Journal of Organizational and End User Computing
Volume 33 • Issue 5 • September-October 2021

Deep Neural Network and Time Series Approach for Finance Systems:
Predicting the Movement of the Indian Stock Market

Praveen Ranjan Srivastava, Indian Institute of Management, Rohtak, India
https://orcid.org/0000-0001-7467-5500

Zuopeng (Justin) Zhang, University of North Florida, USA

Prajwal Eachempati, Indian Institute of Management, Rohtak, India

ABSTRACT

The stock market is an aggregation of investor sentiment that affects daily changes in stock prices. Investor sentiment remained a mystery and challenge over time, inviting researchers to comprehend the market trends. The entry of behavioral scientists in and around the 1980s brought in the market trading's human dimensions. Shortly after that, due to the digitization of exchanges, the mix of traders changed as institutional traders started using algorithmic trading (AT) on computers. Nevertheless, the effects of investor sentiment did not disappear and continued to intrigue market researchers. Though market sentiment plays a significant role in timing investment decisions, classical finance models largely ignored the role of investor sentiment in asset pricing. For knowing if the market price is value-driven, the investor would isolate components of irrationality from the price, as reflected in the sentiment. Investor sentiment is an expression of irrational expectations of a stock's risk-return profile that is not justified by available information. In this context, the paper aims to predict the next-day trend in the index prices for the centralized Indian National Stock Exchange (NSE) deploying machine learning algorithms like support vector machine, random forest, gradient boosting, and deep neural networks. The training set is historical NSE closing price data from June 1st, 2013-June 30th, 2020. Additionally, the authors factor technical indicators like moving average (MA), moving average convergence-divergence (MACD), K (%) oscillator and corresponding three days moving average D (%), relative strength indicator (RSI) value, and the LW (R%) indicator for the same period. The predictive power of deep neural networks over other machine learning techniques is established in the paper, demonstrating the future scope of deep learning in multi-parameter time series prediction.

KEYWORDS

Deep Neural Network, Gradient Boosting, NSE, Random Forest, Stock Market, Time-Series Prediction

DOI: 10.4018/JOEUC.20210901.oa10

1. INTRODUCTION

The stock market is an aggregation of sellers and buyers and serves as a platform for exchanging different companies' stocks. A group of companies constitutes a dedicated stock exchange index aggregated country-wise. Digitization of the stock exchange has facilitated algorithmic trading (AT) on state-of-the-art-machines. According to a World Economic Forum report, over 50% of the trades are on AT. Investors can now complete stock market transactions rapidly.

In this context, researchers are attempting to measure the implications of trading with modern techniques and tools like Artificial Intelligence and robotics in a market where human biases dominate. According to a report of the World Economic Forum, the global market is witnessing a new trend. Presently, 10% of trades are performed by retail traders without the involvement of AT, 40% trades are based on decisions to invest in stock market/index funds/ETFs, and the balance 50% trades are conducted on automatic trading using computers, also known as algorithmic trading (AT). Prediction of the futures markets in the context of AT is an exciting area for further research.

To predict stock market trends, we need to identify the determinants that cause daily stock price changes. One set of factors that reflect the movement in the stock market is technical indicators. These are leading indicators of variation in stock market trends since they perform mathematical transformations on historical stock prices.

In the context of algorithmic trading, many technological advances have taken place, leading to the development of robust algorithms. We implement these algorithms to determine whether the stock market follows an upward trend (bull phase) or tanks (bear phase).

The model enables real-time decision-making of whether to buy or sell or hold onto the stocks. Machine learning techniques are imperative to adopt in real-time decisions with a high level of confidence and accuracy. In machine learning (Demir et al., 2020; Gan et al., 2020), the computer systems are trained on previous instances (also called 'train set') and implicitly programmed to respond to similar real-time scenarios (test set) for decision-making. The machine learning algorithms consider the historical stock prices and leading technical indicators as factors and train the model for future trend predictions (Manickavasagam, Visalakshmi & Apergis, 2020).

The paper aims to develop a stock price prediction model from the above factors training using Support Vector Machine, Random Forest, Gradient Boosting, and Deep Neural Networks for the Indian stock market index (NSE NIFTY 50). We compare the algorithms' predictive performance and compute the variable importance to identify the most critical factors' index trend. The rest of the paper is structured as follows.

Section 2 reviews the literature on technical analysis, technical indicators, and machine learning techniques. The data and research methodology is in Section 3. Section 4 discusses the results of classifier performance and variable importance. Finally, in Section 5, we discuss the conclusions of the research followed by references.

2. LITERATURE REVIEW

2.1 Reading Investor Behavior Through Technical Analysis

Technical analysis is followed closely by professional stock market observers, including trained managers of mutual funds and other chartists who attempt to read the market pattern based on chart walks as observed from time to time or some statistical derivatives made from the daily market data. These methods constitute their instincts.

As stated earlier, technical analysts advocate that a market price reflects all relevant information, and thus their analysis starts at the history of a security's trading pattern, ignoring all other external drivers like economic, fundamental factors, and news events. This is based on the premise that all relevant information is already reflected and discounted in prices. Investors' response to changes in market price also tends to repeat itself because investors collectively lean towards patterned behavior in tune with their universal goal of accomplishing profit maximization. Investors follow all known and unknown market paths by deploying self-generated and borrowed biases and ideas, and the experience gained out of self-inflicted losses. This tendency alerts technicians' focus on identifiable trends and conditions. This aspect of investing is not captured in Cutler, Poterba, and Summers (1988).

Technical analysts (Piñeiro-Chousa, López-Cabarcos, & Ribeiro-Soriano, 2020) believe that current investors collectively repeat the investors' behavior that preceded them. This is captured in recognizable and predictable price patterns. Technical analysis is not just charting; it always involves charting and considers price trends and causality. Investor surveys and Sentiment analysis help monitor the direction of investor sentiment. Furthermore, they gauge the market participants' attitudes in all-weather conditions, whether they are bearish or bullish or in a non-committed mode. Technicians use these surveys to help determine the direction of the market trend from time to time. A change in the market trend predicted when the survey reports detect extreme positions of investor sentiment. If the surveys reflect bullishness fever, it shows that an uptrend after a point of time may reverse as investors have already bought out in anticipation of a further increase in prices and go into a cashless state with few buyers remaining. After that, a general feeling that the market turned costly engulfs the investor community. The feeling is reinforced when one studies fundamental indicators like PE multiple (exceeding typical pattern, turning the market either costly or cheaper) and PEG indicators (PE ratio/Earning growth rate is less than 1). This leaves leverage for potential sellers than buyers, amidst even the bullish sentiment, thus paving the way for prices to go down when supply exceeds demand. Here is an example of contrarian trading when an intelligent investor looks for ways and means to go against the herding effect to milk profits.

In light of the need for technical analysis, the following technical indicators are the following.

2.2 Technical Indicators

The technical indicators (Alonso-Monsalve *et al*., 2020; Dai *et al*., 2020) chosen for the study are defined below:

1. **Moving Average (Huang, 2020):** MV is the average of previous n day stock prices that normalizes the closing prices. Mathematically, moving average is:

$$MA\left(i\right)=1 / t * \sum tcp\left(i\right) \tag{1}$$

where cp(i) is the closing price for the i^{th} period, and 't' signifies the number of periods considered for the study.

For instance, moving average MA(3), i.e., the moving average for three periods, is [(1*Closing price of the first period) + (2* Closing price of the second period) + (3* Closing price of the third period)]/3; say the closing price of the first period is 12456, the closing price of the second period is 12348 and closing price of the third period is 12287, the moving average for the third period=[(1*12456) +(2*12348) +(3*12287)]/3= 24671.

2. **Moving Average Convergence/Divergence (Dai et al., 2020):** This technical indicator, denoted by MACD, shows the relationship between the exponential moving average (EMA) for the 12^{th} and the 26^{th} trading period (buy and sell points chosen by convention by traders) by examining the point of convergence or divergence. The formula measures this:

$$MACD = EMA\left(12\right) - EMA\left(26\right) \tag{2}$$

Exponential Moving Average (EMA) is a weighted moving average that assigns higher weightage to recent prices than past prices. This differs from the Moving Average, which assumes equal weightage for all prices.

Exponential Moving Average is computed iteratively from the past period (yesterday) and closing price by the following formula:

$$\text{EMA(i)=} \left(cp\left(i\right)-EMA\left(i-1\right)\right)*multiplier+EMA\left(i-1\right) \tag{3}$$

where:

$$\text{multiplier=}\ 2\,/\left(number\ \ of\ \ days\ \ considered\ \ for\ \ average+1\right) \tag{4}$$

and cp is the closing price for the ith period.

