk are important objectives for managing a portfolio. Macedo et al. [6] proposed a model to enhance technical trading rule indicator based on Moving average convergence/divergence, Relative Strength Index, Bollinger Bands and Contrarian Bollinger Bands. Artificial Neural Network (ANN) has been widely used in predicting stock for financial markets. Zhang and Wu [17] proposed optimization technique with back propagation ANN to predict stock prices and indices.

Technical indicators like moving average, moving average convergence and divergence, relative strength index and commodity channel index have been used to predict the stock price [15]. Performing feature extraction could help to reduce the redundant features, which can reduce the measurements, storage requirements and the running time of classifiers. It also avoids the curse of dimensionality and improves prediction performance as well as facilitate data visualization and understanding [14]. Ticknor [13] proposed artificial neural network approach to improve the prediction performance. Preis et al. [11] hypothesized that investors may use a Google hits ratio of pages are used to take the decision to predict stock price. Macroeconomic factors are believed to influence stock market movements. Machine learning methods, which are data-driven and assumption free have become more popular in stock market prediction [16].

3 Research Data

In this paper, stock data are collected from <http://www.nseindia.com>. The data contains information about stock such as stock day open price, day low price, day high price and day close price. We have considered ICICI Bank and State bank of India stocks for the experiment. The dataset range is obtained from the year 2008 to 2018.

In Table 1 the abbreviations used are as follows:

$n \rightarrow$ Total number of days.

$H_t \rightarrow$ day high stock price at period t .

$L_t \rightarrow$ day low stock price at period t .

$C_t \rightarrow$ day close stock price at period t .

4 Proposed Work

Overall Framework of the proposed model is described in Fig. 1. The data are retrieved from NSE. We have considered 33 different combinations of technical

indicators and evaluated based on formulas which are described in Table 1. We have used Boruta feature selection method to select the optimal feature of technical indicators. The step by step proposed Boruta feature selection algorithm is stated below.

Step 1: Create duplicate copies of technical indicators feature.

Step 2: Do the random Shuffle original technical indicators and duplicate copies of technical indicator to remove their correlations with the outcome variable.

Step 3: Apply random forest algorithm to find important technical indicator feature based on higher mean values.

Step 4: Calculate Z score by using Mean/Std deviation.

Step 5: Find the maximum Z score on duplicates technical indicator feature.

Step 6: Remove technical indicator feature if Z is less than Technical indicator feature.

Step 7: Repeat the above steps till iteration completes.

