The current issue and full text archive of this journal is available on Emerald Insight at:
https://www.emerald.com/insight/2514-9288.htm

Data mining–based stock price prediction using hybridization of technical and fundamental analysis

Jasleen Kaur
Chitkara Business School, Chitkara University, Punjab, India, and
Khushdeep Dharni
School of Business Studies, Punjab Agricultural University, Ludhiana, India

Received 5 April 2022
Revised 21 October 2022
5 December 2022
16 January 2023
Accepted 6 February 2023

Abstract

Purpose – The stock market generates massive databases of various financial companies that are highly volatile and complex. To forecast daily stock values of these companies, investors frequently use technical analysis or fundamental analysis. Data mining techniques coupled with fundamental and technical analysis types have the potential to give satisfactory results for stock market prediction. In the current paper, an effort is made to investigate the accuracy of stock market predictions by using the combined approach of variables from technical and fundamental analysis for the creation of a data mining predictive model.

Design/methodology/approach – We chose 381 companies from the National Stock Exchange of India's CNX 500 index and conducted a two-stage data analysis. The first stage is identifying key fundamental variables and constructing a portfolio based on that study. Artificial neural network (ANN), support vector machines (SVM) and decision tree J48 were used to build the models. The second stage entails applying technical analysis to forecast price movements in the companies included in the portfolios. ANN and SVM techniques were used to create predictive models for all companies in the portfolios. We also estimated returns using trading decisions based on the model's output and then compared them to buy-and-hold returns and the return of the NIFTY 50 index, which served as a benchmark.

Findings – The results show that the returns of both the portfolios are higher than the benchmark buy-and-hold strategy return. It can be concluded that data mining techniques give better results, irrespective of the type of stock, and have the ability to make up for poor stocks. The comparison of returns of portfolios with the return of NIFTY as a benchmark also indicates that both the portfolios are generating higher returns as compared to the return generated by NIFTY.

Originality/value – As stock prices are influenced by both technical and fundamental indicators, the current paper explored the combined effect of technical analysis and fundamental analysis variables for Indian stock market prediction. Further, the results obtained by individual analysis have also been compared. The proposed method under study can also be utilized to determine whether to hold stocks for the long or short term using trend-based research.

Keywords Data mining techniques, Artificial neural networks, Support vector machines, Decision tree J48, National Stock Exchange, Stock market prediction, Technical analysis, Fundamental analysis

Paper type Research paper

1. Introduction

Stock markets generate massive datasets from a variety of financial companies. Investors are looking for a quick and easy way to obtain relevant stock information. Furthermore, with easy access to stock data, these investors are anticipating stocks in order to maximize their gains. Due to their efficiency in extracting information from non-linear stock market data, data mining techniques have risen to prominence in the field of stock market prediction. These techniques have grown in importance as a study field (Aggrawal *et al.*, 2019; Fayyad *et al.*, 1996; Weiss and Indurkhya, 1998; Chen *et al.*, 2006). The stock

Funding: The authors acknowledge the support of Indian Council of Social Science Research (ICSSR) (grant number RFD/2016-17/GEN/MGT/094) for carrying out this study.

Conflict of interest statement: We have no conflicts of interest to disclose.

Data Technologies and Applications

2514-9288
DOI 10.1108/DTA-04-2022-0142

market is influenced by a variety of factors. Researchers and investors use a variety of models to forecast stock market trends and prices, including single and dual sources of information (Mitilineos and Artikis, 2017; Chen *et al.*, 2018). Single source of information either consider numerical information or semantic information extracted from news. On the other hand, dual sources of information consider historical prices and volume data as well.

Technical indicators have been used to investigate the dynamics of stock price movement by analysing the past sequence of stock prices. The technical analysis method is popular among investors as a primary analytical approach for efficient investment decision-making (Agrawal *et al.*, 2019). Technical analysis studies the market's generated data, that is price and volume, to investigate the price movement's future direction (Chandra, 2008). This approach to investing is essentially a reflection of the idea that prices move in trends that are determined by the attitudes of investors towards economic, political, monetary and psychological forces. For technical analysis, several academics have employed indicators, which include moving average, exponential moving average, bias, moving average convergence divergence, stochastic %K, stochastic %D, on balance volume, momentum, William's %R, A/D oscillator, price rate of change (ROC), disparity 5 days, disparity 10 days, commodity channel index, price oscillator and relative strength index (Versace *et al.*, 2004).

Besides the technical analysis methods, fundamental analysis has also caught the attention of researchers. Fundamental analysts concentrate on the economic forces that cause security prices to rise, fall or remain constant (Murphy, 1999). Every aspect that can affect the value of a security is studied under this analysis, which includes both company-specific factors (e.g. financial condition and management) and macroeconomic factors (e.g. overall economy and industry conditions), for example interest rates, exchange rates, stock returns, dividend yield, T-bill rate, producer price index, industrial production index, consumer price index, money stock, growth in net profit, growth in assets, equity growth, fixed assets/total assets, current assets/total assets, equity/tangible assets, equity/assets, net profit or current assets, return on assets, earnings per share, return on equity, price-earnings ratio, market to book value, news articles and so on (Ayodele *et al.*, 2012; Boyacioglu and Avci, 2010; Emir *et al.*, 2012; Vanstone *et al.*, 2010). Therefore, there are multiple variables that can be considered in fundamental analysis.

Furthermore, technical and fundamental indicators can be collectively employed to forecast the stock markets. A few examples of combined indicators are available (Zhai *et al.*, 2007; Schumaker and Chen, 2009; Fung *et al.*, 2002). As stock prices are influenced by both technical and fundamental indicators, considering both types has the potential to give better results. The focus of this paper is to investigate the accuracy of stock market predictions by using the combined approach of variables from technical and fundamental analysis for the creation of data mining predictive models.

The proposed work assists investors in making profitable investment selections. The model also includes a decision indicator. The proposed method can also be utilized to determine whether to hold stocks for the long or short term using trend-based research. The rest of this paper is organized as follows. Section 2 presents a review of related literature. The experimental setup and methodology used in the study are discussed in Section 3. Section 4 presents results and performance analysis for the evaluation of the proposed work. Section 5 concludes with recommendations for the future.

2. Background of the study

Before the computer age, people usually traded stocks and commodities based on their gut feelings. As the level of investing and trading grew, people searched for tools and methods that would increase their gains while minimizing their risk (Upadhyay *et al.*, 2012). Though analysing stock movement behaviour is a challenging task, data mining techniques can guide an investor in identifying and segmenting high-performance securities so as to make superior investment decisions. Data mining approaches have gained prominence in the field of stock market prediction due to the inherent constraints of conventional forecasting techniques in developing a model to predict future values with accuracy. Traditional approaches have several key flaws, including a faulty forecasting model, the wrong number of variables and faulty coefficient values for these parameters. Data mining techniques can be used to address these problems. These techniques have gained the interest of researchers, practitioners and academicians and have become an increasingly important research area (Weiss and Indurkhya, 1998; Chen *et al.*, 2006; Gupta *et al.*, 2019; Kohli *et al.*, 2019).

Although there are a number of data mining techniques available for stock market predictions to deal with the non-linearity of data, artificial neural networks (ANN) (Vaisla and Bhatt, 2010, Hargreaves and Hay, 2013; Gupta and Sharma, 2014) and support vector machines (SVM) (Hou *et al.*, 2013; Lahmiri, 2011; Hou *et al.*, 2013) are definitely among the most popular choices (Kaur and Dharni, 2022). In the literature, there is a clear difference of opinion regarding the choice of technique. Given the accuracy in terms of prediction direction, neural networks have been chosen as the preferred method (Altay and Satman, 2005; Hammad *et al.*, 2007; Majumder and Hussian, 2007; Tjung *et al.*, 2010; Vojinovic *et al.*, 2001). Support for the technique SVM is also evident in the literature (Cao and Tay, 2001; Kim, 2003; Hou *et al.*, 2013; Arasu *et al.*, 2014; Kumar and Thenmozhi, 2006). Decision trees make them ideal for knowledge discovery due to their classifier building ability without any parameter settings and domain information. Multidimensional data can also be handled by decision trees with a high level of accuracy (Han *et al.*, 2012). Various researchers have supported the use of decision trees in the domain of stock market prediction (Adebimpe *et al.*, 2012; Hargreaves and Hay, 2013; Gupta and Sharma, 2014; Tsai and Wang, 2009; Saeedmanesh *et al.*, 2010). In the current study, ANN, SVMs and decision trees were employed for stock prediction.

