IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received June 3, 2021, accepted June 6, 2021, date of publication June 14, 2021, date of current version June 22, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3088999

Novel Stock Crisis Prediction Technique–A Study on Indian Stock Market

NAGARAJ NAIK, (Graduate Student Member, IEEE), AND BIJU R. MOHAN
Department of IT, National Institute of Technology Karnataka, Mangalore 575025, India

Corresponding author: Nagaraj Naik (nagaraj21.naik@gmail.com)

This work was supported by the Visvesvaraya Ph.D Scheme for Electronics and IT the Departments of MeitY, Government of India.

ABSTRACT A stock market crash is a drop in stock prices more than 10% across the major indices. Stock crisis prediction is a difficult task due to more volatility in the stock market. Stock price sell-offs are due to various reasons such as company earnings, geopolitical tension, financial crisis, and pandemic situations. Crisis prediction is a challenging task for researchers and investors. We proposed a stock crisis prediction model based on the Hybrid Feature Selection (HFS) technique. First, we proposed the HFS algorithm to removes the irrelevant financial parameters features of stock. The second is the Naive Bayes method is considered to classify the strong fundamental stock. The third is we have used the Relative Strength Index (RSI) method to find a bubble in stock price. The fourth is we have used moving average statistics to identify the crisis point in stock prices. The fifth is stock crisis prediction based on Extreme Gradient Boosting (XGBoost) and Deep Neural Network (DNN) regression method. The performance of the model is evaluated based on Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error(RMSE). HFS based XGBoost method was performed better than HFS based DNN method for predicting the stock crisis. The experiments considered the Indian datasets to carry out the task. In the future, the researchers can explore other technical indicators to predict the crisis point. There is more scope to improve and fine-tune the XGBoost method with a different optimizer.

INDEX TERMS Recursive feature elimination, Boruta feature selection, stock price bubble, XGBoost, DNN, stock crisis.

I. INTRODUCTION

Due to the volatility in stock prices, predicting stock crisis movements is challenging. The stock price crisis is nothing but a significant drop in stock price more than 10% within a few days due to heavy selling in stocks [43]. The reasons for the selloff in the stock market are listed below.

1) The stock is overpriced.
2) Company posts the bad earnings.
3) Global market slowdown due to trade war.
4) Geopolitical Tension.
5) Pandemic situations like COVID19.

Early prediction of stock prices makes trading more profitable [3]. The Efficient Market Hypothesis (EMH) states that traded assets' prices, such as stocks, already reflect all publicly available information [35]. There has been an increasing number of studies [7], [34], [42], [47] that provide evidence contrary to what is suggested by the EMH. These studies show that the stock market can be predicted to some extent. Stock crisis prediction helps investors to exit the market at the right time.

The stock market crisis can emerge due to variations in global and local market economic policy and macroeconomic data. For example, a financial market slowdown in 2008 [16] is initially originated in the United States (US), and subsequently, it has affected the economy of other countries. It has been observed that the crisis may originate from a large size economy, and the impact of the crisis will affect smaller economies as well. For example, the subprime crisis originated in the US and evolved into a sovereign debt crisis in European countries. This crisis affected the Asian market, as well. Crisis prediction has been used in the banking sector, business, investments, and other areas. Crisis prediction is critical for the financial market, and this attracted many researchers and academicians. Predicting a crisis is one of the critical issues in the proposed work.

However, Chatzis *et al.* [12] proposed stock crisis events based on deep learning classification methods. The study

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

considered the less than one percentile of stock returns as a stock crisis point. This study has not been considered a fundamental analysis to identify stock quality because stock technical analysis performs better on the strong fundamental stock. Stock forms a bubble before it crashes. The study stock crisis is predicted without considering the bubble in stock prices. Hence identification of stock bubbles is one of the objectives in this paper.

Stock price crisis identification is challenging because the crisis can cause a financial market meltdown, geopolitical tension, and COVID 19 medical emergency. Hence it is not very easy to predict future stock prices. Therefore it creates an opportunity to study stock crisis-related issues.

The contribution of this paper is as follows.

1) Proposed Hybrid Feature Selection(HFS) algorithm to remove irrelevant financial ratio features. This is the first approach for stock crisis prediction based on the HFS method.
2) Naive Bayes classification method is considered to find the strong fundamental stock.
3) Stock price bubble identification using the RSI method. Usually, RSI is used for overbought and oversold movement identification. This is the first approach for stock price bubble identification using the overbought method.
4) Identification of stock crisis events using moving average statistics.
5) Forecasting of the future stock price crisis using the XGBoost and DNN regression method. This is the first approach using the regression method for stock crisis prediction.

The structure of this paper is as follows. Section 2 describes related work on stock crisis prediction. Section 3 presents the HFS based stock crisis prediction model. Section 4 presents the experimental results and discussion. Section 5, we conclude with a relevant discussion about the proposed method.

II. RELATED WORK

Tsai [49] discussed to find the relationship between stock price and exchange rate. The study found that the variables are negatively correlated. Therefore the least square estimation is not able to find the accurate correlation between exchange rate and stock. To overcome this, the authors proposed a quantile regression model. Andersen *et al.* [4] proposed a framework for intraday trading and stock price forecast based on volatility and return distribution. The study highlighted that forecasting based on asset return volatility and its distribution is complicated to fit in the Auto-Regressive Conditional Heteroscedasticity (ARCH) model due to negative correlations.