The multiplier is a weightage factor assigned to the recent past period 'i-1' to increase its prominence in determining the future movement.

For instance, in (2), the exponential moving average of the 12^{th} period is computed from (3) and (4):

EMA(12)= [(closing price in 12^{th} period)- EMA(11^{th} period)]*(2/13)+ EMA(11^{th} period)

where EMA(11) is iteratively computed in turn from (3) as:

EMA(11)= [(closing price in 11^{th} period)- EMA(10^{th} period)]*(2/12)+ EMA(10^{th} period)

and so, on till:

EMA(0) =0 and EMA (1) = closing price in 1^{st}period*(2/2) = cp(1^{st} period)

3. **K(%) Oscillator (Pramudya, 2020):** This oscillator (also known as an indicator) gives a measure of the price velocity providing the relative position of the current closing price in a particular period:

$$\text{K(\%)} = \left(cp\left(i\right)-Lt\right)/\left(Ht-Lt\right)*100 \tag{5}$$

where cp(i) is the closing price of the most recent period 'i,' and H_t and L_t represent the average high and low average prices of the last 't' periods.

For instance, to compute K for the 14^{th} period, the formula (5) is applied as:

K (%) = (closing price of 14^{th} period- an average of all low prices from 1^{st} to 14^{th} period)
/(Average of all high prices from 1^{st} to 14^{th} period- an average of all low prices from
1^{st} to 14^{th} period) * 100

4. **D(%) Indicator (Tan et al., 2020):** The 3-day moving average of the K(%) oscillator gives the D(%) indicator denoted by the formula:

$$\mathrm{D}(\%) = \left[K\left(\%\right)\left(i-2\right) + K\left(\%\right)\left(i-1\right) + K\left(\%\right)\left(i\right)\right] / \ 3 \tag{6}$$

where 'i' denotes the latest time period.

For instance:

D(%) for the 3rd period = [K% (1th period) + K% (2nd period)+ K% (3rd period)]/3

5. **Relative Strength Index (Tan, 2020):** It is a momentum indicator that measures the relative strength in terms of up and down patterns of the closing prices. Mathematically:

$$\mathrm{RSI}= \ 100 - \left[\frac{100}{1+RS}\right] \tag{7}$$

RS= average of 't' days up closing prices / Average of 't' days down closing prices

For instance:

RSI for 5 days =[100-100/(1+RS)]

where RS (5 days)= Average closing prices of 5 days (where closing prices have increased) divided by Average closing prices of 5 days (where closing prices have decreased from the previous period).

6. **Larry William's R% (Jiang et al., 2020):** This stochastic oscillator is the additive inverse of the K(%) oscillator measured by the formula:

$$\mathrm{R}(\%)= \left(Ht - cp\left(i\right)\right) / \left(Ht - Lt\right) * 100 \tag{8}$$

also denoted as (1-K(%)).

For instance, from the above example for K(%) oscillator:

R (%) = (Average of all the low prices from 1st to 14th period-closing price of 14th period)
/(Average of all the high prices from 1st to 14th period- an average of all the low prices
from 1st to 14th period) * 100

The technical indicators and their applicability are thus summarized in Table 1.

2.3 Rationale For Adopting Machine Learning Techniques

Initially, we adopt baseline machine learning techniques like Support Vector Machine to predict the stock price trend. Support Vector Machine (SVMs) is adopted to achieve higher generalization performance, as it utilizes an induction principle called structural risk minimization (SRM) principle (Chandaka, 2009; Longato et al., 2020).

The predictions are further validated by more robust machine learning algorithms Random Forest, a powerful ensemble predictor, and Deep Neural Networks (DNN).

Random Forest is an ensemble model that combines the predictive power of simple models and is applicable since it can be analyzed easily (Jiang et al., 2020; Kamarajan et al., 2020) is a robust predictor.

Table 1. Summary of Technical Indicators adopted in the study

Technical Indicator	Definition	Purpose of the indicator
Moving Average(MA)	MA is the average of previous 'n' day stock prices that normalizes the closing prices.	The indicator helps in identifying crossover of buy and sell points.
Moving Average Convergence/ Divergence (MACD)	MACD denotes the relationship between the exponential moving average (EMA) for the 12^{th} and the 26^{th} trading period by examining the point of convergence or divergence.	The indicator is used to find new trends and to signal the end of a trend.
K(%) Oscillator	This oscillator (also known as an indicator) gives a measure of the price velocity providing the relative position of the current closing price in a particular period.	The indicator is used to generate overbought and oversold trading signals.
D(%) Indicator	The 3-day moving average of the K(%) oscillator gives the D(%) indicator.	This indicator measures the momentum of the price movement in terms of 3 days moving average.
Relative Strength Index	It is a momentum indicator that measures the relative strength in terms of up and down patterns of the closing prices.	This indicator analyzes the bullish and bearish momentum patterns.
Larry William's R%	This stochastic oscillator is the additive inverse of the K(%) oscillator.	It is used to identify overbought/oversold levels, momentum confirmations as well as finding trade signals.

Neural networks are generally better at processing and providing a generic output (Zhou *et al*., 2002). Neural networks, incredibly deep neural networks (Loureiro *et al*., 2018), process the data through multiple hidden layers, considering which process the input features through an activation function that maps the semantic context essential to understanding the dynamic stock market situation. There is a need to store previous records of stock price trends to provide more accurate recommendations, which is enabled by the recurrent neural networks due to its memory gate that stores past information. Deep neural networks optimize each hidden layer's output through optimizers/ solvers, eliminating outliers to a great extent.

Further details about the rationale for adopting each machine learning algorithm in the paper is provided in Section 3.2.

2.4 Existing Studies in Machine Learning and Finance

The existing studies in machine learning and finance were initially reviewed below.

One of the first machine learning algorithms to be implemented in the stock market context is the Support Vector Machines(SVM). This technique is used in both classification and regression context for linear classifiers. Existing studies like Madge and Bhatt (2015) have predicted stock behavior for 34 technology sector stocks using SVM. The technical indicators like momentum and price volatility have been considered as individual predictors for the analysis. However, we record the low accuracy of 55-60% with only the two technical indicators as input. SVM algorithm has also been adapted to other variations like least square support vector machine (LSSVM) by Wang and Shang (2014) for predicting the China Security Index 300b (CSI 300) using ten technical indicators, which boosted the accuracy and was easier to train.

Further, ensemble techniques were adopted for stock price prediction like Random Forest and Gradient Boosting. Random Forest was adopted by Tratkowski (2020) for predicting the stock market in terms of formulating long-term investment strategies. It outperforms other baseline classifiers.

Gradient Boosting was adopted by Nabi (2020) to predict the stock market movement with a feature engineering variation (GB-FE) from historical data for the Nasdaq and S&P500. A mean square error (MSE) of 0.041% is attained, proving its effectiveness.

The need for higher predictive accuracy and an algorithm that stores previous patterns for better pattern recognition prompted research using the Deep Learning architecture. We use Deep learning in an economic context in some of the following studies.

The application of Deep Learning to analyze time-series data of the stock market was in Takeuchi & Lee (2013), who applied deep learning to enhance momentum trading strategies in stocks. An autoencoder composed of stacked RBMs is used to extract features from the history of individual stock prices. Data on individual US stocks from CRPE prices to NYSE shares from January 1965 to December 1989 and feature extraction of 12 monthly returns and 20 day-wise returns and compute returns for DEC10 and DEC 1 stocks for momentum effect validation. Vengertsev(2014) deployed a Deep Learning architecture for univariate time series forecasting by pre-training with continuous (RBM) and Conditional RBM.

Batres-Estrada (2015) implemented deep learning with a Theano backend in Python. A Deep Belief Network with a Multi-layer Perceptron (MLP) was trained for sector-wise stock selection and short-term predictions and regression. Financial services firms use robotic advisors who use Artificial Intelligence techniques to form portfolios (Heaton, Polson, & Witte, 2017).

Lee (2020) utilized Long Short Term (LSTM) deep neural networks with only technical indicators to predict the prices for the Taiwan Stock Exchange (TSE) and could attain 75% prediction accuracy. Parray (2020) also incorporated technical indicators and historical price data to predict the stock prices for the NSE NIFTY50 using neural networks and could predict up to 76% accuracy.

The limitations of the existing studies are thereby discussed and reviewed in Section 2.5.