Statistical techniques, technical analysis, fundamental analysis, chaos theory, linear regression and time series analysis are the few techniques that have been adopted to predict the stock markets (Ravichandran *et al.*, 2005). Technical analysis and fundamental analysis are the two most popular techniques among researchers in the current arena. Various evidences of stock market prediction, predominantly engaging the variables of technical analysis, are available (Kara *et al.*, 2011; Kimoto *et al.*, 1990; Chang and Liu, 2008; Chavarnakul and Enke, 2008; Versace *et al.*, 2004; Altay and Satman, 2005; Choudhry and Garg, 2008; Umbarkar and Nandgaonkar, 2015; Cruz *et al.*, 2003; Huang *et al.*, 2008; Kumar and Thenmozhi, 2006; Zhai *et al.*, 2007; Atsalakis and Valavanis, 2009; Lahmiri, 2011; Teixeira and Oliveira, 2010; Kwon and Moon, 2007; Dai *et al.*, 2012; Hsieh *et al.*, 2011; Kim, 2006; Kim and Han, 2000; Kim, 2003; Cao and Tay, 2001). Few researchers have also employed fundamental analysis (Ayodele *et al.*, 2012; Boyacioglu and Avci, 2010; Emir *et al.*, 2012; Vanstone *et al.*, 2010; Enke and Thawornwong, 2005; Thawornwong and Enke, 2004; Hong *et al.*, 2007), but the majority of studies have ignored the impact of fundamental analysis.

In the current study, we have explored the combined effect of technical analysis and fundamental analysis variables for Indian stock market prediction, given the combined approach has more potential to enhance the decision-making quality as compared to the single technique (De Souza *et al.*, 2018; Ayodele *et al.*, 2012; Zhai *et al.*, 2007; Schumaker and Chen, 2009; Fung *et al.*, 2002).

3. Methods

We evaluated the performance of data mining techniques for a selected Indian stock exchange, that is the National Stock Exchange of India (NSE), using ANN, SVM and decision trees.

3.1 Sampling

We selected companies from the NSE's CNX 500 index. The CNX 500 is India's first broad-based capital market benchmark. As of 31 March 2015, the CNX 500 Index represented 95.77 per cent of the free float market capitalization of the stocks listed on the NSE. Automobiles, construction, textiles, consumer goods, fertilizers and pesticides, financial services, energy, pharmaceuticals, industrial manufacturing, information technology, chemicals and other sectors are represented in the CNX 500. Out of all sectors, we excluded banking sector further analysis as the performance measurement variables of banking sector are different as compared to variables of other sectors under consideration.

3.2 Data collection and preprocessing

For the period 1 April 1998 to 31 March 2017, we collected data on 52 fundamental variables from all companies. The information was gathered from the Ace Equity Database and the Reserve Bank of India's official website. We used 12 variables from previous studies by Enke and Thawornwong (2005), Thawornwong and Enke (2004), Fan and Palaniswami, 2001, Emir *et al.* (2012), Boyacioglu and Avci (2010), Ayodele *et al.* (2012), Hong *et al.* (2007), Vanstone *et al.* (2010), Naik and Padhi (2012), Yu *et al.* (2009) and another 40 variables as additional variables. The initial collection of 52 fundamental variables is shown in Table I.

3.3 Data analysis

We followed a two-stage data analysis procedure. The first stage entails identifying key fundamental variables and constructing a portfolio based on that analysis. The second stage entails applying technical analysis to forecast price movements in the companies included in the portfolios.

We used factor analysis for the identification of important fundamental variables. Factor analysis attempts to identify a set of dimensions that are not directly observable in a large set of variables. This analysis is used to summarize a majority of the information in a dataset in terms of relatively fewer new categories, called factors. The major use of factor analysis is to group redundant variables so that a smaller number of variables can be selected for further analysis. Factor analysis begins with the construction of a new set of variables based on the relationships in the correlation matrix. Principal component analysis (PCA) can be used to transform a set of variables into a new set of composite variables or principal components that are not correlated with each other. These linear combinations of variables, called factors, account for the variance in the data as a whole. The best combination makes up the first component, which is the first factor. The second principal component is defined as the best linear set of variables for explaining the variance not accounted for by the first factor. There may be many factors, each being the best linear combination of variables not yet accounted for by previous factors.

Mathematically, each variable is expressed as a linear combination of factors. The covariation among the variables is described in terms of a small number of common factors plus a unique factor for each variable.

$$X = \mu + A \times F + U,$$

where

X = set of p observed variables $X_1, X_2, \ldots, X_p$ with mean vector μ ($p \times 1$) and covariance matrix $\Sigma(p \times p)$,

F = A set of s unobserved variables called common factors $F_1, F_2, \ldots, F_s$,

U = A set of p unique but unobserved factors $U_1, U_2, \ldots, U_p$.

S. No.	Variable	S. No.	Variable
1	Cash profit	27	Cash flow from finance activities
2	Investments	28	Free cash flow
3	ROCE (%) (return on capital employed)	29	ROE (%) (return on equity)
4	Price/book value (*x*)	30	PATM (%) (profit after tax margin)
5	Equity dividend (%)	31	Market capitalization
6	PE (*x*) (price to earning ratio)	32	Debt to equity (*x*)
7	EPS (Rs) (earnings per share)	33	Current ratio (*x*)
8	Book value (Rs)	34	Enterprise value
9	Dividend pay out ratio (%)	35	CEPS (Rs) (cash earning per share)
10	ROA (%) (return on asset)	36	DPS (Rs) (dividend per share)
11	M3 (money supply)	37	Adjusted DPS (Rs)
12	Average CPI (consumer price index)	38	Adjusted book value (Rs)
13	Gross sales	39	EBITM (%) (earning before interest and taxes margin)
14	Total income	40	Pre-tax margin (%)
15	Total expenditure	41	Fixed asset turnover (*x*)
16	PBIDT (profit before interest, depreciation and taxes)	42	Sales/working capital (*x*)
17	PBIT (profit before interest and taxes)	43	Quick ratio (*x*)
18	PBT (profit before taxes)	44	Receivable days
19	PAT (profit after taxes)	45	Inventor days
20	Reserves and surplus	46	PAT growth (%)
21	Net worth	47	Adjusted EPS growth (%)
22	Capital employed	48	Bank rate
23	Gross block	49	Profit before interest, depreciation and tax margin (PBIDTM (%))
24	Total current liabilities	50	CPM (%) (cash profit margin)
25	Total assets	51	Adjusted EPS
26	Cash flow from operations	52	Returns

Note: First 12 variables are taken from previous studies and next 40 variables are additionally considered for the study

Table I.
Initial set of 52 fundamental variables

We considered variables in the reduced dataset as independent variables and bidirectional variables of returns, that is increase and decrease, as dependent variables. Also, we considered the period from 1 April 1998 to 31 March 2013, 2010–13 training set, as a training period and three test sets of 1 year each as test sets, that is 2013–14, 2014–15 and 2015–16. We built models using ANN, SVM and decision tree J48 and tested the models for test sets, which resulted in stocks being classified into two categories (above median return and below median return) of 20 stocks each selected on the basis of fundamental variables. For this purpose, we calculated the medians of the respective years and compared them with the actual returns. We considered companies with returns greater than the median return as “above the median return companies” and the remaining companies as “below the median return companies.” Further, we constructed two sets with above and below returns based on the median of the respective years. Both sets included 20 companies, each selected on the basis of consistent prediction of directions in ANN, SVM and decision trees J48.

For the second stage of analysis, we collected daily data from prowess on a subset of 40 companies and transformed it into the technical indicators listed in Annexure I . We built predictive models for all companies using ANN and SVM techniques. For the data of each company, we used 80 per cent of the records for training the model and the remaining 20 per cent for testing the model. After the model building, we calculated returns using trading decisions based on the outcome of the model and then compared the returns with the buy-and-hold returns and the return of the NIFTY 50 set as a benchmark. The overall methodology is represented in Figure 1.