Kim [27] proposed the Support Vector Machine (SVM) model to forecast stock prices. KOSPI(Korean composite stock price index) daily stock price data were considered in the experiment. The study considered two variables based on the stock price movements, i.e., one indicates up, and 0 indicates down. Two thousand nine hundred twenty-eight trading samples have been considered, in that 20% of data is used for holdout and 80 % data for training. The original data scaled by performing normalization [-1.0,1.0]. To accomplish the work, 12 technical indicators raw data are given to the SVM model. The experimental work investigates SVM parameters upper bound C and kernel parameter sigma square.

An Artificial Neural Network(ANN) is a popular method for classification and pattern identification in stock prices. ANN has been adopted in most of today's applications to design smart and intelligent machines for business and science purposes. The ANN is learned from training data and identifies the future pattern based on experience. The ANN model is flexible because it can handle nonlinear data without prior knowledge of the relationship between input and output data. Deep neural network are considered for stock price prediction [22], [54].

The Autoregressive Integrated Moving Average(ARIMA) model is widely used to find a linear relationship in the time-series application. However, most researchers found that the ARIMA model cannot identify the nonlinear pattern in data. Therefore, most of the methods considered SVM and ANN. Pai *et al.* [40] proposed a hybrid ARIMA and SVM model for stock price prediction. The residuals are obtained using ARIMA and given input to SVM for forecasting. Blair *et al.* [6] considered two variables daily stock returns and volatility index(VIX). The ARCH model is applied to estimate the volatility between daily returns and VIX. The study concludes that VIX outperformance for forecasting the volatility compared to daily returns.

Tsaih *et al.* [50] proposed a hybrid artificial intelligence method for stock price forecasting. This approach combines a rule-based method and a neural network to predict the daily stock prices. The results are compared with backpropagation(BP) and perceptron. The S&P 500 stock is considered in the experiment. The advantage of reasoning neural networks is the fast learning rate compared to BP, and the number of hidden nodes is less compared to BP.

Wavelet Denoising-based Back Propagation network (WDBP) has been proposed to forecast the stock prices. MAE, RMSE metric is considered to evaluate the performance of the WDBP model. The shanghai stock exchange data has been considered for the experiment from 1993 to 2009. The data are split into two parts, training and testing. 80% of data is used for training and 20 % for testing. In this approach, data are decomposed into many layers using wavelet transformation. The representation of generated signals can be either low frequency or high frequency. BP neural network is considered to predict the future value based on wavelet transformation frequency [51].

Most of the work considered the Support Vector Regression(SVR) for predicting stock price [21], [29], [55]. One of the biggest problems of SVR is parameters estimation of the kernel function. However, in existing work, parameter estimation is carried out manually, i.e., trial and error. These manual estimations are not accurate. To overcome this,

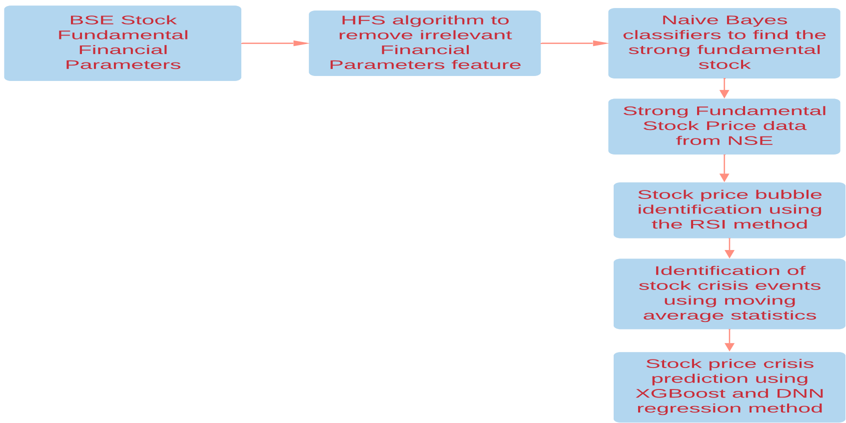


FIGURE 1. Overall proposed work.

TABLE 1. Related works.

Author	Techniques	Outcome
David Enke et.al [15]	Neural Network	Stock Price Prediction
Chenn Huanget et.al [23]	ANN, SVM	Stock Price Classification
Mehmet Orhan et al. [39]	GARCH	Stock Index Prediction
Nikolay Y. Nikolaev et al. [37]	GARCH	Currency volatility Prediction
Jan Wosnitzag et al. [52]	LPPL	Liquidity crisis (Physics-based study)
Qun Zhang et al. [58]	Quantile Regressions of Log- Periodic Power Law	Financial Crises(Physics-based study)
Chong Li [30]	LPPL	Chinese market bubble identification(Physics based study)
Sotirios P Chatzis et al. [12]	DNN	Stock Crises Classification
Werner Kristjanpoller et al. [28]	ANN-GARCH	Currency Price Prediction
Maji et.al[33]	Linear, Quadratic and Cubic curve model	Portfolio Management Framework
Chandar [11]	Elman Neural Network	Stock Price Prediction
Chandar [10]	TDNN,RBFNN and BPNN	Stock Price Prediction

the authors proposed multiple kernel learning methods to optimizes the SVR parameters [53].