2.5 Limitations of Prior Studies

First, the existing studies like Wang and Shang (2014) and Tratkowski (2020) have investigated the impact of technical indicators on stock price independently but not in combination with historical stock prices. Technical indicators help provide an estimation of the price trend but, by itself, are not sufficient to forecast the future price trend accurately. There is also a need to know the previous stock prices better to understand the relative movement of the future price movement. For instance, if a technical indicator like Moving Average predicts the stock price to increase by ten basis points purely based on the Moving Average indicator’s absolute values, this only presents a one-dimensional market trend. However, the stock price movement may, on the contrary, decrease by ten basis points. This is because stock prices are not governed by a single factor but by different forces acting from different directions. The net impact of these different market forces determines the accurate stock price movement. This multi-dimensional perspective can be presented by factoring in historical stock prices. In the above example, though technical indicators predicted the stock price to increase by ten basis points, the historical stock prices from previous periods showed a downward trend. This proves that the historical prices were a more prominent force than the current technical indicators and thus brought down the net stock market index by ten basis points due to the negative momentum generated. Thus, technical oscillators or indicators incorporated additionally with historical prices help provide a better idea of the future direction of stock price movement and, thus, boost the predictive accuracy in determining the correct trend.

Second, though Parray (2020) used artificial neural networks for stock market forecasting in the Indian stock market, deep neural networks and the comparison of their predictive power viz-viz machine learning techniques were not performed in the Indian stock market context. The predictive and decision-making power of deep neural networks from historical data and oscillator history has not been established.

Lastly, the variable importance has not been computed across machine learning models in all the above existing studies (Madge & Bhatt, 2015; Lee, 2020) to determine the significance of technical indicators and technical analysis for stock market movement. Moreover, the variable importance is not further validated from a statistical data modeling perspective. This would help understand the significance of the predictor variables to make real-time decisions about which technical indicators are essential to retaining to discard indicators that do not impact the stock market movement.

For overcoming the above limitations, a stock price prediction model using machine learning techniques trained on historical data and technical indicators is proposed in Figure 1 to compare the predictive power of state-of-the-art machine learning techniques and variable importance.

3. RESEARCH METHODOLOGY

3.1 Data Collection

The data from the NSE website of 8 years was considered in the post-digitization period from January 1st, 2013, up to June 30th, 2020. The data contained open, close, high, and low prices; however, the closing prices were then considered for subsequent analysis. The technical indicators explained above in Section 2.2 were then computed from the closing prices, i.e., Moving Average (MA), Moving Average Convergence-Divergence (MACD), K(%) oscillator and its corresponding 3-day moving average D(%), Relative Strength Indicator(RSI) value and the LW(R%) indicator.

The working procedure of the machine learning algorithms is illustrated in Figure 2.

From Figure 2, the working algorithm of the machine learning techniques is detailed.

Random Forest is a group of decision trees that computes each tree's best split and aggregates each tree's output by a majority voting procedure. This procedure ensures that the majority class upvoted by the decision trees is chosen. The model with the least prediction error (computed by Root Mean Square Error) is chosen as the optimal model. Gradient Boosting is an ensemble machine learning technique that combines a robust classifier (like Random Forest) and a weak classifier (Decision Tree) to train the data and ensure the least prediction and overfit error to predict the stock price trend.

Figure 1. Research Gap

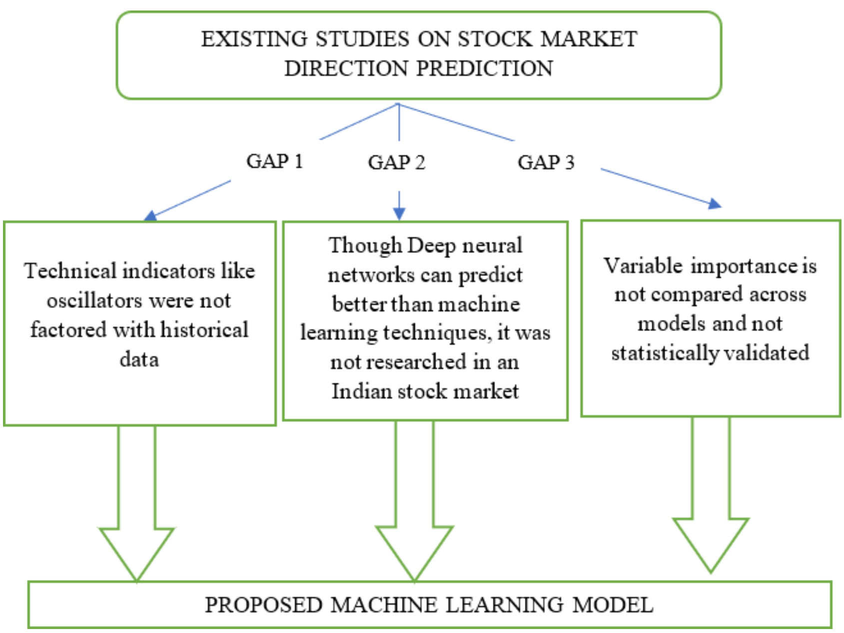
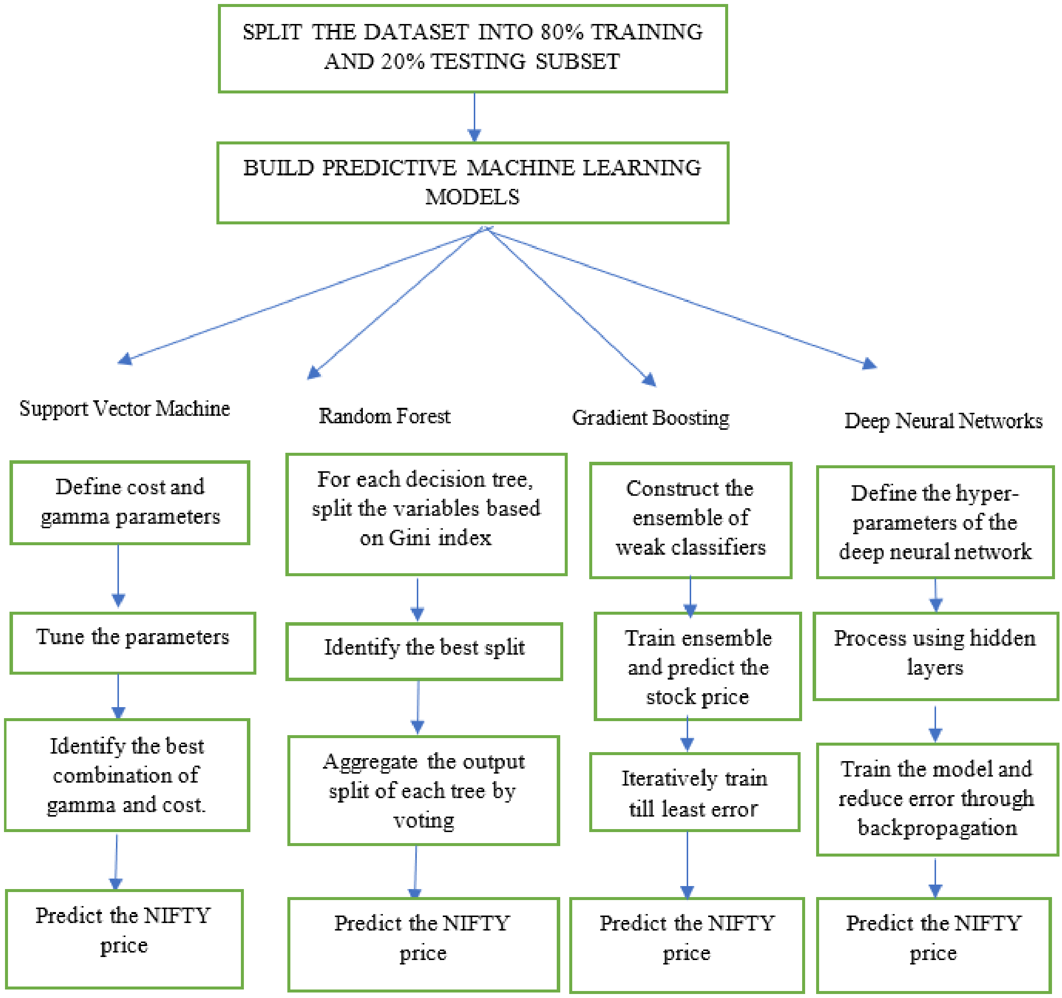


Figure 2. Working of the ML predictive models



Deep Neural Networks processes the data with multiple hidden layers by applying a suitable activation function on the input normalized data and minimizing the error through a backpropagation algorithm. Hyperparameters like hidden layers, choice of the optimizer, choice of activation function, and learning rate (for minimizing error) are tuned to get the optimal accuracy model.

The proposed model is discussed below.