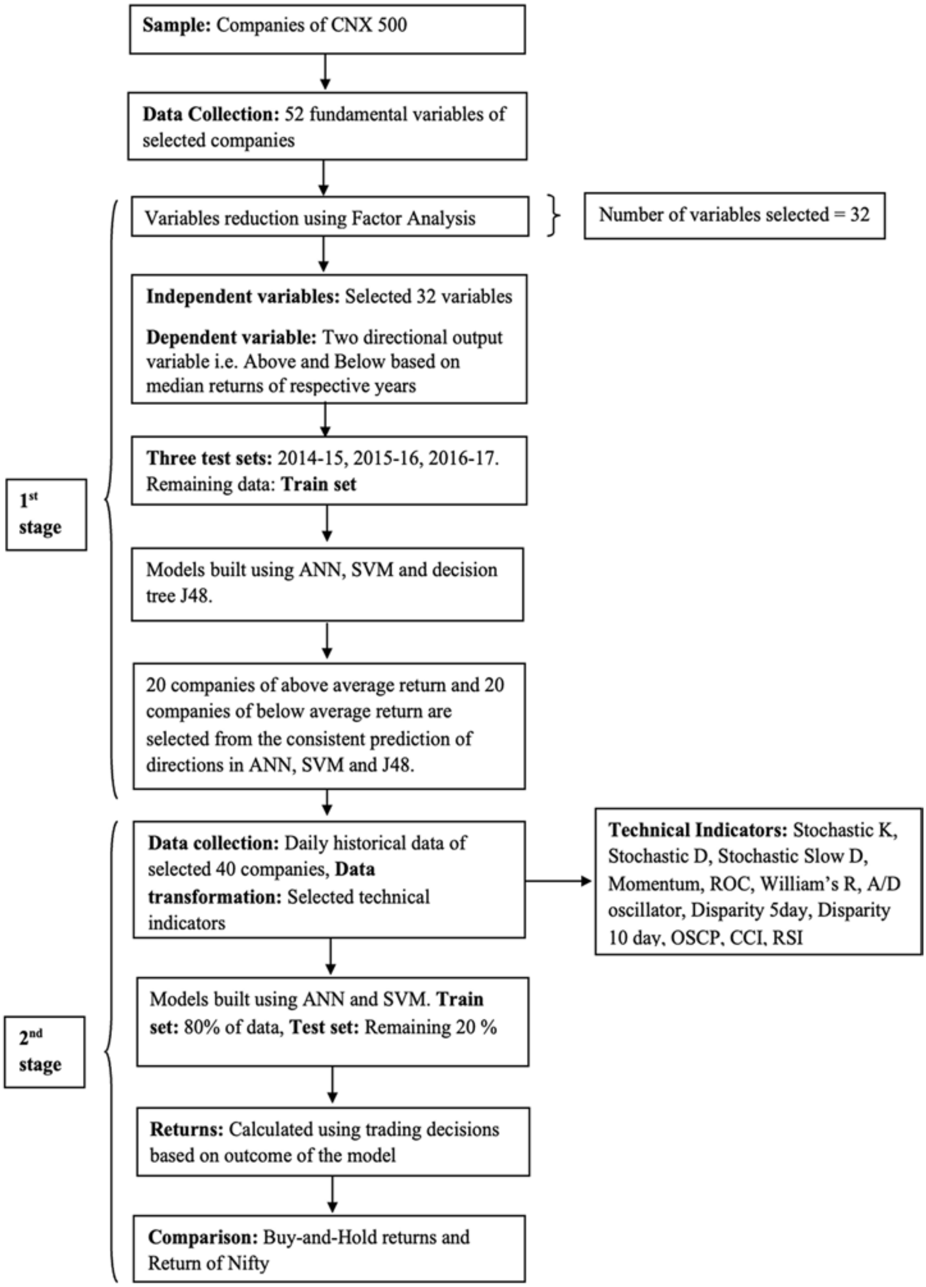

Figure 1.
Two stage methodology followed in the study

3.3.1 Artificial neural networks. We used a multilayer perceptron classifier with backpropagation to categorize instances in this study. The experiment was carried out with "a" hidden layer, and the network's nodes are all sigmoid. The hidden layer's level, "a," is equal to (attributes + classes)/2. The data under consideration has 14 attributes and two classes. Learning rate (lr), momentum constant (mc) and the number of epochs to train through are all set to 0.3, 0.2 and 500, respectively.

3.3.2 Support vector machines. Because of its capacity and ease of implementation, we utilized the radial basis function kernel as the kernel function of SVM in this study. Weka software was used to implement John Platt's sequential minimal optimization approach. The levels for the various parameters evaluated in this study, such as the complexity parameter (c), ε parameter, tolerance parameter and gamma of kernel function, are 1, 10^{-12}, 0.001 and 0.01, respectively.

3.3.3 Decision tree J48. We implemented the algorithm for building decision trees in Weka as a classifier called J48. The classifier is represented as J48 –C 0.25 –M 2. The value of confidence factor was 0.25. Confidence factor was used for pruning, which indicates that smaller values incur more pruning. Num folds value was taken as 3. Num folds determined the amount of data used for reduced error pruning. One fold was used for pruning, the rest for growing the tree. The minimum number of instances per leaf was 2.

3.3.4 Trading strategy. If an increase in stock price/index value is predicted in the directional forecast, the trading method involves deciding to BUY the stock. A position that is taken and kept until the opposing directional prediction is met. The position was squared off by selling when the opposite directional outcome (i.e. from increase to decline) was received. Similarly, when the directional prediction was for a fall, a SELL position was taken and held until the opposite result (i.e. a rise) was achieved. The BUY–SELL cycle was treated as a single trade. In addition, short selling of stocks was not considered when computing the returns. All trades had their own returns, which were then added together to calculate the average return of all companies.

4. Results

To meet the requirements of data analysis techniques and to ensure the uniformity of variables and data availability, a few companies were excluded from the consideration set. Findings discussed in the following sections are based on 381 companies with time period ranging from 1 April 1998 to 31 March 2016. The process of selecting explanatory variables is given in the following section.

4.1 Data reduction

We used factor analysis for variable reduction to identify the key input variables for building predictive models. Factor analysis results in the exclusion of less relevant fundamental variables and provides a set of variables that produce more information. For all CNX500 companies, we initially included 52 variables (shown in Table I). We run PCA to identify the final set of input variables. For the identification of these variables, we employ a series of factor analysis procedures in an iterative mode. After implementing the first PCA, we eliminated variables with loading values of less than 0.7 so as to maximize the total variance explained. Factor analysis was run sequentially with stepwise elimination of variables with lower loadings, that is loadings less than 0.8. Optimum results were obtained at the end of the fifth factor analysis run on the sequence. With the elimination of variables, an increase in total variance was observed, and the process of elimination of variables continued till a peak of total variance explained was observed. Hence, we considered the set of variables with the maximum total variance explained and the acceptable value of Kaiser–

Meyer–Olkin (KMO) as the final set of input variables. For factor analysis, we used the varimax rotation method with Kaiser normalization. The final set of variables, thus identified, are shown in italics in Table I.

Table II represents the values of the KMO measure of sampling adequacy and Bartlett's test of sphericity for the factor analysis run for the identification of variables. The value of KMO measures of sampling adequacy came out to be 0.809, which indicates that sample size is adequate (William *et al.*, 2010). The value of Chi-square for Bartlett's test of sphericity is 452,383.9191. This value came out to be significant at 1 per cent level of significance with 496 degrees of freedom. Therefore, 32 variables obtained with the help of the PCA extraction method are used for further analysis.

Eight factors are obtained with the help of PCA extraction method represented in Table III. These eight factors can explain 83.469 per cent of the total variance.

The factors extracted are explained as follows:

(1) *Profitability and size indicators*: This factor is related to various variables related to income and expenditure performance, sources of funds and applications of the funds. This factor includes profit before interest, depreciation and taxes (PBIDT), profit before interest and taxes (PBIT), profit before taxes (PBT), profit after taxes (PAT), cash profit, reserves and surplus, net worth, capital employed, gross block, total assets and market capitalization. This factor explains 32.059 per cent of the total variance.

(2) *Margin ratios*: Margin ratios factor is related to various margin ratios like Profit before interest, depreciation and tax margin (PBIDTM) (per cent), profit after tax margin (PATM) (per cent), cash profit margin (CPM) (per cent), earning before interest and taxes margin (EBITM) (per cent) and pre-tax margin(per cent). This factor is able to explain 13.020 per cent of the total variance.

(3) *Dividend measures*: This factor includes equity dividend per cent, dividend per share (DPS) (Rs) and adjusted DPS (Rs) and explains 8.871 per cent of the total variance.

(4) *Earning per share measures:* This factor includes adjusted earnings per share (EPS), cash earning per share (CEPS) (Rs) and adjusted book value (Rs) and explains 8.753 per cent of the total variance.

(5) *Liquidity measures*: Liquidity measures factor includes only two variables, that is current ratio (*x*) and quick ratio (*x*) and explains 6.307 per cent of the total variance.

(6) *Growth measures*: Growth measures factor includes only PAT growth (per cent) and adjusted EPS growth (per cent) explaining 5.909 per cent of the total variance.

(7) *Money supply and inflation*: This factor is related to money supply M3 and average consumer price index (CPI), and it is able to explain 4.528 per cent of the total variance.

(8) *Bank rate*: Last factor is related to bank rate only, which explains 4.022 per cent of the total variance.