Maji *et al.* [33] studied portfolio management framework using linear, quadratic and cubic curve model. In this work, different industry stocks are classified into clusters. The performance of the curve model is evaluated using the R-squared metric. Online textual news [5] were considered to predict the stock prices. Naïve Bays classifier was used to classify the news sentiments.

Chandar [10] considered Delay Neural Networks (TDNN), Radial Basis Function Neural Networks (RBFNN), and Back Propagation Neural Network (BPNN) for predicting the stock prices. The study concluded that the BPNN model was performed better than other models. Safari and Ghavifekr [45] investigated the stock price prediction using the ANN method. The RMSE metrics were considered to evaluate the performance of the model.

Chandar [11] considered Elman Neural Network(ENN) model for stock price prediction, and the Grey Wolf optimizer method was considered to optimize the parameters of the ENN model. Moreover, the DNN model is performed better than ANN and SVM techniques [32]. DNN has more than three layers of neural network that helps the model to learn more accurately compared to the ANN.

The overall related work is described in Table 1. Most of the work considered stock price prediction [18], [25], [31], [56], [46]. There is limited work on stock crisis prediction. Sotirios *et al.* [12] is proposed stock crisis prediction based on classification method. Log Periodic Power Law(LPPL) method is considered for stock crisis identification [30], [52]. Therefore it creates an opportunity to study the stock crisis-based prediction.

III. METHODOLOGY

Zhang *et al.* [57] proposed the LPPL method to identify the bubble in stock prices. The bubble is nothing but the exponential growth of stock prices. This study has not been considered a fundamental analysis to recognize the quality of stock. Benjamn Grahman [19] earned money in the stock market, but he has lost money during the stock crisis in 1929. Later he has published a book [20] on security analysis based on stock fundamentals. According to Grahman, the stock price's

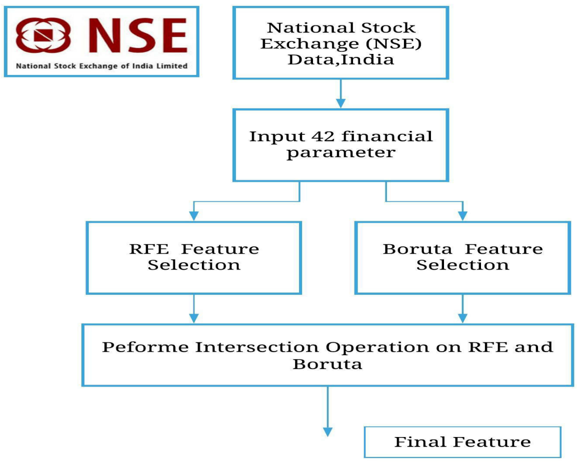


FIGURE 2. HFS feature selection.

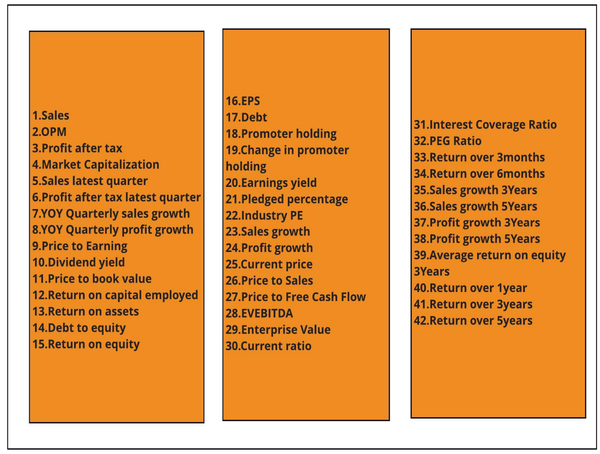


FIGURE 3. Stock financial parameters list.

fair value is based on company earnings, dividend, and asset value. Therefore we have considered financial parameters to recognize the quality of stock.

The overall flow of the proposed work is depicted in Fig 1. We proposed a Hybrid Feature Selection(HFS) algorithm for forecasting the future stock price crisis using the XGBoost and DNN regression method.

A. HYBRID FEATURE SELECTION (HFS) ALGORITHM TO REMOVE IRRELEVANT FINANCIAL PARAMETERS

The fair value of the stock price depends on stock financial parameters. There are many financial parameters such as price to earnings, company returns, company dept, etc. Identification of relevant stock parameters is a challenging task. Therefore we proposed Hybrid Feature Selection(HFS) technique to select an essential financial parameters feature. HFS technique combines two individual algorithms, namely Recursive Feature Elimination(RFE) and Boruta Feature Selection(BFS). Also, we performed an intersection operation to the outcome of this combination. The proposed work is depicted in Fig 2. The step-by-step proposed HFS technique is described in Algorithm 1.