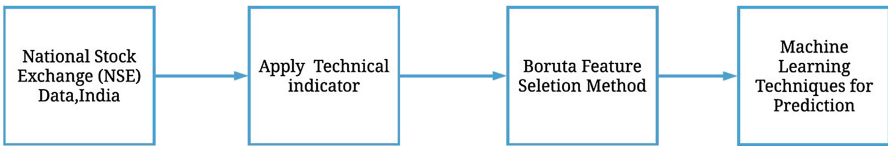


Fig. 1. Overall framework of proposed prediction model

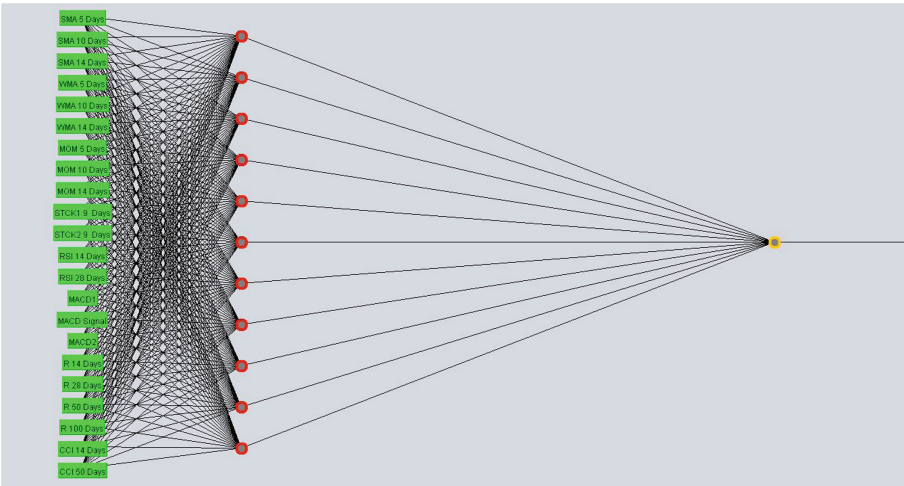


Fig. 2. The proposed ANN regression prediction model

Table 2. Stock ICICI Bank after feature selection.

Technical indicator feature	meanImp	medianImp	minImp	maxImp	normHits	Decision
SMA 5 days	5.905820	5.805600	3.5773017	8.330396	1.00000000	Confirmed
SMA 10 days	4.412296	4.450179	2.5664467	6.381897	0.94949495	Confirmed
SMA 14 days	3.904888	3.865653	1.7642646	5.652071	0.86868687	Confirmed
WMA 5 days	5.108720	5.116738	1.9129550	7.403865	0.94949495	Confirmed
WMA 10 days	4.715484	4.789195	2.7125061	6.435136	0.95959596	Confirmed
WMA 14 days	4.357254	4.259337	1.9938488	6.654934	0.91919192	Confirmed
MOM 5 days	18.692034	18.659231	16.9670413	20.344669	1.00000000	Confirmed
MOM 10 days	10.406941	10.374657	9.1582195	11.926020	1.00000000	Confirmed
MOM 14 days	9.074784	9.058466	7.0379555	10.879017	1.00000000	Confirmed
STCK1 9 days	21.269620	21.339074	18.4123667	23.683564	1.00000000	Confirmed
STCK2 9 days	25.729514	25.897507	22.7613795	28.723707	1.00000000	Confirmed
RSI 14 days	15.487287	15.438833	13.8918381	16.732855	1.00000000	Confirmed
RSI 28 days	8.608997	8.602433	7.0550301	10.168651	1.00000000	Confirmed
MACD1 (26,13,9) days	16.152372	16.184757	14.2102377	18.254305	1.00000000	Confirmed
MACD SIGNAL	9.164158	9.182269	7.2546778	10.961601	1.00000000	Confirmed
MACD (15,25,15)	6.046035	5.977311	3.7503675	8.219024	1.00000000	Confirmed
R 14 days	41.157543	41.103670	38.5168811	44.780878	1.00000000	Confirmed
R 28 days	24.057435	24.017315	22.3425494	26.501515	1.00000000	Confirmed
R 50 days	19.071730	19.049960	17.2172337	21.637499	1.00000000	Confirmed
R 100 days	12.560288	12.563706	10.8720538	14.235407	1.00000000	Confirmed
CCI 14 days	17.287549	17.246896	15.4191912	19.279900	1.00000000	Confirmed
CCI 50 days	9.203503	9.272982	7.1159672	11.020916	1.00000000	Confirmed
CCI 100 days	6.259018	6.366987	3.7899522	8.369810	0.98989899	Confirmed

Prediction Model

ANN Regression Model. The ANN can provide a more accurate prediction model for the larger amount of data, nonlinear data, nonstationary and has been a popular approach for stock market prediction. Feature selection performed on technical indicator using Boruta algorithm and selected technical indicator feature is given as input to the ANN regression model and it is described in Fig. 2. In this work, the ANN is used to predict the stock prices. ANN has three layers, each layer is connected to the other. The neurons represent the technical indicators. The sigmoid function activation function is used in the ANN regression model. The threshold value 0.5 has been set. A gradient descent momentum parameters are considered to determine the weights and to reduce the global minimum.