3.2 Proposed Model For Workflow

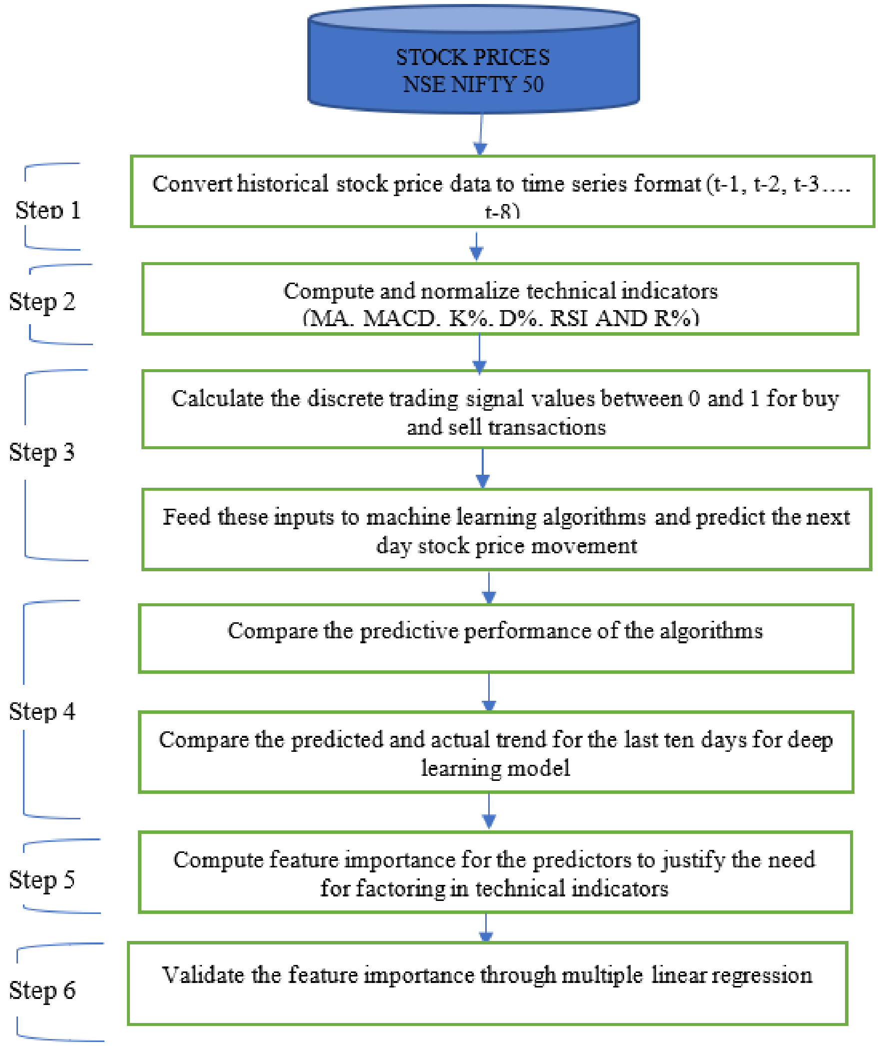
The following research methodology was adopted using machine learning techniques to determine the next day trend from technical indicators and other historical data (see Figure 3).

In Figure 3, the following steps were undertaken:

Step 1: Collect historical data of NIFTY 50.

The stock price data collected from the NSE website is first converted to a time series format using Python's pre-defined functions. From January 1st, 2013 to June 30th, 2020, the daily data is

Figure 3. Research Methodology adopted in the study



converted to time series of 't-1' to 't-8' where each time series is the daily data for the particular year. For instance, the time series data for the previous year (2019-2020) contains daily closing price data from June 30th, 2019 to June 30th, 2020, and similarly, for other years, each data series is considered an input feature for machine learning models.

Step 2: Computing the technical indicators for the above historical data series.

The following six technical indicators are computed using the above Equations (1) to (8). Since all the technical indicators are in different ranges and are used to predict the price trend measured in the range of 0 to 1 (0 representing negative trend and 1 representing positive trend), all the indicators are first normalized to the same range of 0 to 1 for more accurate prediction before they are fed as input to the machine learning algorithm. The indicators are: Moving Average (MA), Moving Average Convergence-Divergence(MACD), K(%) oscillator, its corresponding three-day moving average D(%), Relative Strength Indicator(RSI) value, and the LW(R%) indicator.

Step 3: Calculate the discrete signal values in the same range (0-1).

The discrete trading signals(TS) in the same range of 0-1 for representing sell or buy (bias of 0.5 added) respectively are also computed to be fed to the baseline machine learning models and deep neural network models from which the trend is predicted.

Step 4: Compare the predictive performance of machine learning classifiers.

The predictive performance of the classifiers Support Vector Machine, Random Forest, Gradient Boosting, and Deep Neural Networks is compared on the test set in terms of index prices. The actual trend and predicted trend for the deep learning model are compared for the last ten days.

Step 5: Compute the variable importance and compare it across classifiers.

The predictors' feature importance (historical data and technical indicators) are also computed and compared across the classifiers to determine and establish technical indicators in determining market movement.

Step 6: Formulate a multiple linear regression model to validate whether the same variables are significant. The results of the regression model are illustrated in sub-section 4.5.

The above-proposed model in Figure 3, implementing the working procedure in Figure 2, computes the predictive performance and variable importance across all the machine learning classifiers to predict the stock prices and stock price movement for the next 1-day period.

The detailed implementation of the machine learning techniques and rationale are discussed as follows:

- **Random Forest-based predictive model (Bendazzoli et al., 2019):** The Random Forest is an ensemble learning technique that builds on a simple decision tree's functionality by aggregating multiple decision trees' results a voting rule (Qamar *et al*., 2016). It has two main advantages: one, the random forest resamples the training data with replacement and reduces the variance in classification and restricts the number of features in a tree, thus providing a more optimized variable importance result. However, in unbalanced dataset scenarios, there is a need to use under-sampling or oversampling or hybrid techniques to balance the dataset, and it tends to cause over-fitting in specific scenarios where data is not aligned with the problem domain. For this model, all predictors are converted to numeric, and the target variable, *i.e*., the rating, is predicted, as shown in section 4.1.
- **Support Vector Machine (Pradhan & Sameen, 2020):** The support vector regression method is also adopted to achieve linear separability and predict the ratings. The results of this model are given in section 4.2.
- **Gradient Boosting predictive model (Hubáček et al., 2019):** In this paper, Gradient Boosting is adopted due to its built-in ability to handle unbalanced datasets. The gradient boost ensemble model is also run to predict rating to boost the predictive accuracy and interpretability. For this model, all predictors are converted to numeric, and the output variable, *i.e*., the rating, is predicted, as shown in section 4.3.
- **Deep Neural Network-Based Model (Hassanpour et al., 2019):** A Deep Neural Network (DNN) model simulates the inner workings of a human brain. A typical architecture is layer-wise: the input layer takes the normalized input data, and the last layer provides the output. In between, the processing of inputs is performed in hidden layers (one or more), which process the input values

and compute an activation function (preferably sigmoid) based on the importance of variables. A DNN is trained to train the input weights and generate the output incrementally. DNNs effectively handle large numbers of multiple input variables, for example, in big data scenarios. In this paper, a DNN has been constructed with three hidden layers due to the minimum root mean square error (RMSE) of 0.07. This also minimizes the probability of over-fitting the deep neural network model, i.e., the model fitting only on some data points and under-performing on other data points (Loureiro *et al*., 2018). The number of input nodes is six, considering six individual predictors. The variable importance of the predictors and the model's performance measured in terms of accuracy are compared and illustrated in sub-sections 4.4.

The reasons for this multi-stage approach (validation of the machine learning techniques through multiple linear regression) are as follows:

1. The model building phase aims to predict the rating by multiple layered deep neural networks.
2. Though machine learning techniques compute the variable importance, there is a need to confirm these variables' significance to provide more personalized and reliable recommendations to stakeholders. In this context, a valid mechanism to confirm the significance is the multiple regression techniques.

The results of the study are discussed below.

4. RESULTS AND DISCUSSION

The Results section compares the machine learning classifiers' predictive performance from the above inputs (historical data, technical indicators, and discrete trading signal values) as follows.

The dataset format fed as input to the machine learning classifiers is shown in Table 2.

The parameter Trading Signal (TS) is a derived variable that is also incorporated in the dataset to analyze the impact of the trading signal (buy/sell decision) on the NIFTY price in period 't'. The TS indicates both an improvement in market trend (uptrend) and downward movement in trend (downtrend) computed [as outlined above in Step 3] by the following equations.