Table III represents that maximum loading for factor 1 is from PBIDT, that is 0.974. For factor 2, PATM (per cent) has maximum loading, that is 0.945. Similarly, for factor 3,

Table II. Values for KMO and bartlett's test for fundamental variables

Kaiser–Meyer–Olkin measure of sampling adequacy	0.809
Bartlett's test of sphericity	
Approximately Chi-square	452,383.919
df	496
Sig.	0.000

Factor	Eigen value	% of variance	Items	Item loading
Profitability and size indicators	10.259	32.059	PBIDT	0.974
			PBIT	0.958
			PBT	0.941
			PAT	0.940
			Cash profit	0.961
			Reserves and surplus	0.949
			Net worth	0.951
			Capital employed	0.878
			Gross block	0.857
			Total assets	0.915
			Market capitalization	0.853
			Enterprise value	0.899
Margin ratios	4.167	13.020	PBIDTM (%)	0.844
			PATM (%)	0.945
			CPM (%)	0.944
			EBITM (%)	0.855
			Pre-tax margin (%)	0.942
Dividend measures	2.839	8.871	Equity dividend (%)	0.970
			DPS (Rs)	0.976
			Adj DPS(Rs)	0.966
Earnings per share measures	2.801	8.753	Adjusted EPS	0.980
			CEPS (Rs)	0.953
			Adjusted book value (Rs)	0.961
Liquidity measures	2.018	6.307	Current ratio (*x*)	0.991
			Quick ratio (*x*)	0.992
Growth measures	1.891	5.909	PAT growth (%)	0.966
			Adj. EPS growth (%)	0.967
Money supply and inflation	1.449	4.528	M3	0.836
			Average CPI	0.861
Bank rate	1.287	4.022	Bank rate	0.956

Table III. Factors related to fundamental variables

maximum loading is from DPS (Rs), that is 0.976; for factor 4, maximum loading is from adjusted EPS, that is 0.980 and for factor 5, maximum loading is from quick ratio, that is 0.992. Adjusted EPS growth has maximum loading of 0.967 for factor 6. Average CPI has maximum loading of 0.861 for factor 7 and bank rate has 0.956 loading in factor 8.

4.2 Model built using ANN, SVM and decision trees (J48) for portfolio creation

After selecting the variables, we considered these fundamental variables independent variables. We calculated the annual returns of 381 stocks from 1 April 1998 to 31 March 2016. These returns were arranged in descending order. We marked stocks with returns greater than the returns of the second quartile as "above average return" and stocks with returns less than the returns of the second quartile as "below average return."

Further, we built ANN, SVM and decision tree (J48) models using binary outcome variables of "above median return" and "below median return". We considered a training set. Table IV summarises the parameters of summary statistics, that is hit ratio, Kappa statistics, mean absolute error (MAE), root mean square error (RMSE), precision recall (PRC), F-measure, Matthews correlation co-efficient (MCC), receiver operating characterstics (ROC) area, and PRC area of models built using ANN, SVM and decision tree (J48) for three test sets pertaining to 2013–14, 2014–15 and 2015–16.

With the help of the predicted outcomes of ANN, SVM and decision tree (J48), we selected two sets of companies, which are represented in Tables V and VI. The first set was made up of 20 companies which were generating returns higher than the average return (labelled as Portfolio1). The second set was made up of 20 companies which were generating returns less than the average returns (labelled as Portfolio 2).

4.3 Results of predictive models of portfolio stocks

In the second stage of data analysis, prices of individual stocks were predicted using technical analysis variables. The daily data starting from 1 April 2014 to 31 March 2017 of the above 40 companies was considered and transformed into technical indicators. These technical indicators were considered independent variables. Predictive models were built for all companies using ANN and SVM techniques where the dependent variables were considered as bidirectional variables, that is increases and decreases in value of returns. Initial 80 per cent of the data were used as a training set, with the remaining 20 per cent used to test the model. The performance of all models was evaluated on the basis of various parameters like hit ratio, MAE and RMSE.

Table V shows that in the case of Portfolio 1, a maximum hit ratio of 0.958 is obtained for the ANN prediction model of Shriram Transport Finance Company Ltd. and a minimum hit ratio of 0.813 is obtained for the ANN model of Gujarat Alkalies and Chemicals Ltd. The ANN model of Swan Energy Ltd. has obtained the maximum values of MAE and RMSE, that is 0.247 and 0.413, respectively. The minimum values of MAE and RMSE were obtained for the ANN model of Shriram Transport Finance Company Ltd., that is 0.066 and 0.198, respectively. The ANN model of Swan Energy Ltd. generates the maximum return, that is 84.633, and the minimum return is generated by the ANN model of Godrej Properties Ltd., that is 30.177. Table V also shows that a maximum hit ratio of 0.938 is obtained for the SVM prediction model in the case of Hindustan Zinc Ltd.'s portfolio 1, and a minimum hit ratio of 0.785 is obtained for the SVM model of Swan Energy Ltd. The SVM model of HEG Ltd. has obtained the maximum values of MAE and RMSE, that is 0.153 and 0.391, respectively. A minimum value of MAE and RMSE was obtained for the SVM model of Hindustan Zinc Ltd., that is 0.063 and 0.250, respectively. The SVM model of Swan Energy Ltd. generates the maximum return, that is 92.902, and the minimum return is generated by the SVM model of Bombay Dyeing & Manufacturing Company Ltd., that is 31.070.

Table VI indicates that in the case of Portfolio 2, a maximum hit ratio (0.931) is obtained for the ANN prediction model of Tata Consultancy Services Ltd. and a minimum hit ratio (0.826) is obtained for the ANN model of Siemens Ltd. The maximum value of MAE is obtained by the ANN model of United Breweries Ltd. (0.181) and the maximum value of

Table IV.
Summary statistics of models: ANN, SVM and J48

Technique	Test set	Hit ratio	Kappa stat	MAE	RMSE	Precision	Recall	F-measure	MCC	ROC area	PRC area
ANN	2013–14	50.663	0.011	0.476	0.506	0.517	0.507	0.407	0.019	0.683	0.642
	2014–15	55.438	0.107	0.472	0.499	0.592	0.554	0.502	0.141	0.710	0.681
	2015–16	59.055	0.180	0.464	0.485	0.633	0.591	0.554	0.219	0.708	0.683
SVM	2013–14	51.724	0.032	0.483	0.695	0.573	0.517	0.400	0.069	0.516	0.509
	2014–15	50.133	0.000	0.499	0.706	0.501	0.501	0.391	0.001	0.500	0.500
	2015–16	49.344	0.011	0.493	0.703	0.523	0.507	0.390	0.022	0.505	0.502
Decision tree (J48)	2013–14	59.151	0.183	0.414	0.579	0.592	0.592	0.591	0.183	0.611	0.579
	2014–15	59.416	0.206	0.416	0.567	0.734	0.594	0.621	0.244	0.649	0.707
	2015–16	61.417	0.206	0.435	0.548	0.724	0.614	0.640	0.232	0.625	0.693

Table V. Summary statistics and returns of all stocks of Portfolio 1

Companies	ANN Hit ratio	MAE	RMSE	Returns	No. of trades	Annual return	SVM Hit ratio	MAE	RMSE	Returns	No. of trades	Annual return
Balrampur Chini Mill Ltd.	0.854	0.148	0.350	133.576	36	46.817	0.896	0.104	0.323	137.958	34	48.352
Bombay Burmah Trading Corporation Ltd.	0.868	0.139	0.327	169.376	30	59.364	0.875	0.125	0.354	165.405	34	57.972
Bombay Dyeing & Manufacturing Company Ltd.	0.847	0.151	0.359	192.427	30	67.443	0.875	0.125	0.354	88.648	34	31.070
E.I.D. Parry (India) Ltd.	0.889	0.111	0.299	105.352	35	36.925	0.889	0.111	0.333	105.266	34	36.894
Godrej Properties Ltd.	0.861	0.149	0.334	86.100	30	30.177	0.910	0.090	0.301	90.268	31	31.638
Gujarat Alkalies and Chemicals Ltd.	0.813	0.180	0.384	123.704	25	43.357	0.875	0.125	0.354	132.477	28	46.432
Gujarat State Fertilizers & Chemicals Ltd.	0.861	0.145	0.332	156.020	29	54.683	0.861	0.139	0.373	152.722	31	53.527
HEG Ltd.	0.854	0.153	0.339	133.428	30	46.765	0.847	0.153	0.391	128.336	32	44.980
Hindustan Zinc Ltd.	0.924	0.099	0.265	138.922	37	48.690	0.938	0.063	0.250	138.583	40	48.571
Indiabulls Real Estate Ltd.	0.903	0.103	0.275	168.857	36	59.182	0.924	0.076	0.276	167.780	39	58.805
Jindal Saw Ltd.	0.889	0.117	0.312	172.688	33	60.525	0.882	0.118	0.344	171.348	29	60.055
JSW Steel Ltd.	0.840	0.179	0.363	89.999	30	31.543	0.875	0.125	0.354	96.833	34	33.939
KNR Construction Ltd.	0.847	0.159	0.337	106.207	28	37.224	0.854	0.146	0.382	108.487	27	38.023
MOIL Ltd.	0.931	0.097	0.262	160.613	30	56.293	0.875	0.125	0.354	163.695	26	57.373
Rashtriya Chemicals & Fertilizers Ltd.	0.910	0.084	0.265	152.662	31	53.506	0.924	0.076	0.276	148.654	34	52.101
Shriram Transport Finance Company Ltd.	0.958	0.066	0.198	112.404	32	39.396	0.931	0.069	0.264	104.615	35	36.666
SREI Infrastructure Finance Ltd.	0.840	0.151	0.354	136.791	34	47.944	0.861	0.139	0.373	136.431	36	47.817
Swan Energy Ltd.	0.764	0.247	0.413	241.471	24	84.633	0.785	0.215	0.464	265.085	22	92.909
Tata Sponge Iron Ltd.	0.868	0.139	0.315	115.059	31	40.327	0.868	0.132	0.363	116.612	31	40.871
Vedanta Ltd.	0.875	0.112	0.289	152.975	34	53.616	0.910	0.090	0.301	151.385	36	53.058