To carried out the set of experiments, we have considered NIFTY50 stocks. Fundamentals of NIFTY50 stock financial parameters are retrieved from Bombay Stock Exchange(BSE), India [9]. Stock financial parameters list are depicted in Fig 3. We have considered 42 various stock financial parameters of NIFTY50 stock. The next task is identifying relevant stock financial parameters using the Recursive Feature Elimination (RFE) method. The 42 stock financial parameters are given as the input to the RFE to selects the best feature. The way RFE works is, it is a backward-compatible way of making feature selection. It starts with initially all the features, builds the model using a random forest regression method. Here price to earning (P/E) financial parameters is considered as target variable for regression. The next step is to remove the feature based on the Root mean square error (RMSE) score and build the model again. The RMSE score greater than nine is considered irrelevant, and each feature RMSE score is depicted in Fig 4. Based on the RFE method, 35 features are obtained as relevant, and it depicted in Fig 5.

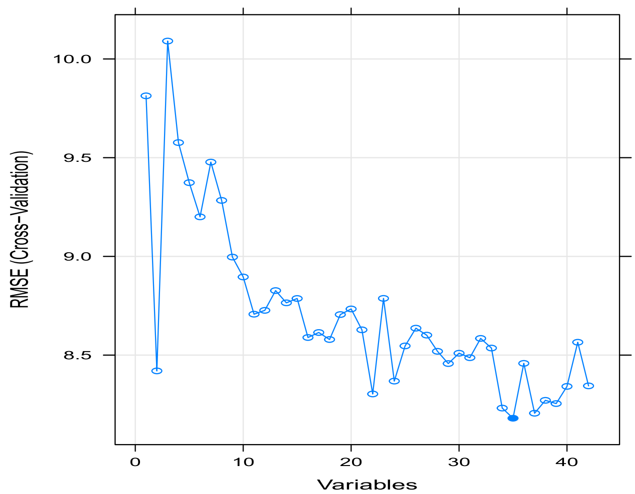


FIGURE 4. Feature RMSE score based on RFE.

Boruta feature selection(BFS) method is used to remove the irrelevant feature. The 42 stock financial parameters are given as the input to the BFS to selects the best feature. The way boruta works is that it creates the shadow financial parameters, duplicates the dataset, and shuffles each column's values. Here Price to Earning(P/E) financial parameters is considered as target variable for regression. Next step is train the model using random forest regression to find important financial parameter features. Important feature-based BFS method is depicted in Fig 6. Based on the BFS method, 17 features are obtained as relevant, and the final feature list is depicted in Fig 7.

Algorithm 1 Hybrid Feature Selection Algorithm

Input 42 stock financial parameter.
Select the best feature by performing a feature selection algorithm, namely Information gain(IG), RFE, and Boruta.
RFE Algorithm as follows:
For each financial parameter feature T_i, $i = 1..T$
Calculate model performance and financial parameter using random forest.
Keep important financial parameter feature T_i and remove weak financial parameters.
end RFE.
Boruta algorithm as follows:
Duplicates the dataset and shuffles the values in each column. These values are called shadow features.
Creates a shadow or duplicates feature of the financial parameter.
Train the random forest classifier to find important financial parameter features.
Each iteration, compare the original feature with the shadow feature Z score.
Remove the feature with the least Z score.
end Boruta.
Perform intersection operation RFE $\cap$ Boruta.
Output best feature.

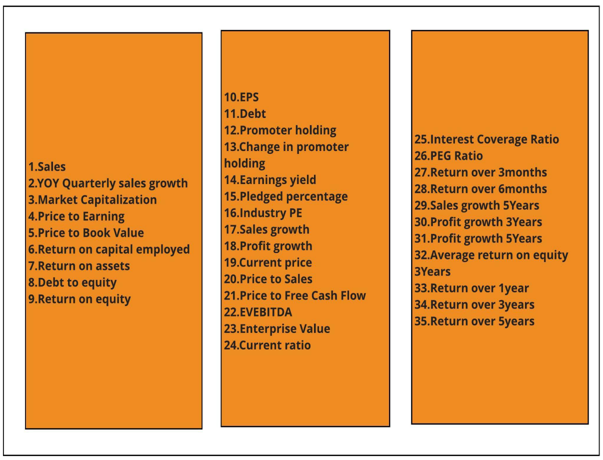


FIGURE 5. Final Selected Feature based on RFE.

We have performed an intersection operation on the outcome of the RFE and BFS feature selection method. Finally, we have obtained 13 features, depicted in Fig 8.

The final selected features of the HFS method are given as input to the Naive Bayes to classify the quality stock.

B. NAIVE BAYES CLASSIFIERS METHOD TO FIND THE STRONG FUNDAMENTAL STOCK

Naïve Bayes is widely used in text classification and sentiment analysis to identify positive and negative sentiments

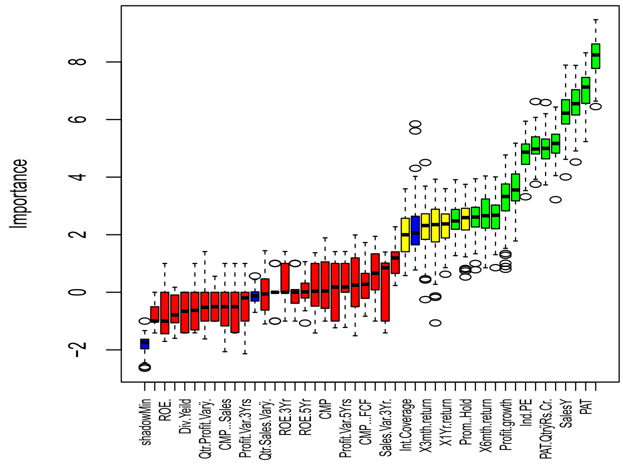


FIGURE 6. Important feature based on BFS.