Table 2. Input Normalized Sample Dataset

NIFTY t-8	NIFTY t-7	NIFTY t-6	NIFTY t-5	NIFTY t-4	NIFTY t-3	NIFTY t-2	NIFTY t-1	MA	MACD	K (%)	D (%)	RSI	LW (R%)	TS
5993.25	6221.15	8395.45	7791.30	8179.50	10442.20	10792.50	12182.5	0.00	0.01	0.39	0.37	0	0.61	0.05
6009.50	6211.15	8378.40	7784.65	8192.25	10443.20	10672.25	12282.2	0.01	0.01	0.41	0.40	0	0.59	0.50
6016.15	6191.45	8127.35	7741.00	8190.50	10504.80	10727.35	12226.65	0.01	0.02	0.48	0.42	0	0.52	0.00
5988.40	6162.25	8102.10	7568.30	8273.80	10558.85	10771.80	11993.05	0.01	0.03	0.44	0.44	0	0.56	0.00
6001.70	6174.60	8234.60	7601.35	8243.80	10623.60	10802.15	12052.95	0.02	0.03	0.43	0.44	0	0.57	0.00
5971.50	6168.35	8284.50	7563.85	8236.05	10637.00	10855.15	12025.35	0.02	0.05	0.46	0.44	0	0.54	0.14
5968.65	6171.45	8323.00	7510.30	8288.60	10632.20	10821.60	12215.9	0.02	0.05	0.46	0.44	0	0.54	0.01
5951.30	6272.75	8299.40	7562.40	8380.65	10651.20	10794.95	12256.8	0.03	0.05	0.48	0.46	1	0.52	0.50
6024.05	6241.85	8277.55	7536.80	8407.20	10681.25	10737.60	12329.55	0.03	0.05	0.45	0.46	0	0.55	0.50
6056.60	6320.90	8494.15	7437.80	8400.35	10741.55	10886.80	12362.3	0.04	0.06	0.56	0.50	0	0.44	0.17
6001.85	6318.90	8513.80	7351.00	8412.80	10700.45	10890.30	12343.3	0.05	0.06	0.50	0.51	0	0.50	0.00

- For uptrend:

Trading signal (TS)= $\left(cp - min\left(cp\right)\right) / \left(max\left(cp\right) - min\left(cp\right)\right) * 0.5 + 0.5$ (9)

where cp=Closing Price in period 't-1' (NIFTY $_{t-1}$).

- For downtrend:

TS = $\left(cp - min\left(cp\right)\right) / \left(max\left(cp\right) - min\left(cp\right)\right) * 0.5$ (10)

where:

cp= $NIFTY_{t-1}$

The dataset contains daily historical prices for each period ($NIFTY_{t-8}$ to $NIFTY_{t-1}$), normalized technical indicators, and normalized trading signals (both normalized between 0 and 1).

For instance, in Table 2, 2nd row, the closing prices for $NIFTY_{t-8}$ (January 2nd, 2013) is Rs.6009.50, $NIFTY_{t-7}$ (January 2nd, 2014) is Rs.6009.50, $NIFTY_{t-6}$ (January 2nd, 2015) is Rs.6211.15, $NIFTY_{t-5}$ (January 2nd, 2016) is Rs.8378.40, $NIFTY_{t-4}$ (January 2nd, 2017) isRs.7784.65, $NIFTY_{t-3}$(January 2nd, 2018) is Rs.8192.2, $NIFTY_{t-2}$(January 2nd, 2019) is Rs.10443.20 and $NIFTY_{t-1}$ (January 2nd, 2020) is Rs.12282.2. The normalized technical indicators and trading signal (TS) values are: MA(0.01), MACD (0.01), K(%) (0.41), D (%) (0.40), RSI(0), LW(R%) (0.59) and TS (0.50).

The dataset containing 1725 data points for the period from January 1st, 2013 to June 30th,2020 (7.5 years and assuming 230 days as the trading period) is split into 80% for training (1380 data points) and 20% for testing (345 data points) before running machine learning algorithms.

4.1 Results Based on a Support Vector Machine

Initially, the support vector machine (SVM) algorithm was adopted on the dataset. The SVM was implemented using a k-fold cross-validation method with five splits to test the classifier's effectiveness. The number of splits vary according to data size and should not be too small or too large. Therefore, five splits are chosen for boosting performance. The performance metrics were computed as an average of the five splits (Fang et al., 2020).

The accuracy of the Support Vector Machine classifier was found to be 86%, indicating that 86% of the predicted stock prices and stock price movement corresponded to the actual price movement. Additionally, the precision 0.85, which indicates that 85% of the stock prices predicted to increase also increased in actual. This indicates that stocks that increased as predicted can be filtered and identified for transactions, while stocks that did not increase in reality (15%) were discarded from future analysis.

The recall was found to be 0.82, which implies that 82% of the data points (data points which were predicted to increase increased while those predicted to decrease decreased in actual) were found relevant (i.e., increased as expected). A model with high recall is useful in this context to identify selling points since points where stock prices increases are indicative of a bull market which motivate traders to buy and sell more frequently than hold onto their stocks.

4.2 Results Based on Random Forest

Another algorithm considered in this work for the prediction of the stock market price movement is Random Forest. Outcomes of the random forest-based model generated using the 'randomForest' package of the R tool are in Table 1. The parameters, *i.e*., the number of predictors 'mtry' and the number of trees 'ntree,' are tuned to choose the best model. A sample of the tuning dataset is in Table 3.

The 'mtry' parameter denotes the number of parameter splits is varied from 1 to $n-1$. The number of optimal predictors is considered $n/3$, where n is the number of variables considered in the model (Adam *et al*., 2014), mtry values vary from 1 to 5 due to 6 predictors adopted.

At the optimal accuracy of 93.4%, 93% precision, and 89% recall, the corresponding 'mtry' value achieved=2, which is in line with the theoretical proposition that optimal mtry ~ = number of predictors/3, which in this case = (6/3 ~=) 2.

This signifies that random forest could predict more accurately than the support vector machine (baseline classifier) due to its generalizability and majority voting rule applicable in decision-making scenarios.

4.3 Results Based on Gradient Boosting

Another algorithm considered in this work for the prediction of the stock market price movement is Gradient Boosting. The ideal size of the classification tree is n_tree = 100. This is an improvement over the Random Forest model due to an ensemble of techniques and aggregation of output from multiple decision trees. The Root Mean Square Error (RMSE) is reported in Table 4.

An accuracy of 97.8% was achieved with 98% precision and 92% recall. This signifies that ensembling through gradient boosting further improved the prediction accuracy due to the combination of multiple classifier outputs.

4.4 Results Based on Deep Neural Network

The DNN based model performance is illustrated below. K-fold cross-validation of 5 folds is implemented using Grid Search, which iteratively searches and identifies the best parameters with the highest accuracy. The perceptron is trained for 200 iterations, and the RELU activation function with Adam Optimizer is found to be the best hyperparameters.

Based on Figure 4, the minimum weight decay of 0.1, and the minimum cross-validation RMSE is attained at hidden layer=5, and therefore, five hidden layers have been adopted in the deep neural network model.

An accuracy of 98.7% was achieved with 99.5% precision and 93% recall. This signifies that ensembling the addition of hidden layers for training the model contributing to better understanding the stock price trend and proving to be more effective in identifying instances where 'buy and sell' transactions are possible with higher confidence.

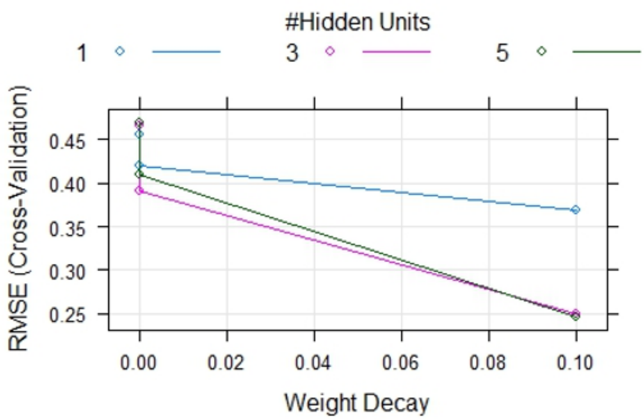
Table 3. Predictive Accuracy of the Random Forest Model

mtry	ntree	RMSE
2	100	6.45
2	200	6.28
4	500	7.56

Table 4. Tuning parameters of the Gradient Boosting Model

n_trees	RMSE
50	198.34
100	120.5
150	176.49

Figure 4. Neural Network RMSE v/s Weight Decay Curve



The performance metrics (accuracy, precision, and recall) are compared and summarized for each classifier in Table 5.

Further, the actual NIFTY prices and predicted NIFTY prices by all the machine learning algorithms are compared in Figure 5.