Table VI.
Summary statistics and returns of all stocks of Portfolio 2

Companies	ANN Hit ratio	MAE	RMSE	Returns	No. of trades	Annual return	SVM Hit ratio	MAE	RMSE	Returns	No. of trades	Annual return
Asian Paints Ltd.	0.917	0.102	0.279	74.952	36	26.270	0.903	0.097	0.312	75.319	38	26.398
Bajaj Auto Ltd.	0.847	0.149	0.341	58.543	36	20.518	0.924	0.076	0.276	64.073	40	22.457
Castrol India Ltd.	0.896	0.108	0.298	−14.501	36	−5.082	0.924	0.076	0.276	−20.064	41	−7.032
Colgate-Palmolive (India) Ltd.	0.868	0.138	0.339	55.944	31	19.608	0.833	0.167	0.408	55.546	30	19.468
Dabur India Ltd.	0.882	0.120	0.318	60.338	36	21.148	0.889	0.111	0.333	60.909	33	21.348
Dr. Reddys Laboratories Ltd.	0.868	0.148	0.347	50.624	28	17.743	0.868	0.132	0.363	49.407	32	17.316
Eicher Motors Ltd.	0.896	0.107	0.287	103.834	32	36.393	0.889	0.111	0.333	106.915	35	37.472
HCL Technologies Ltd.	0.840	0.152	0.332	69.662	33	24.416	0.889	0.111	0.333	72.553	34	25.429
Hero MotoCorp Ltd.	0.868	0.137	0.312	69.861	30	24.486	0.910	0.090	0.301	71.782	35	25.159
Hindustan Unilever Ltd.	0.882	0.122	0.309	53.367	33	18.704	0.903	0.097	0.312	55.410	35	19.420
Housing Development Finance Corporation Ltd.	0.854	0.148	0.356	77.761	33	27.254	0.868	0.132	0.363	77.136	33	27.035
Infosys Ltd.	0.868	0.146	0.352	59.881	37	20.987	0.882	0.118	0.344	62.949	38	22.063
ITC Ltd.	0.875	0.130	0.328	80.030	39	28.049	0.889	0.111	0.333	78.772	37	27.608
MRF Ltd.	0.882	0.110	0.286	31.301	30	10.970	0.875	0.125	0.354	130.767	38	45.832
Siemens Ltd.	0.826	0.173	0.384	71.068	32	24.909	0.813	0.188	0.433	65.952	33	23.115
Sun Pharmaceutical Industries Ltd.	0.861	0.129	0.321	80.301	33	28.144	0.903	0.097	0.312	79.722	36	27.941
Tata Consultancy Services Ltd.	0.931	0.080	0.250	73.653	31	25.814	0.944	0.056	0.236	75.916	31	26.607
Tata Motors Ltd.	0.854	0.139	0.341	95.991	37	33.643	0.868	0.132	0.363	95.190	39	33.363
United Breweries Ltd.	0.819	0.181	0.379	60.548	29	21.221	0.771	0.229	0.479	53.288	30	18.677
Wipro Ltd.	0.861	0.155	0.332	53.889	34	18.887	0.833	0.167	0.408	52.033	32	18.237

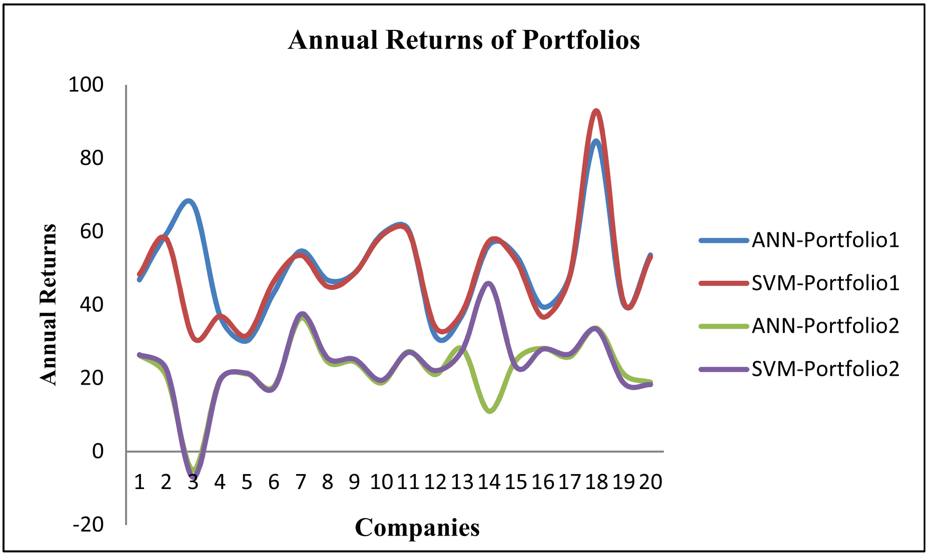

Figure 2.
Annual returns for companies under Portfolio 1 and Portfolio 2

RMSE is obtained by the ANN model of Siemens Ltd. (0.384). A minimum value of MAE and RMSE is obtained for the ANN model of Tata Consultancy Services Ltd., that is 0.080 and 0.250, respectively. The ANN model of Eicher Motors Ltd. is generating the maximum return (36.393) and the ANN model of Castrol India Ltd. is generating negative returns, that is −5.082. Table VI also indicates that for portfolio 2, a maximum hit ratio of 0.944 is obtained for the SVM prediction model in the case of Tata Consultancy Services Ltd. and a minimum hit ratio of 0.813 is obtained for the SVM model of Siemens Ltd. The SVM model of Siemens Ltd. has obtained the maximum values of MAE and RMSE, that is 0.188 and 0.433, respectively. The minimum values of MAE and RMSE were obtained for the SVM model of Tata Consultancy Services Ltd., that is 0.056 and 0.236, respectively. The SVM model of MRF Ltd. generates the maximum return of 45.832 and the SVM model of Castrol India Ltd. generates negative returns, that is −7.032. Annual returns obtained from the models built using ANN and SVM for both the portfolios are graphically represented in Figure 2.

4.4 Comparison of returns with buy-and-hold return

The trading returns and calculated Sharpe ratio based on data mining techniques, that is ANN and SVM, for both portfolios are reported in Table VII. The returns and Sharpe ratio of a buy-and-hold trading strategy for that particular period are also provided. Table VII reveals that both techniques generate higher returns in both portfolios as compared to buy-and-hold returns. Number of trades, based on prediction from data mining models, are expected to be more than that of buy-and-hold strategy (single selling transaction). Procedure of accounting for multiple transaction cost is realistic because higher number of transactions will attract higher transaction costs in real stock market. Besides, the overall returns from data mining model–based predictions remain higher (as compared to buy-and-hold) even after higher transaction costs. Transaction cost of 0.05 per cent has been estimated by considering the charges of stock exchanges for buying/selling equities and brokerage charges.

Specifically, ANN generates the highest return in the first portfolio, and SVM generates the highest return in the second portfolio. Moreover, both the portfolios generate higher returns as compared to the return generated by NIFTY (4.550). The comparison of returns and Sharpe ratio for both the portfolios are graphically represented in Figures 3 and 4.

The average returns from models of companies in portfolios built using ANN and SVM were examined using the Sharpe ratio. The Sharpe ratio is a risk-adjusted return on investment and can be used to evaluate the total performance of an investment portfolio. Also, it is said that the higher the Sharpe ratio, the higher the return and the lower the volatility (Rodriguez *et al.*, 2000; Enke and Thawornwong, 2005). Table VII illustrates that a portfolio of companies with an above average return has the highest Sharpe ratio for ANN models. A Sharpe ratio value of greater than 3 is considered excellent by investors. The Sharpe ratio for the SVM model is also greater than 3 for Portfolio 1. The buy-and-hold strategy has obtained the minimum value of the Sharpe ratio. Portfolio 2 has obtained the highest value of Sharpe ratio for ANN models, that is 2.583, followed by SVM models with a Sharpe ratio of 2.372. The minimum Sharpe ratio has been obtained by the buy-and-hold strategy, that is 0.093.