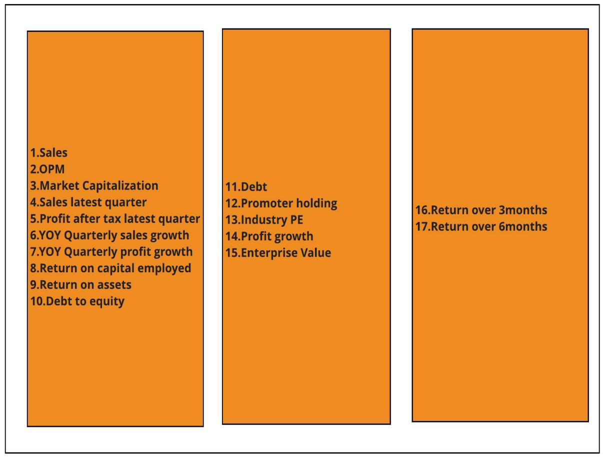


FIGURE 7. Final Selected Feature based on BFS.

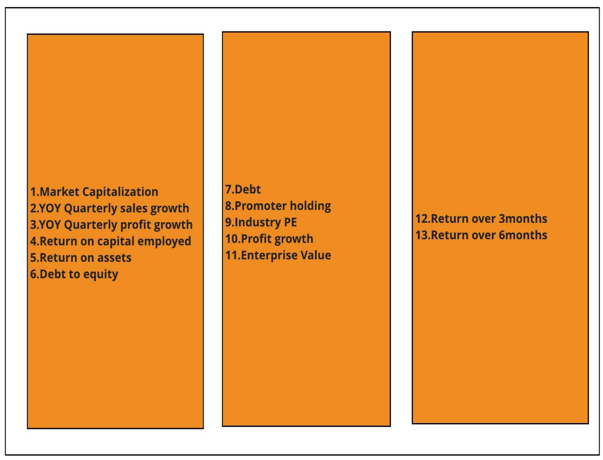


FIGURE 8. Final Selected Feature based on HFS.

[2], [36], [48]. In this paper, the NB classification method is considered to find the best strong fundamental stock based on stock financial parameters. Here price to earning(P/E) financial parameters is considered as target variable, i.e., probability of p(quality of stock). The Next is to calculate the

TABLE 2. Stock price bubbles are captured using RSI statistics.

Stock Name	RSI Value	Bubble Point Date	Bubble Price
Kotak Bank	76	16/Dec/2019	1698
	81	04/Oct/2010	255
	83	10/Dec/2007	348
ICICI Bank	81	23/Dec/2019	543
	79	01/Dec/2014	332
	81	08/Nov/2010	231
	79	29/Oct/2007	243
Axis Bank	75	03/June/2019	815
	72	06/Sep/2016	634
	81	02/Mar/2015	645
	82	04/Feb/2013	298
	80	04/Oct/2010	315
	86	14/Jan/2008	240
Yes Bank	78	31/Jul/2017	369
	81	04/Feb/2013	105
	75	10/Dec/2007	53
Bandhan Bank	83	06/Aug/2018	766
Indusland Bank	77	25/Jun/2018	2006
	79	27/May/2013	522
	87	10/Dec/2007	133
Tcs	79	15/Jan/2007	338
Infosys	83	20/Jan/2014	472
	79	03/Jan/2011	431
	83	02/Jan/2007	297
Hcl Tech	81	17/Feb/2020	616
	71	09/Mar/2015	525
	79	27/Mar/2006	82
Wipro	80	25/Feb/2019	291
	73	02/Mar/2015	253
	71	19/Feb/2007	154
Tech Mahindra	78	17/Feb/2020	844
	79	07/Feb/2015	738
	95	22/Jan/2007	465
Mind Tree	81	16/Feb/2020	1044
	80	04/Jan/2010	185
	79	19/Mar/2007	228
Hero Moto	75	04/Sep/2017	4005
	76	01/Dec/2014	3226
Eicher	77	04/Sep/2017	33021
	76	08/Oct/2007	493
TVS	89	01/Jan/2018	775
	82	27/Jan/2015	315
	88	13/Sep/2010	79
	85	03/Apr/2006	89
Sun Pharma	75	03/Sep/2018	666
	81	06/Apr/2015	1170
	80	28/Apr/2008	149
Dr.Reddy	76	17/Aug/2015	4949
	83	08/May/2006	849
Cipla	71	17/Sep/2018	680
	71	25/Aug/2008	240
Torrent Pharma	74	21/Jan/2019	1893
	72	28/May/2007	132
AuroBindo	76	10/Sep/2018	822
	82	08/Nov/2010	129
	73	25/Jan/2007	83
Biocon	77	26/Feb/2018	323
	86	01/Nov/2010	72
	81	12/Nov/2007	55
Cadila	74	31/Jul/2017	549
	71	19/Oct/2015	447
	73	11/Jul/2011	196
	82	02/May/2006	106

individual probability of each stock financial parameter with the target variable. The probability of Financial Parameters (FP) and Quality Stock(QS) is defined in Equation 1. Here higher probability of stock is considered as fundamentally strong stock. From the NIFTY50 stock, the top high probability of fundamentally strong stock is considered in the experiments based on the Naive Bayes classifier.