From Table 5 and Figure 5, it is evident that Deep Neural Networks outperforms the baseline Support Vector Machine and the ensemble algorithms(Random Forest, Gradient Boosting, and Deep Neural Network) in terms of all three metrics (Accuracy, Precision, and Recall). Further, the Deep Neural Network algorithm, the stock price direction is also predicted. This trend is compared with the observed trend as a validation of predictive performance. Therefore, the observed trend and predicted trend from June 21st to July 1^{st} are compared to determine and provide recommendations regarding 'buy and sell' for the test set for the Deep Neural Network model (see Table 6).

Here, the predicted trend is DOWN if the Predicted signal is less than 0.5 and UP if the signal is more significant than 0.5. The Trend Coincidence that maps both observed and predicted trends are found to be 100%, which confirms that deep neural networks outperform other classifiers in determining stock trends.

The relative variable importance is summarized across all the four machine learning models illustrated in Figure 6.

The top 5 predictors for all the four machine learning algorithms are found to be NIFTY(t-1) [previous period price], NIFTY(t-2)[next to previous period price], Moving Average(MA), K%, and Moving Average Convergence and Divergence (MACD).

4.5 Analysis of Multiple Linear Regression-Based Models

The variation in variable importance from the above importance results of the above machine learning techniques is examined by validating the regression models' results in Figure 7.

Table 5. Performance Comparison of Machine learning Models

Algorithm	Accuracy(%)	Recall	Precision
Support Vector Machine	86	0.85	0.82
Random Forest	93.4	0.93	0.89
Gradient Boosting	97.8	0.98	0.92
Deep Neural Network	98.7	0.995	0.93

Figure 5. Comparison of Machine learning techniques

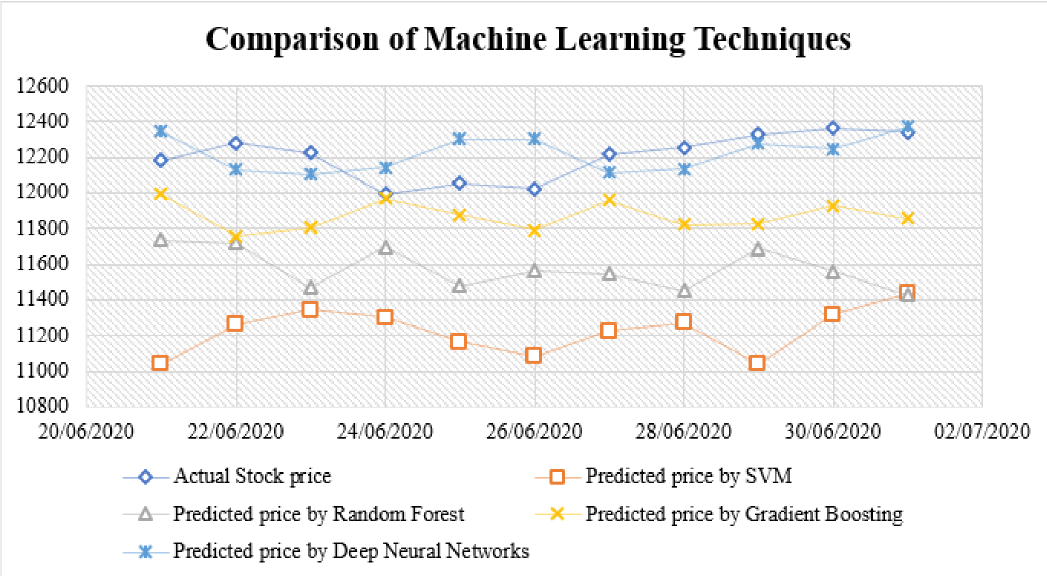


Table 6. Trend and Trading Signal Predicted Sample

Predicted Trend	Observed Trend	Trading Points	Trend Coincidence
1	1	SELL	CORRECT
0	0	HOLD	CORRECT
0	0	HOLD	CORRECT
1	1	SELL	CORRECT
1	1	SELL	CORRECT
1	1	SELL	CORRECT
1	1	SELL	CORRECT
1	1	SELL	CORRECT
1	1	SELL	CORRECT
0	0	HOLD	CORRECT
0	0	HOLD	CORRECT
0	0	HOLD	CORRECT

Regression Analysis is built by factoring in the above variables considered for NIFTY price prediction. The regression model ensures that all the robustness tests for the assumption of linear regression, namely multi-collinearity, linearity, auto-correlation and homoscedasticity are validated (Abadie et al., 2020). The robustness tests, namely Durbin-Watson, Langrange Multiplier (LM Coefficient), and Variance Inflation Factor (VIF) are run to ensure the reliability of the model variable significance (Yin et al., 2020).

The regression model with the six inputs is given in Table 7.

Figure 6. Relative importance across machine learning models

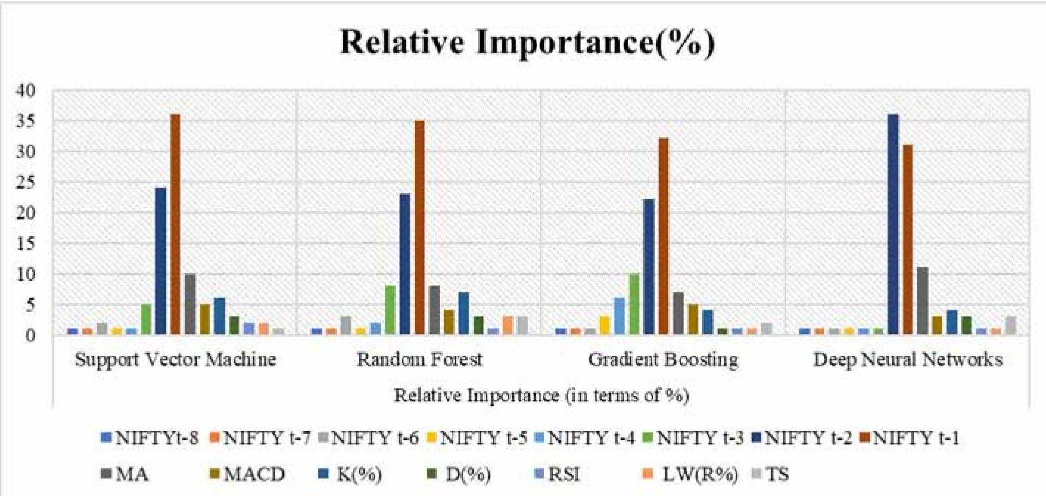


Table 7. Linear Regression model (Adj. R-Squared=0.91)

Variable	Coefficients	Standard Error	p-value	VIF
Intercept	-6651.238895	5546.99069	0.233	
NIFTY t-7	0.893889837	0.308266202	0.545	2.06
NIFTY t-6	0.433552504	0.280415504	0.125	1.12
NIFTY t-5	-0.773748336	0.235525407	0.137	1.54
NIFTY t-4	-0.171276517	0.337270608	0.613	2.97
NIFTY t-3	1.293658798	0.305153749	0.471	1.28
NIFTY t-2	1.257967935	0.227639282	0.002	1.31
NIFTY t-1	-0.586156295	0.246967039	0.001	1.04
NORMALIZED MA	6504.245187	4200.293283	0.001	1.01
NORMALIZED MACD	-14486.15783	4338.645285	0.005	1.63
NORMAL K(%)	-610.1465008	588.1184862	0.003	1.04
NORMALIZED D(%)	3300.581013	588.2328341	0.000	1.2
NORMALIZED RSI	-27.50150747	191.644151	0.886	3.26
NORMALIZED LW(R%)	0.45	0.125	0.768	3.47
DW Statistic	1.91(0.56)			
LM Statistic	2.21(0.64)			

At a 95% significance level, NIFTY t-2, NIFTY t-1, NORMALIZED MA, NORMALIZED MACD, NORMAL K(%), and NORMALIZED D(%) [highlighted in red color] are the most significant technical indicators based on coefficient and p-value. The value of adjusted R-squared is 0.91, i.e., 91% is explained.

According to the Durbin Watson test's thumb rule, the value of the Durbin Watson statistic (DW) must lie between 2 and 4 with a value tending closer to 2, implying that auto-correlation is not present in the dataset. Moreover, the significance value rho must be closer to 0. In this context, the Durbin Watson Statistic (DW) is reported to be 1.91 with a p-value of 0.56. The significance value of rho is 0.001. Both these statistics imply that there is no presence of autocorrelation in the dataset.