5. Discussion

The accuracy of models built using technical variables is far better than that of models built using fundamental variables. There has been an enduring debate regarding the comparative performance of technical analysis and fundamental analysis. Few researchers agree with the

Table VII. Returns and sharpe ratio of Portfolio 1, Portfolio 2 and buy-and-hold

	Portfolio 1 Return	Sharpe ratio	Portfolio 2 Return	Sharpe ratio
SVM	138.529	3.478	68.179	2.372
ANN	142.432	3.934	63.351	2.583
Buy-and-hold	41.627	1.177	1.468	0.093

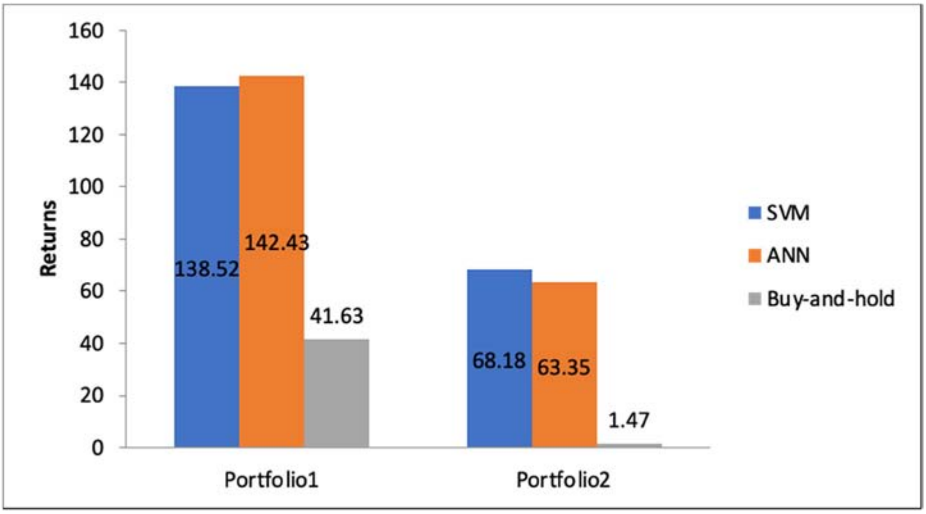

Figure 3. Comparison of returns for Portfolio 1 and Portfolio 2

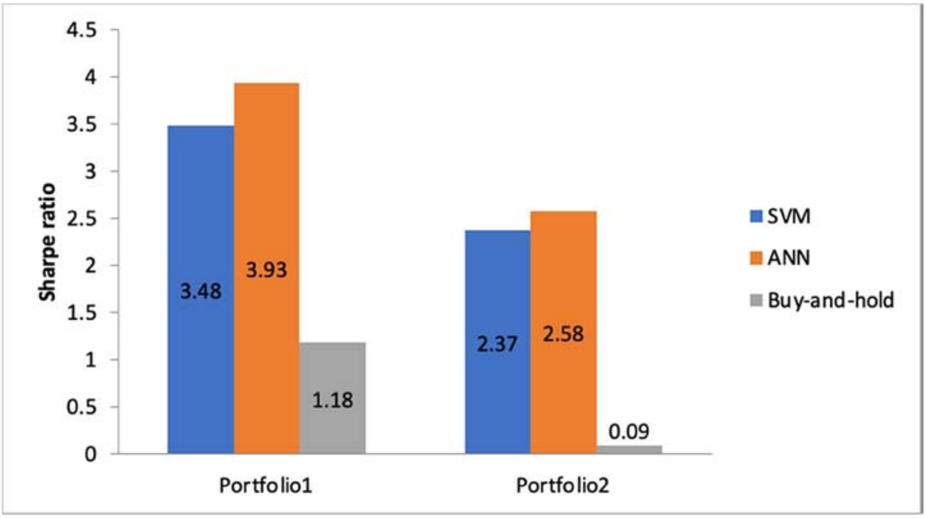

Figure 4.
Comparison of sharpe ratio for Portfolio 1 and Portfolio 2

former (Teixeira and Oliveira, 2010; Kwon and Moon, 2007; Dai *et al.*, 2012; Hsieh *et al.*, 2011), while others prefer the latter (Chen *et al.*, 2003; Fan and Palaniswami, 2001; Mittal and Goel, 2012; Schumaker and Chen, 2009; Vanstone *et al.*, 2010; Yumlu *et al.*, 2005). Evidence of comparison of both techniques is also available (Emir *et al.*, 2012; Neely and Weller, 2012; Taylor and Allen, 1992). Emir *et al.* (2012) gave contradicting results as compared to the results of the current study, which shows that fundamental analysis performs better than technical in terms of accuracy. However, others have supported the results of the current study that technical indicators are better as compared to fundamental indicators (Neely and Weller, 2012; Taylor and Allen, 1992). Also, a few instances have been observed in which both technical and fundamental variables are combined for the creation of a neural network predictive model, and this results in improved accuracy as compared to single technical analysis (Ayodele *et al.*, 2012; Bettman *et al.*, 2009).

In the current study, fundamental variables considered include variables from financial statements, various ratios and industry-specific and economic-specific variables. Investors rely largely on a small list of financial variables to make investment decisions. Financial statements contain useful information for predictability (Bauman, 1996). In this study, the variables included in the study are a mix of variables considered by various researchers, including Huang *et al.* (2004) (financial statement and financial ratios), Enke and Thawornwong (2005) and Thawornwong and Enke (2004) (financial and economic variables), Emir *et al.* (2012) (financial ratios), Campbell and Mankiw (1987), Cochrane (1988) (GNP and GDP), Huizinga (1987) (exchange rate) and Leung *et al.* (2000) (macroeconomic variables).

Both ANNs and SVMs were used to classify the stocks of the CNX 500 into two categories termed "portfolios," that is above average return stocks and below average return stocks. The results have shown that data mining techniques are able to discriminate between both the portfolios of above and below average returns, leading to two entirely different sets of stocks, with one set of stocks having an above average return and the other set of stocks having below average returns.

The performance of both the portfolios in terms of returns has improved with the usage of data mining techniques. The return for the buy-and-hold strategy for Portfolio 1 was observed to be 41.627, which increased to 138.529 in the case of SVM and 142.432 in the case of ANN. Similarly, the return for the buy-and-hold strategy for Portfolio 2 was 1.468, which increased to 68.179 for SVM and 63.351 for ANN. Therefore, the returns of both the portfolios are higher than the benchmark buy-and-hold strategy return. It can be concluded that data mining techniques give better results, irrespective of the type of stock, and have the ability to make up for poor stocks. The comparison of returns of portfolios with the return of NIFTY as a benchmark also indicates that both the portfolios are generating higher returns as compared to the return generated by NIFTY.

5.1 Limitations of the study

(1) The study includes the secondary data which have been downloaded from databases. The accuracy of data is limited by the accuracy of data source.

(2) For the purpose of meaningful comparison, standard methods of data mining techniques have been used for prediction purpose like ANN, SVM, association rules and decision trees. Optimization of these techniques based on individual data sets may yield outcomes with higher accuracy.

(3) Procedures used in conducting factor analysis and identification of portfolio use expost facto data in part, that is from year 2014 to 2016. Given the nature of fundamental analysis, the "variable selection" based on data reduction is expected to hold for a significant duration of time. But still, the results may lack generalization on account of use of expost facto data.

6. Conclusion

The study's findings conclusively demonstrate that the returns of both portfolios outperform the return of the buy-and-hold strategy. It can be concluded that data mining techniques give better results, irrespective of the type of stock, and have the ability to make up for poor stocks. The comparison of returns of portfolios with the return of NIFTY as a benchmark also indicates that both the portfolios are generating higher returns as compared to the return generated by NIFTY.

ORCID iDs

Jasleen Kaur https://orcid.org/0000-0002-8559-106X
Khushdeep Dharni https://orcid.org/0000-0003-3422-3731

References

Adebimpe, L.A., Olusola, A. and Longe, O.B. (2012), "Forecasting portfolio investment using data mining", *African Journal of Computing & ICT*, Vol. 5 No. 4, pp. 101-106.

Agrawal, M., Khan, A.U. and Shukla, P.K. (2019), "Stock price prediction using technical indicators: a predictive model using optimal deep learning", *International Journal of Recent Technology and Engineering (IJRTE)*, Vol. 8 No. 2 July, pp. 2297-2305.

Altay, E. and Satman, M.H. (2005), "Stock market forecasting: artificial neural network and linear regression comparison in an emerging market", *Journal of Financial Management and Analysis*, Vol. 18 No. 2, pp. 18-33.

Arasu, B.S., Jeevananthan, M., Thamaraiselvan, N. and Janarthanan, B. (2014), "Performances of data mining techniques in forecasting stock index - evidence from India and US", *Journal of National Science Foundation Sri Lanka*, Vol. 42 No. 2, pp. 177-191.