$$P(FP \mid QS) = \frac{P(QS \mid FP)\,P(FP)}{P(QS)} \tag{1}$$

C. STOCK PRICE BUBBLE IDENTIFICATION USING THE RSI METHOD

Relative Strength Index (RSI) statistics are used to find the bubble in stock price. The RSI technical indicator value ranges from 0 to 100. The RSI values below 30 indicate that the stock price is oversold, and RSI values above 70 indicate the overbought levels. When the RSI indicator value reaches above 70, there is a high chance that stock price is falling. Due to overprice in stock. We have considered the first 22 fundamentally strong stocks for computing the RSI. To compute the RSI value, we required the historical stock price data collected from the National Stock Exchange(NSE) portal. We have considered historical stock data from 2007 to December 2020. Next is to calculate the RSI value based on stock price using the below equations.

$$RSI(\#Days) = 100 - (100/(1 + Avg(Gain)/Avg(Loss))) \tag{2}$$

Most of the existing work RSI computed based on 14 days [26], [41]. However, in our approach, we have considered 200 days in RSI to find the stock price bubble. The reason is 14 days is used for intraday trading and not for the long term. The overprice in stock is nothing but a stock price bubble. The bubbles are captured using RSI statistics, and it is described in Table 2. The next step is the identification of stock crisis points based on the bubble of the stock price.

D. IDENTIFICATION OF STOCK CRISIS EVENTS USING MOVING AVERAGE STATISTICS

After identification of bubble in stock price, the next step is stock price crisis identification. The identification of stock crisis points is carried out by using the moving average technique. We have considered two moving averages, 50 days and 200 days. The moving average is computed based on the stock price. The first moving average of 50 days indicates the stock price's short movements, and the second moving average of 200 days indicates long movements of the stock price. The short movements of stock price trades below its long price movements indicate the downtrend in stock price. Such data points are considered stock crisis starting points, and it is depicted in Fig 9. In Fig 9 shows that the blue line indicates 50 days moving average, and the pink line indicates 200 days moving average. When 50 days moving average trades below 200 days, it is considered a stock crisis starting point. The stock crisis point is captured using moving average statistics, and it is described in Table 3. Here we have identified the stock crisis point. The next step is a prediction of future stock crisis points using XGBoost and the DNN regression method.