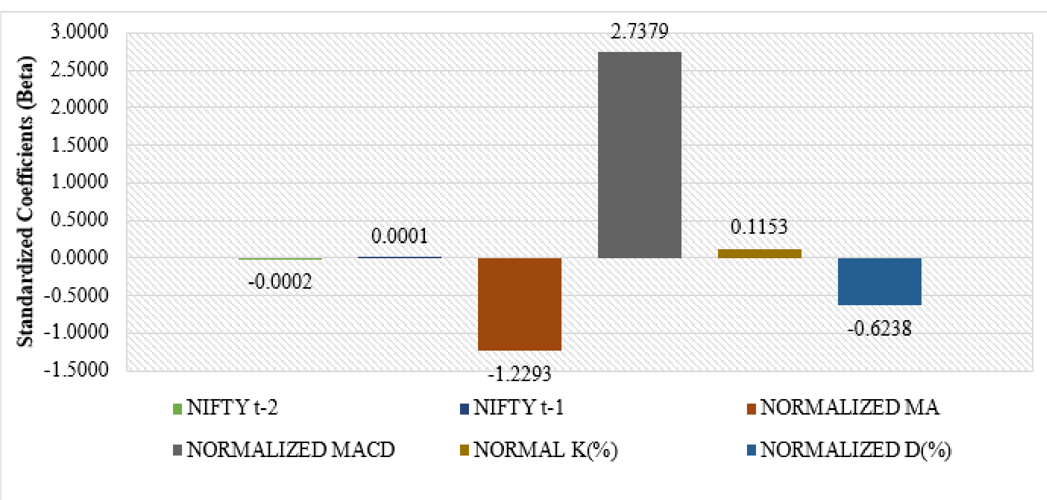
Similarly, for the Langrange Multiplier (LM) test, if the p-value statistic is greater than the level of significance 'alpha', the null hypothesis of homoscedasticity is validated. The Langrange Multiplier (LM) is reported to be 2.21 with a p-value of 0.64, greater than the level of significance alpha=0.05 (5% significance). This implies that the dataset is homoscedastic. The VIF (Variance Inflation Factor) for all the predictors is found to be <10, indicating that there is no multi-collinearity in the data.

Figure 7 depicts the most significant (in terms of p-value and coefficients) features for the dataset to predict the NIFTY price for the period 't'.

For the feature contribution chart, the significant variables' standardized coefficients are considered by normalizing the original coefficient. This normalization is performed by dividing the original coefficient by the sum of all the significant variable model coefficients (highlighted in Table 6 in red) and multiplying by 100. For instance, the original coefficient of $NIFTY_{t-1}$ is -0.586 in Table 6. The sum of all significant variable coefficients (all significant direct effect variables)=1.257967935-0.586156295+6504.245187-14486.15783-610.1465008+3300.581013= -5291. The standardized feature coefficient for variable 'Background' is thus, normalized and plotted above in Figure 7 as = -0.586/-5291=0.0001. Similarly, all standardized coefficients are plotted in Figure 7. This standardization is performed for clearer plotting to scale in this scenario.

In Figure 7, the presence of the factor of the NORMALIZED MACD(Moving Average Convergence and Divergence) (grey color) and NORMAL_K% (mustard color) is found to be significant positive drivers of NIFTY price with respective standardized coefficients 2.737 and 0.115. This implies that keeping other factors constant, one unit of increase in both NORMALIZED MACD and NORMAL_K% can increase in 2.737 units and 0.115 units increase in NIFTY price. On the other hand, NORMALIZED MA (Moving Average) (colored dark brown) is found to be a significant negative driver of NIFTY price, indicating that keeping other factors constant, one unit of increase in MA can lead to a -1.2293 decrease in a unit of NIFTY price.

Figure 7. Feature contribution chart from the Multiple Linear Regression Model



5. DISCUSSION

5.1 Theoretical Implications of the Study

This study makes the following four contributions to the literature: First, this study extends the literature in stock price prediction by factoring in multiple technical indicators with historical stock market data and deploying state-of-the-art machine learning techniques to enhance the prediction accuracy.

Second, in this study, a comparative analysis of random forest, state of the art ensemble ML techniques like Gradient Boosting, and complex Deep Neural Networks have been performed because of their predictive accuracy. It was found that DNN outperforms the other two algorithms (random forest and gradient boosting). Thus, DNN can be recalibrated for different datasets to predict other countries' stock markets in real-time.

Thirdly, the predictors' feature importance is also computed and compared to identify the significant drivers' stock price movement. This is also useful to determine the weightage of historical data and price indicators to identify which technical indicators are significant contributors to the stock price. The Moving Average(MA), K% oscillator, and Moving Average Convergence and Divergence (MACD) are major technical indicators.

Fourthly, the feature importance is validated by a multiple linear regression model, and it is found that $NIFTY_{t-2}$, $NIFTY_{t-1}$, NORMALIZED MA, NORMALIZED MACD, NORMAL K(%), and NORMALIZED D(%) are found to be significant. Interactions between parameters were not considered since the technical indicators are inter-related, and interaction variables would increase the dataset's multi-collinearity and reduce the predictive performance.

5.2 Implications for Practice

The output from the study accurately predicts the stock index price and movement in the paper. The implications for practice are three-fold: first, for the management; second, for traders; and third, investors.

5.2.1 Managerial Implications

The stock market prediction model enables the management to gauge the stock market trend and take calculated decisions regarding whether to buy and sell shares or to hold onto the shares until the market position improves. Also, the Moving Average(MA), K% oscillator, and Moving Average Convergence and Divergence (MACD) are vital technical indicators which are a good proxy for stock market direction and need to be analyzed in detail. The predictive power of deep learning prompts research and development teams in companies to formulate and tune the deep neural network models with different hyperparameters like activation function, the number of hidden layers, nature of the optimizer, and different learning rates on the live indicator and historical price data to improve the accuracy.

Thus, by generating automated "bots" that run these simulations and makes 'buy or sell' decisions, irrespective of whether the traders are active on the stock terminal, the study can contribute to the realistic corporate situation and would be a useful tool for companies to restore efficiency in the market. Future researchers can also exploit the advantages of transfer learning (a process of applying the same deep learning algorithm on different stock market scenarios even outside the Indian stock market context) and generate insights for each of these scenarios in future research.

5.2.2 Implications for Traders

The factors NIFTY(t-1)[previous period price], NIFTY(t-2)[next to previous period price], Moving Average (MA), K%, and Moving Average Convergence and Divergence (MACD) were found to be the most critical features simultaneously contributing to the stock price. This finding can be taken into account by the stock market traders to analyze in greater detail the historical stock market data more recent and closer to the current data. The above technical indicators MA and MACD, are to be analyzed in greater detail to devise live trading crossover strategies. The deep learning model can be simulated and back-tested on live trading strategies to recalibrate and tune the model for achieving better returns. Thus, the historical data and technical indicators are both essential for decision-making. Not only the accuracy but also the deep learning model can be tuned to achieve a high recall rate, which helps in identifying relevant performing stocks to prioritize trading.

5.2.3 Implications for Potential Investors

The model enables potential investors to know the stock market performance from the historical data and technical indicators in real-time. The variable importance results are also significant for the investors to track the previous two periods of stock prices and the Moving Average indicators. K(%) oscillator data is critical for determining the stock price trend.

The deep learning model can also be recalibrated by informed investors to make dynamic buy-sell decisions and be watchful when the stock market is in a bear phase. Also, identifying high performing stocks enables investors to choose suitable portfolios for long term investment.

6. CONCLUSION

The paper proposes a machine learning approach to predict the next day trend of NSE NIFTY prices. Eight years of historical NIFTY50 data with technical indicators were factored in to train the model. The deep neural networks approach was found to outperform the machine learning and ensemble classifiers. The variable importance computed also validated the hypothesis that technical indicators combined with historical indicators help predict the stock market trend closer to the actual trend. The multiple regression models have been used to validate the variable importance computed above, and it is found that the previous two periods NIFTY prices (NIFTY t-1, NIFTYt-2), Moving Average (MA), Moving Average Convergence Divergence (MACD), and K(%) oscillator are the most critical factors determining stock market movement.

The model can be extended by adding more historical data and computing more technical indicators as features. Further, the deep neural network model can be recalibrated, and transfer learning can be applied to train the data of other stock exchanges to replicate the above results. The accuracy of the stock market prediction can be further enhanced by incorporating the stocks' behavioral dimension in the form of textual sources like news and disclosures. These sources help identify investor sentiment and investor reactions to the stock market, which also predict the accurate stock price and direction.