Atsalakis, G.S. and Valavanis, K.P. (2009), “Forecasting stock market short-term trends using a neuro-fuzzy based methodology”, *Expert Systems with Applications*, Vol. 36, pp. 10696-10707.

Ayodele, A., Charles, A., Marion, A. and Sunday, O. (2012), “Stock price prediction using neural network with hybridized market indicators”, *Journal of Emerging Trends in Computing and Information Sciences*, Vol. 3, pp. 1-9.

Bauman, M.P. (1996), “A review of fundamental analysis research in accounting”, *Journal of Accounting Literature*, Vol. 15, pp. 1-33.

Bettman, J.L., Sault, S.J. and Schultz, E.L. (2009), “Fundamental and technical analysis”, *Accounting and Finance*, Vol. 49, pp. 21-36.

Boyacioglu, M.A. and Avci, D. (2010), “An adaptive network-based fuzzy inference system (ANFIS) for the prediction of stock market return: the case of the Istanbul stock exchange”, *Expert Systems with Applications*, Vol. 37, pp. 7908-7912.

Campbell, J.Y. and Mankiw, N.G. (1987), “Permanent and transitory components in macroeconomic fluctuations”, *American Economic Review*, Vol. 77, pp. 111-117.

Cao, L. and Tay, F.E.H. (2001), “Financial forecasting using support vector machines”, *Neural Computing & Applications*, Vol. 10, pp. 184-192.

Chandra, P. (2008), *Investment Analysis and Portfolio Management*, Tata McGraw Hill Education Private, New Delhi, pp. 464-477.

Chang, P.C. and Liu, C.H. (2008), “A neural network with a case based dynamic window for stock trading prediction”, *Expert Systems with Applications*, Vol. 36, pp. 6889-6898.

Chavarnakul, T. and Enke, D. (2008), “Intelligent technical analysis based equivolume charting for stock trading using neural networks stock trading using neural networks”, *Expert Systems with Applications*, Vol. 34, pp. 1004-1017.

Chen, A., Leung, M.T. and Daouk, H. (2003), “Application of neural networks to an emerging financial market: forecasting and trading the Taiwan stock index”, *Computers & Operations Research*, Vol. 30, pp. 901-923.

Chen, W.H., Shih, J.Y. and Wu, S. (2006), “Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets”, *International Journal of Electronic Finance*, Vol. 1, pp. 49-67.

Chen, Y.J., Chen, Y.M., Tsao, S.T. and Hsieh, S.F. (2018), “A novel technical analysis-based method for stock market forecasting”, *Soft Computing*, Vol. 22, pp. 1-18. doi: 10.1007/s00500-016-2417-2.

Choudhry, R. and Garg, K. (2008), “a hybrid machine learning system for stock market forecasting”, *World Academy of Science, Engineering and Technology*, Vol. 15, pp. 315-318.

Cochrane, J.H. (1988), “How big is the random walk in GNP?”, *Journal of Political Economy*, Vol. 9, pp. 893-920.

Cruz, P.F., Rodrıguez, J.,.-A. and Giner, J. (2003), “Estimating GARCH models using support vector machines”, *Quantitative Finance*, Vol. 3, pp. 1-10.

Dai, W., Wu, J.-Y. and Lu, C.-J. (2012), “Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes”, *Expert Systems with Applications*, Vol. 39, pp. 4444-4452.

De Souza, M.J.S., Ramos, D.G.F., Pena, M.G., Sobreiro, V.A. and Kimura, H. (2018), “Examinat ion of the profitability of technical analysis based on moving average strategies in BRICS”, *Financial Innovation*, Vol. 4 No. 3, doi: 10.1186/s40854-018-0087-z.

Emir, Ş., Dinçer, H. and Timor, M. (2012), “A stock selection model based on fundamental and technical analysis variables by using artificial neural networks and support vector machines”, *Review of Economics and Finance*, Vol. 2, pp. 106-122.

Enke, D. and Thawornwong, S. (2005), “The use of data mining and neural networks for forecasting stock market returns”, *Expert Systems with Applications*, Vol. 29, pp. 927-940.

Fan, A. and Palaniswami, M. (2001), "Stock selection using support vector machines," *IJCNN '01. International Joint Conference on Neural Networks, Proceedings (Cat. No.01CH37222)*, Washington, DC, USA, 2001, pp. 1793-1798 Vol. 3, doi: 10.1109/IJCNN.2001.938434.

Fayyad, U., Djorgovski, S.G. and Weir, N. (1996), "Automating the analysis and cataloging of sky surveys", in Fayyad, U., Shapiro, G.P., Smyth P. and Uthurusamy, R. (Eds), *Advances in Knowledge Discovery and Data Mining*, MIT Press, Cambridge, pp. 471-494.

Fung, G.P.C., Yu, J.X. and Lam, W. (2002), "News sensitive stock trend prediction", *Proceedings of the 6th Pacific-Asia Conference on Knowledge Discovery and Data Mining- PAKDD 2002*, Taipei, Taiwan, pp. 289-296.

Gupta, A., Bhatia, P., Dave, K. and Jain, P. (2019), "Stock market prediction using data mining techniques", *2nd International Conference on Advances in Science and Technology (ICAST*), Mumbai, India

Gupta, A. and Sharma, S.D. (2014), "Clustering-classification based prediction of stock market future prediction", *International Journal of Computer Science and Information Technologies*, Vol. 5 No. 3, pp. 2806-2809.

Hammad, A.A.A., Ali, S.M.A. and Hall, E.L. (2007), "Forecasting the jordanian stock prices using artificial neural network", *Intelligent Engineering Systems Through Artificial Neural Networks*, ASME Press, NY, USA.

Han, J., Kamber, M. and Pei, J. (2012), *Data Mining Concepts and Techniques*, 3rd ed., Elsevier, Waltham, USA.

Hargreaves, C. and Hay, Y. (2013), "Prediction of stock performance using analytical techniques", *Journal of Emerging Technologies in Web Intelligence*, Vol. 5 No. 2, pp. 136-142.

Hong, H., Torous, W. and Valkanov, R. (2007), "Do industries lead the stock market?", *Journal of Financial Economics*, Vol. 83, pp. 367-396.

Hou, L., Yang, S. and Chen, Z. (2013), "The use of data mining techniques and support vector regression for financial forecasting", *International Journal of Database Theory and Application*, Vol. 6 No. 4, pp. 145-156.

Hsieh, T., Hsiao, H. and Yeh, W. (2011), "Forecasting stock markets using wavelet transforms and recurrent neural networks: an integrated system based on artificial bee colony algorithm", *Applied Soft Computing*, Vol. 11, pp. 2510-2525.

Huang, C., Yang, D. and Chuang, Y. (2008), "Application of wrapper approach and composite classifier to the stock trend prediction", *Expert Systems with Applications*, Vol. 34, pp. 2870-2878.

Huang, Z.H., Chen, H., Hsu, C.J. and Chen, W.H. (2004), "Credit rating analysis with support vector machines and neural networks: a market comparative study", *Decision Support Systems*, Vol. 37, pp. 543-558.

Huizinga, J. (1987), "An empirical investigation of the long run behavior of real exchange rates", *Carnegie Rochester Conference Series on Public Policy*, Vol. 27, pp. 149-214.

Kara, Y., Boyacioglu, M.A. and Baykan, O.K. (2011), "Predicting direction of stock price index movement using artificial neural networks and support vector machines: the sample of the Istanbul stock exchange", *Expert Systems with Applications*, Vol. 38, pp. 5311-5319.

Kaur, J. and Dharni, K. (2022), "Application and performance of data mining techniques in stock market: a review", *Intelligent Systems in Accounting, Finance and Management*, Vol. 29 No. 4, pp. 219-241. doi: 10.1002/isaf.1518.

Kim, K. (2003), "Financial time series forecasting using support vector machines", *Neurocomputing*, Vol. 55, pp. 307-319.

Kim, K. (2006), "Artificial neural networks with evolutionary instance selection for financial forecasting", *Expert Systems with Applications*, Vol. 30, pp. 519-526.

Kim, K. and Han, I. (2000), "Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index", *Expert Systems with Applications*, Vol. 19, pp. 125-132.

Kimoto, T., Asakawa, K., Yoda, M. and Takeoka, M. (1990), "Stock market prediction system with modular neural network", *Proceedings of the International Joint Conference on Neural Networks*, IEEE, San Diego, CA, pp. 1-6.

Kohli, P.P.S., Zargar, S., Arora, S. and Gupta, P. (2019), "Stock prediction using machine learning algorithms", *Applications of Artificial Intelligence Techniques in Engineering, Advances in Intelligent Systems and Computing book series*, Springer, Vol. 698, pp. 405-414.