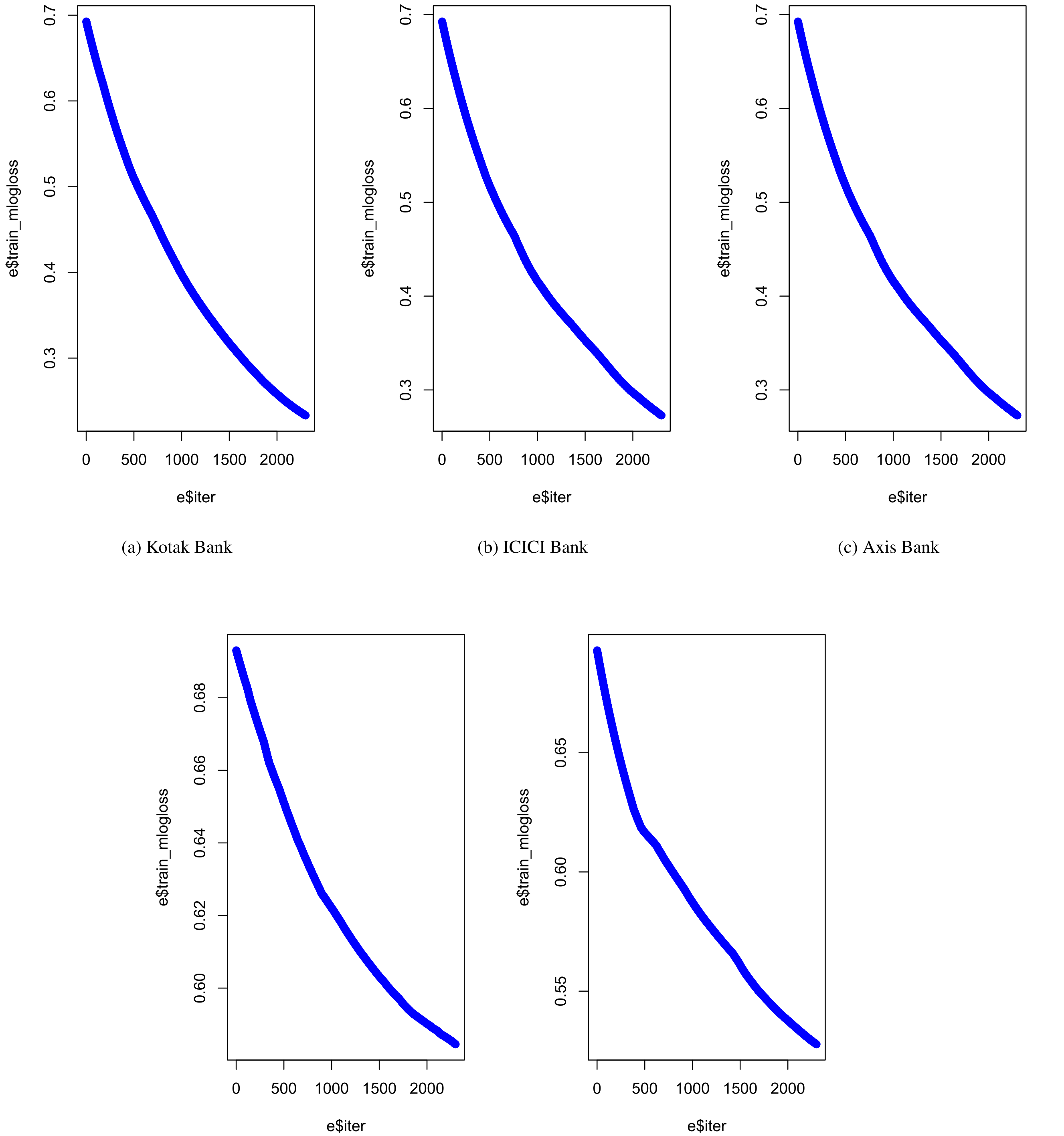
(a) Kotak-Bank

(b) ICICI-Bank

(c) Axis-Bank

(d) Yes-Bank

FIGURE 9. **Stock Crisis Point.**



(a) Kotak Bank (b) ICICI Bank (c) Axis Bank

(d) Infosys (e) Yes Bank

FIGURE 10. Loss function error rate.

E. STOCK CRISIS PREDICTION USING XGBoost REGRESSION TECHNIQUE

The linear regression model can be applied when the data is homoscedastic, i.e., when the error variance is constant [1]. However, stock price data is nonlinear, and the error term is not constant over time. XGBoost [13] and DNN [44] are more popular methods when the type of data is nonlinear. Therefore, we have considered XGBoost and DNN methods.

TABLE 3. Stock price crisis point is captured using moving average statistics.

Stock Name	Crash Point Date	Crash Price
Kotak Bank	02/Mar/2020	1618
	27/Dec/2010	226
	21/Jan/2008	265
ICICI Bank	02/Mar/2020	543
	09/Feb/2015	307
	10/Jan/2011	199
	04/Feb/2008	211.50
Axis Bank	15/Jul/2019	758
	26/Sep/2016	549
	20/Apr/2015	536
	18/Mar/2013	265
	22/Nov/2010	283
	03/Mar/2008	190
Yes Bank	23/Oct/2017	337
	17/Jun/2013	96
	21/Jan/2008	45
Bandhan Bank	24/Sep/2018	590
Indusland Bank	03/Sep/2018	1906
	22/Jul/2013	458
	28/Jan/2008	101
Tcs	28/May/2007	311
Infosys	18/Mar/2014	440
	14/Feb/2011	385
	05/Mar/2007	268
Hcl Tech	02/Mar/2020	568
	27/Apr/2015	452
	02/May/2006	73
Wipro	01/Jul/2019	279
	13/Apr/2015	229
	26/Feb/2007	133
Tech Mahindra	02/Mar/2020	774
	13/Mar/2015	671
	25/Jun/2007	365
Mind Tree	09/Mar/2020	855
	25/Jan/2010	155
	09/Jul/2007	197
Hero Moto	09/Oct/2017	3759
	19/Jan/2015	2929
Eicher	27/Nov/2017	30367
	17/Dec/2007	427
TVS	29/Jan/2018	697
	23/Mar/2015	265
	13/Dec/2010	72
	22/May/2006	68

XGBoost machine learning is one of the methods to solve complex data-driven real-world problems. The stock price crisis data are given as input to the XGBoost method. The closing price of the stock is considered as the target variable for regression. The residuals differences between the observed and predicted values are calculated. Next XGBoost fits a regression tree to the residuals. This fitting is called as XGBoost tree. Each tree starts a single leaf, and all residuals are placed into the leaf. The next step is to calculate the similarity score for residuals to split the tree. The similarity score and output value are defined in Equations 3 and 4. The goal is to find an output value for the leaf is to minimize the residuals. For that, we square the output value from the new tree and scale it with λ. If $\lambda > 0$, then we will shrink the output value. Because we are optimizing the output value from the first tree, we can replace the previous prediction. Here lambda is setting to zero. The loss function is described in Fig10. After 2000 iterations, there is a significant drop in the error.