5 Experimental Results and Discussion

There are two phases, the first phase of experiment considered feature selection technique. The second phase of the experiment is the prediction model. In the

Table 3. Stock State Bank of India after feature selection.

Technical indicator feature	meanImp	medianImp	minImp	maxImp	normHits	Decision
SMA 5 days	5.4606723	5.5251714	3.1273954	7.653046	1.00000000	Confirmed
SMA 10 days	3.6382503	3.6526033	0.8677842	6.344588	0.81818182	Confirmed
WMA 5 days	4.8639987	4.9798041	2.1810807	7.662506	0.96969697	Confirmed
WMA 10 days	4.2772306	4.2978051	2.0651554	8.159167	0.91919192	Confirmed
WMA 14 days	3.7222162	3.9236859	0.8468174	6.263128	0.87878788	Confirmed
MOM 5 days	19.2684238	19.1960889	17.7018349	21.289981	1.00000000	Confirmed
MOM 10 days	8.5963907	8.6058484	6.7912856	9.896962	1.00000000	Confirmed
MOM 14 days	9.1535623	9.1827829	7.7671871	10.944682	1.00000000	Confirmed
STCK1 9 days	22.8154629	22.8720799	20.0837955	25.518795	1.00000000	Confirmed
STCK2 9 days	21.6583364	21.6522237	19.0834711	23.759105	1.00000000	Confirmed
RSI 14 days	15.1007526	15.0648255	13.6517412	17.075092	1.00000000	Confirmed
RSI 28 days	9.0990375	9.0610509	6.6856226	10.547252	1.00000000	Confirmed
MACD1 (26,13,9) days	14.6987380	14.6169198	12.7518900	16.634019	1.00000000	Confirmed
MACD SIGNAL	8.3677429	8.4367255	6.3049964	10.389434	1.00000000	Confirmed
MACD (15,25,15)	6.1898880	6.2728086	4.1248029	7.845896	1.00000000	Confirmed
R 14 days	37.9879805	37.9176749	34.9232519	40.963526	1.00000000	Confirmed
R 28 days	25.7820969	25.7832651	23.5718441	27.630123	1.00000000	Confirmed
R 50 days	18.4300230	18.4008101	16.4257141	20.670375	1.00000000	Confirmed
R 100 days	14.0062758	13.9214321	11.6389235	16.356005	1.00000000	Confirmed
A/D	17.2844799	17.2619585	15.0965778	19.177774	1.00000000	Confirmed
CCI 14 days	7.1505888	7.1446862	5.2038171	9.449850	1.00000000	Confirmed
CCI 50 days	8.7992939	8.8263404	6.7923020	10.430373	1.00000000	Confirmed

first phase of the experiment, we have used the Boruta method to select the best feature based on the Z Score. ICICI Bank and State bank of India stocks after feature selection are described in Tables 2 and 3. These selected features of the stock technical indicator are given as input to the ANN regression model. The MAE and RMSE are used to evaluate the performance of the prediction model and it is described in Eqs. 1 and 2. The proposed ANN Regression prediction

model performance is compared with existing work and it is shown in Table 4. All tests are trained with a 10 fold cross-validation based model. From this study, we can conclude that technical indicator feature is important to predict the stock prices.

$$MAE = \frac{1}{n} \sum_{t=1}^n |e_t| \tag{1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \tag{2}$$

Table 4. Result comparison with existing work.

Stock Name	Prediction Model	MAE	RMSE
ICICI Bank	ANN [5]	27.0583	36.0444
ICICI Bank	Proposed ANN Model	15.1221	19.9444
State Bank of India	ANN [5]	27.7392	36.4834
State Bank of India	Proposed ANN Model	17.4341	23.1585

6 Conclusion

The paper focused on stock prediction for short-term trading. The stock data is collected from the National Stock Exchange (NSE). The paper has two objectives, first is the technical indicator feature selection and identification of the relevant technical indicators by using Boruta feature selection techniques. The second objective is an accurate prediction model for stocks. The performance for stock ICICI Bank MAE is 15.12 and stock State Bank of India MAE is 14.4. The experimental prediction model ANN outperforms compares to existing work by decreasing the error rate in the prediction. The future work can be identified the microeconomics, macroeconomics factor and fundamental analysis find the quality of stocks.