Kumar, M. and Thenmozhi, M. (2006), "Forecasting stock index movement: a comparison of support vector machines and random forest", *Indian Institute of Capital Markets 9th Capital Markets Conference Paper*, 2005, http://ssrn.com/abstract=876544

Kwon, Y.-K. and Moon, B.-R. (2007), "A hybrid neurogenetic approach for stock forecasting", *IEEE Transactions on Neural Networks*, Vol. 18, pp. 851-864.

Lahmiri, S. (2011), "Neural networks and investors sentiment measures for stock market trend prediction", *Journal of Theoretical Applied Information Technology*, Vol. 27, pp. 1-10.

Leung, M.T., Daouk, H. and Chen, A.S. (2000), "Forecasting stock indices: a comparison of classification and level estimation models", *International Journal of Forecasting*, Vol. 16, pp. 173-190.

Majumder, M. and Hussian, M.D. (2007), "Forecasting of Indian stock market index using artificial neural network", *Information Science*, pp. 98-105.

Mitilineos, S.A. and Artikis, P.G. (2017), "Forecasting of future stock prices using neural networks and genetic algorithms", *International Journal of Decision Sciences, Risk and Management*, Vol. 7 Nos 1/2, pp. 2-25.

Mittal, A. and Goel, A. (2012), "Stock prediction using Twitter sentiment analysis", *Stanford University Working Paper*, CS229, Stanford, CA, USA, p. 15.

Murphy, J.J. (1999), *Technical Analysis of Financial Markets: a Comprehensive Guide to Trading Methods and Applications*, Executive Tax Reports, New York, pp. 1-22.

Naik, P.K. and Padhi, P. (2012), "The impact of macroeconomic fundamentals on stock prices revisited: evidence from Indian Data", *Eurasian Journal of Business and Economics*, Vol. 5, pp. 25-44.

Neely, C.J. and Weller, P.A. (2012), " Technical analysis in the foreign exchange market.", in James, J., Marsh, I. and Sarno, L. (Eds), *Wiley Handbook of Exchange Rates*, Wiley Publishing, Hoboken, NJ, John Wiley, pp. 343-374.

Ravichandran, K.S., Thirunavukarasu, P., Nallaswamy, R. and Babu, R. (2005), "Estimation of return on investment in share market through ANN", *Journa l of Theoretical and Applied Information Technology*, Vol. 3, pp. 44-54.

Rodriguez, F., Gonzalez-Martel, C. and Sosvilla-Rivebo, S. (2000), "On the profitability of technical trading rules based on artificial neural networks: evidence from the Madrid stock market", *Economics Letters*, Vol. 69, pp. 89-94.

Saeedmanesh, M., Izadi, T. and Ahvar, E. (2010), "HDM: a hybrid data mining technique for stock exchange prediction", *Proceedings of International Multiconference of Engineers and Computer Scientists*, Vol. 1, IMECS, Hong Kong, pp. 587-592.

Schumaker, R.P. and Chen, H. (2009), "Textual analysis of stock market prediction using breaking financial news: the azfin text system", *ACM Transaction Information Systems*, Vol. 27, pp. 1-19.

Taylor, M.P. and Allen, H.L. (1992), "The use of technical analysis in the foreign exchange market", *International Journal of Money and Finance*, Vol. 11, pp. 304-314.

Teixeira, L.A. and Oliveira, A.L.I. (2010), "A method for automatic stock trading combining technical analysis and nearest neighbor classification", *Expert Systems Applications*, Vol. 37, pp. 6885-6890.

Thawornwong, S. and Enke, D. (2004), “The adaptive selection of financial and economic variables for use with artificial neural networks”, *Neurocomputing*, Vol. 56, pp. 205-232.

Tjung, L.C., Kwon, O., Tseng, K.C. and Bradley-geist, J. (2010), “Forecasting financial stocks using data mining”, *Global Economy and Finance Journal*, Vol. 3 No. 2, pp. 13-26.

Tsai, C.-F. and Wang, S.-P. (2009), “Stock price forecasting by hybrid machine learning techniques”, *Proceedings of the International MultiConference of Engineers and Computer Scientists*, Vol. 1 No. 755.

Umbarkar, S.S. and Nandgaonkar, P.S.S. (2015), “Using association rule mining: stock market events prediction from financial news”, *International Journal of Science and Research (IJSR)*, Vol. 4 No. 6, pp. 1958-1963.

Upadhyay, A., Bandyopadhyay, G. and Dutta, A. (2012), “Forecasting stock performance in Indian market using multinomial logistic regression”, *Journal of Business Studies Quarterly*, Vol. 3, pp. 16-39.

Vaisla, K, S and Bhatt, A, K (2010), “An Analysis of the performance of Artificial Neural Network technique for stock market forecasting”, *International Journal of Computer Science and Engineering*, Vol. 2, pp. 2104-2109.

Vanstone, B., Finnie, G. and Hahn, T. (2010), “Stock market trading using fundamental variables and neural networks”, *Proceedings of ICONIP 2010: 17th International Conference on Neural Information Processing*, Information Technology Papers, School of Information Technology, Sydney, Australia.

Versace, M., Bhatt, R., Hinds, O. and Shiffer, M. (2004), “Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks”, *Expert Systems with Applications*, Vol. 27, pp. 417-425.

Vojinovic, Z., Kecman, V. and Seidel, R. (2001), “A data mining approach to financial time series modelling and forecasting”, *International Journal of Intelligent Systems in Accounting, Finance & Management*, Vol. 10, p. 225239.

Weiss, S.H. and Indurkhya, N. (1998), *Predictive Data Mining: A Practical Guide*, Morgan Kaufmann Publishers, San Francisco, CA.

Williams, B., Onsman, A. and Brown, T. (2010), “Exploratory factor analysis: a five-step guide for novices”, *Journal of Emergency Primary Health Care (JEPHC)*, Vol. 8, pp. 1-13.

Yu, L., Chen, H., Wang, S. and Lai, K.K. (2009), “Evolving least squares support vector machines for stock market trend mining”, *IEEE Transactions on Evolutionary Computation*, Vol. 13, pp. 87-102.

Yumlu, S., Gurgen, F.S. and Okay, N. (2005), “A comparison of global, recurrent and smoothed-piecewise neural models for istanbul stock exchange (ISE) prediction”, *Pattern Recognition Letters*, Vol. 26, pp. 2093-2103.

Zhai, Y., Hsu, A. and Halgamuge, S.K. (2007), “Combining news and technical indicators in daily stock price trends prediction”, *Proceedings of the 4th international symposium on Deep Neural Networks: Advances in Neural Networks, Part III*, Springer-Verlag, Nanjing, pp. 1087-1096.

Further reading

Shapiro, G.P., Frawley, W.J. and Matheus, C.J. (1992), “Knowledge discovery in databases: an overview”, *AI Magazine*, Vol. 13, pp. 57-70.

ANNEXURE I

List of Technical variables

Dependent variables	Formula
Stochastic %K	$\frac{C_t - LL_{t-n}}{HH_{t-n} - LL_{t-n}} \times 100$ where C_t is closing price, LL_t is lowest low and HH_t is highest high in t days
Stochastic %D	$\frac{\sum_{i=0}^{n-1} \%K_{t-1}}{\text{n}}$
Stochastic slow %D	$\frac{\sum_{i=0}^{n-1} \%D_{t-1}}{\text{n}}$
Momentum	$C_t - C_{t-n}$ where n=10, C_t is closing price today
ROC (rate of change)	$\frac{C_t}{C_{t-n}} \times 100$
LW %R (Larry William's %R)	$\frac{H_n - C_t}{H_n - L_n} \times 100$
A/D oscillator (accumulation/ distribution oscillator)	$\frac{H_t - C_{t-1}}{H_t - L_t} \times 100$
Disparity 5-days	$\frac{C_t}{MA_5} \times 100$ where MA_5 is 5-day moving average
Disparity 10-days	$\frac{C_t}{MA_{10}} \times 100$ where MA_{10} is 10-day moving average
OSCP (price oscillator)	$\frac{MA_5 - MA_{10}}{MA_5}$
CCI (commodity channel index)	$\frac{H_t + L_t + C_t - ADP_{t-1}}{0.015 \times AvgDev_{t-1}}$ where $ADP_t = \frac{\sum_{i=t-n+1}^{t}(H_i + L_i + C_i)}{n}$. $AvgDev_t = \frac{\sum_{i=t-n+1}^{t} \lvert H_i + L_i + C_i - ADP_{t-1})}{n}$
RSI (relative strength index)	$100 - \frac{100}{1 + RS}$ where $RS = \frac{AU}{AD}$ AU = the total of the upward price changes during the past 14 days AD = the total of the downward price changes (used as positive numbers) during the past 14 days

Corresponding author

Jasleen Kaur can be contacted at: jazz831@gmail.com