REFERENCES

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-Based versus Design-Based Uncertainty in Regression Analysis. *Econometrica*, *88*(1), 265–296. doi:10.3982/ECTA12675

Alonso-Monsalve, S., Suárez-Cetrulo, A. L., Cervantes, A., & Quintana, D. (2020). Convolution on neural networks for high-frequency trend prediction of cryptocurrency exchange rates using technical indicators. *Expert Systems with Applications*, *149*, 113250. doi:10.1016/j.eswa.2020.113250

Bendazzoli, S., Brusini, I., Damberg, P., Smedby, Ö., Andersson, L., & Wang, C. 2019. Automatic rat brain segmentation from MRI using statistical shape models and random forest. In Medical Imaging 2019: Image Processing (Vol. 10949, p. 109492O). International Society for Optics and Photonics. doi:10.1117/12.2512409

Chandaka, S., Chatterjee, A., & Munshi, S. (2009). Cross-correlation aided support vector machine classifier for the classification of EEG signals. *Expert Systems with Applications*, *36*(2), 1329–1336. doi:10.1016/j.eswa.2007.11.017

Cutler, D. M., Poterba, J. M., & Summers, L. H. (1988). What moves stock prices? (No. w2538). National Bureau of Economic Research. doi:10.3386/w2538

Dai, Z., Dong, X., Kang, J., & Hong, L. (2020). Forecasting stock market returns: New technical indicators and two-step economic constraint method. *The North American Journal of Economics and Finance*, *53*, 101216. doi:10.1016/j.najef.2020.101216

Demir, S., Mincev, K., Kok, K., & Paterakis, N. G. (2020). Introducing technical indicators to electricity price forecasting: A feature engineering study for linear, ensemble, and deep machine learning models. *Applied Sciences (Basel, Switzerland)*, *10*(1), 255. doi:10.3390/app10010255

Gan, L., Wang, H., & Yang, Z. (2020). Machine learning solutions to challenges in finance: An application to the pricing of financial products. *Technological Forecasting and Social Change*, *153*, 119928. doi:10.1016/j.techfore.2020.119928

Hassanpour, S., Tomita, N., De Lise, T., Crosier, B., & Marsch, L. A. (2019). Identifying substance use risk based on deep neural networks and Instagram social media data. *Neuropsychopharmacology*, *44*(3), 487–494. doi:10.1038/s41386-018-0247-x PMID:30356094

Heaton, J. B., Polson, N. G., & Witte, J. H. (2017). Deep learning for finance: Deep portfolios. *Applied Stochastic Models in Business and Industry*, *33*(1), 3–12. doi:10.1002/asmb.2209

Huang, J. Z., & Huang, Z. J. (2020). Testing moving average trading strategies on ETFs. *Journal of Empirical Finance*, *57*, 16–32. doi:10.1016/j.jempfin.2019.10.002

Hubáček, O., Šourek, G., & Železný, F. (2019). Learning to predict soccer results from relational data with gradient boosted trees. *Machine Learning*, *108*(1), 29–47. doi:10.1007/s10994-018-5704-6

Jiang, M., Liu, J., Zhang, L., & Liu, C. (2020). An improved Stacking framework for stock index prediction by leveraging tree-based ensemble models and deep learning algorithms. *Physica A*, *541*, 122272. doi:10.1016/j.physa.2019.122272

Kamarajan, C., Ardekani, B. A., Pandey, A. K., Kinreich, S., Pandey, G., Chorlian, D. B., Meyers, J. L., Zhang, J., Bermudez, E., Stimus, A. T., & Porjesz, B. (2020). Random Forest Classification of Alcohol Use Disorder Using fMRI Functional Connectivity, Neuropsychological Functioning, and Impulsivity Measures. *Brain Sciences*, *10*(2), 115. doi:10.3390/brainsci10020115 PMID:32093319

Lee, M. C., Liao, J. S., Yeh, S. C., & Chang, J. W. (2020, January). Forecasting the Short-term Price Trend of Taiwan Stocks with Deep Neural Network. In *Proceedings of the 2020 11th International Conference on E-Education, E-Business, E-Management, and E-Learning* (pp. 296-299). Academic Press.

Longato, E., Acciaroli, G., Facchinetti, A., Maran, A., & Sparacino, G. (2020). Simple linear support vector machine classifier can distinguish impaired glucose tolerance versus type 2 diabetes using a reduced set of CGM-based glycemic variability indices. *Journal of Diabetes Science and Technology*, *14*(2), 297–302. doi:10.1177/1932296819838856 PMID:30931604

Loureiro, A. L., Miguéis, V. L., & da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, *114*, 81–93. doi:10.1016/j.dss.2018.08.010

Manickavasagam, J., Visalakshmi, S., & Apergis, N. (2020). A novel hybrid approach to forecast crude oil futures using intraday data. *Technological Forecasting and Social Change*, *158*, 120126. doi:10.1016/j.techfore.2020.120126 PMID:32518424

Nabi, R. M., & Soran Ab, M. S., & Harron, H. (2020). A Novel Approach for Stock Price Prediction Using Gradient Boosting Machine with Feature Engineering (GBM-wFE). *Kurdistan Journal of Applied Research*, *5*(1), 28–48. doi:10.24017/science.2020.1.3

Parray, I. R., Khurana, S. S., Kumar, M., & Altalbe, A. A. (2020). Time series data analysis of stock price movement using machine learning techniques. *Soft Computing*, 1–9.

Piñeiro-Chousa, J., López-Cabarcos, M. Á., & Ribeiro-Soriano, D. (2020). Does investor attention influence water companies' stock returns? *Technological Forecasting and Social Change*, *158*, 120115. doi:10.1016/j.techfore.2020.120115

Pradhan, B., & Sameen, M. I. (2020). Modeling traffic accident severity using neural networks and support vector machines. In *Laser Scanning Systems in Highway and Safety Assessment* (pp. 111–117). Springer. doi:10.1007/978-3-030-10374-3_9

Pramudya, R. (2020). Technical Analysis to Determine Buying and Selling Signal in Stock Trade. *International Journal of Finance & Banking Studies, 9*(1), 58-67.

Qamar, U., Niza, R., Bashir, S., & Khan, F. H. (2016). A majority vote-based classifier ensemble for web service classification. *Business & Information Systems Engineering*, *58*(4), 249–259. doi:10.1007/s12599-015-0407-z

Takeuchi, L., & Lee, Y. Y. (2013). *Applying deep learning to enhance momentum trading strategies in stocks*. Working paper, Stanford University.

Tan, X., Li, S., Wang, C., & Wang, S. (2020, March). Enhancing High Frequency Technical Indicators Forecasting Using Shrinking Deep Neural Networks. In *2020 6th International Conference on Information Management (ICIM)* (pp. 249-256). IEEE. doi:10.1109/ICIM49319.2020.244707

Tratkowski, G. (2020). Construction of Investment Strategies for WIG20, DAX and Stoxx600 with Random Forest Algorithm. In *Contemporary Trends and Challenges in Finance* (pp. 179–188). Springer. doi:10.1007/978-3-030-43078-8_15

Vengertsev, D. (2014). Deep learning architecture for univariate time series forecasting. Technical report, Stanford University

Wong, J. Z. (2020). *On the Effectiveness of Technical Analysis, Time Series Analysis, and Machine Learning Approaches in Forecasting Stock Markets Movement Direction* (Doctoral dissertation). Tunku Abdul Rahman University College.

Yin, Y. (2020). Model-free tests for series correlation in multivariate linear regression. *Journal of Statistical Planning and Inference*, *206*, 179–195. doi:10.1016/j.jspi.2019.09.011

Zhou, Z. H., Wu, J., & Tang, W. (2002). Ensembling neural networks: Many could be better than all. *Artificial Intelligence*, *137*(1-2), 239–263. doi:10.1016/S0004-3702(02)00190-X

Justin Zhang teaches courses of Information Systems and Business Analytics in the Department of Management at University of North Florida. He received his Ph.D. in Business Administration with a concentration on Management Science and Information Systems from Pennsylvania State University, University Park. His research interests include economics of information systems, knowledge management, electronic business, business process management, information security, and social networking. He is the editor-in-chief of the Journal of Global Information Management, an ABET program evaluator, and an IEEE senior member.

Praveen Ranjan Srivastava is an Associate Professor at Indian Institute of Management Rohtak, India in the area of Management Information Systems. His research interests include Analytics, E-commerce, online learning, MCDM, software management, Machine Learning and Nature Computing. He has published more than 75 research Papers in various leading international journals and conferences. He is an awardee of Microsoft Research Grant. He is an Editor in Chief of International Journal of Software Engineering, Technology and Applications (IJSETA), published by Inderscience. He is also a member of Editorial Board of various leading journals and conferences.

Prajwal Eachempati is a research scholar at Indian Institute of Management Rohtak, India in the area of Management Information Systems. He has completed his Bachelor in Computer Science from Osmania University, Hyderabad, India. His research interests include MCDM, Machine Learning, Social Networking, Financial Data Management and Advanced Analytics.