Acknowledgment. This work has been supported by the Visvesvaraya Ph.D Scheme for Electronics and IT (Media Lab Asia), the departments of MeitY, Government of India. The Task carried out at the Department of Information Technology, NITK Surathkal, Mangalore, India.

References

1. Alexander, S.S.: Price movements in speculative markets: trends or random walks. *Ind. Manag. Rev.* **2**(2), 7 (1961). (pre-1986)
2. Berutich, J.M., López, F., Luna, F., Quintana, D.: Robust technical trading strategies using GP for algorithmic portfolio selection. *Expert Syst. Appl.* **46**, 307–315 (2016)
3. Chen, Y., Hao, Y.: Integrating principle component analysis and weighted support vector machine for stock trading signals prediction. *Neurocomputing* **321**, 381–402 (2018)
4. Chourmouziadis, K., Chatzoglou, P.D.: An intelligent short term stock trading fuzzy system for assisting investors in portfolio management. *Expert Syst. Appl.* **43**, 298–311 (2016)
5. Kara, Y., Boyacioglu, M.A., Baykan, O.K.: Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul stock exchange. *Expert syst. Appl.* **38**(5), 5311–5319 (2011)
6. Macedo, L.L., Godinho, P., Alves, M.J.: Mean-semivariance portfolio optimization with multiobjective evolutionary algorithms and technical analysis rules. *Expert Syst. Appl.* **79**, 33–43 (2017)
7. Malkiel, B.G., Fama, E.F.: Efficient capital markets: a review of theory and empirical work. *J. Financ.* **25**(2), 383–417 (1970)
8. Mousavi, S., Esfahanipour, A., Zarandi, M.H.F.: A novel approach to dynamic portfolio trading system using multitree genetic programming. *Knowl.-Based Syst.* **66**, 68–81 (2014)
9. Nakano, M., Takahashi, A., Takahashi, S.: Generalized exponential moving average (EMA) model with particle filtering and anomaly detection. *Expert Syst. Appl.* **73**, 187–200 (2017)
10. Patel, J., Shah, S., Thakkar, P., Kotecha, K.: Predicting stock market index using fusion of machine learning techniques. *Expert Syst. Appl.* **42**(4), 2162–2172 (2015)
11. Preis, T., Moat, H.S., Stanley, H.E.: Quantifying trading behavior in financial markets using Google trends. *Sci. Rep.* **3**, 01684 (2013)
12. Rubio, A., Bermúdez, J.D., Vercher, E.: Improving stock index forecasts by using a new weighted fuzzy-trend time series method. *Expert Syst. Appl.* **76**, 12–20 (2017)
13. Ticknor, J.L.: A Bayesian regularized artificial neural network for stock market forecasting. *Expert Syst. Appl.* **40**(14), 5501–5506 (2013)
14. Tsai, C.-F., Hsiao, Y.-C.: Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches. *Decis. Support Syst.* **50**(1), 258–269 (2010)
15. Tsai, C.-F., Lin, Y.-C., Yen, D.C., Chen, Y.-M.: Predicting stock returns by classifier ensembles. *Appl. Soft Comput.* **11**(2), 2452–2459 (2011)
16. Vaisla, K.S., Bhatt, A.K.: An analysis of the performance of artificial neural network technique for stock market forecasting. *Int. J. Comput. Sci. Eng.* **2**(6), 2104–2109 (2010)
17. Zhang, Y., Lenan, W.: Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network. *Expert Syst. Appl.* **36**(5), 8849–8854 (2009)