FIGURE 11. Stock crisis input data.

$$Similarityscore = \frac{sumofresidual, square}{numberofresidual + \lambda} \tag{3}$$

$$Outputvalue = \frac{sumofresidual}{numberofresidual + \lambda} \tag{4}$$

F. STOCK CRISIS PREDICTION USING DNN REGRESSION TECHNIQUE

A Neural network is most popular to deal with nonlinear data [8], [14], [17], [24]. The single neural network structure is defined in Equation 5. It has one layer and one activation function. The single neural network takes inputs and calculates the weighted sum of the inputs and adds a bias. This calculation is represented in the form of a transfer function. This calculated weighted sum is passed an input to an activation function to generate the output. Here W is weight, B is bias, h is the hidden layer, and δ is the activation function.

$$h = \delta(Wh_1 + B) \tag{5}$$

In the proposed work, we have considered a deep neural network with 2 hidden layers, and it is defined as follows.

$$h_1 = \delta 1(Wh_1 + B1),$$
$$h_2 = \delta 2(Wh_2 + B2),$$

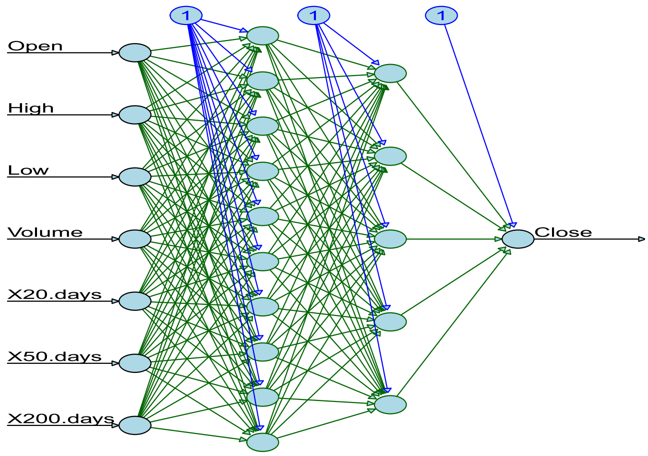


FIGURE 12. DNN method.

The stock price crisis data are given as input to the DNN method. The list of stock price input variables is depicted in Fig 11. The proposed DNN is depicted in Fig 12. Here closing price of a stock is considered as the target variable for the DNN regression method. We have normalized independent input variables by subtracting mean from each value, dividing them by standard deviation. The Rectified linear unit activation function is considered in the hidden layer.

IV. EXPERIMENT AND RESULT DISCUSSION

The set of experiments is carried out using the R Studio environment. Stock Crisis identification is a difficult task due to volatility in the stock market. There are many financial parameters, such as the price to earnings, company returns, company dept, etc. Identification of relevant stock financial parameters is a challenging task. Therefore we proposed the Hybrid Feature Selection technique to select an essential financial parameters feature. The XGBoost and DNN regression method was used to predict the stock crisis event. In this work, we considered a few NIFTY 50 stocks in the experiments from January 2007 to April 2021. To obtain the best results, we have fine-tuned the parameters of the XGBoost regression method. We have varied the size of the decision tree from 100 to 600, and the learning rate increase from 0.001 to 0.3. For the DNN method, we have varied the learning rate from 0.001 to 0.3. To validate the model performance, we have considered the ten cross-fold validation. It is the most popular statistical method to validate the results. In this method, datasets are divided into training sets and test sets, and a test set was used to evaluate the model's performance. In our experiments, we have divided datasets into ten folds. 90% of data is considered for training, and 10% is for testing. For each cross fold, we have validated the results, and at the end, we have considered the average results of 10 cross folds.

The model's performance is evaluated based on MSE, MAE, and RMSE score and it is defined in Equations 6, 7 and 8. Here y_i represents the observed value. x_i represents the predicted value and n is the total number of elements in datasets. The proposed HFS based XGBoost performs better than the DNN method, described in Table 4. Table 4 using HFS based XGBoost method shows that the lowest RMSE value for Kotak Bank, ICICI Bank, Axis Bank, and Infosys are 11.9457,4.43981, 4.237013, and 28.82983.

TABLE 4. Results.

Stock Name	Prediction Model	MSE	MAE	RMSE
Kotak Bank	HFS based XGBoost	142.6998	7.918674	11.9457
ICICI Bank	HFS based XGBoost	19.71191	2.486814	4.43981
Axis Bank	HFS based XGBoost	17.95228	2.440138	4.237013
Infosys	HFS based XGBoost	831.1591	21.00927	28.82983
Yes Bank	HFS based XGBoost	47.69003	5.190128	6.905797
Kotak Bank	HFS based DNN	38518.46	147.2966	196.2612
ICICI Bank	HFS based DNN	310.6973	13.5881	17.62661
Axis Bank	HFS based DNN	564.9369	14.07273	23.7684
Infosys	HFS based DNN	28.28089	4.512494	5.317978
Yes Bank	HFS based DNN	192.8784	7.667211	13.88807

$$mse = (\frac{1}{n})\sum_{i=1}^{n}\left|(y_i - x_i)^2\right| \tag{6}$$

$$mae = (\frac{1}{n})\sum_{i=1}^{n}|y_i - x_i| \tag{7}$$

$$rmse = \sqrt{(\frac{1}{n})\sum_{i=1}^{n}(y_i - x_i)^2} \tag{8}$$

From Fig 13, we observed that XGBoost prediction model data points are fitted better than the DNN model.

We have considered the Friedman test [38] to validate the post results of the DNN and XGBoost method, and it is defined in Equation 9. K is a number of the prediction model, N is the total number of elements, R_j is the sum of the ranks for the j prediction model.

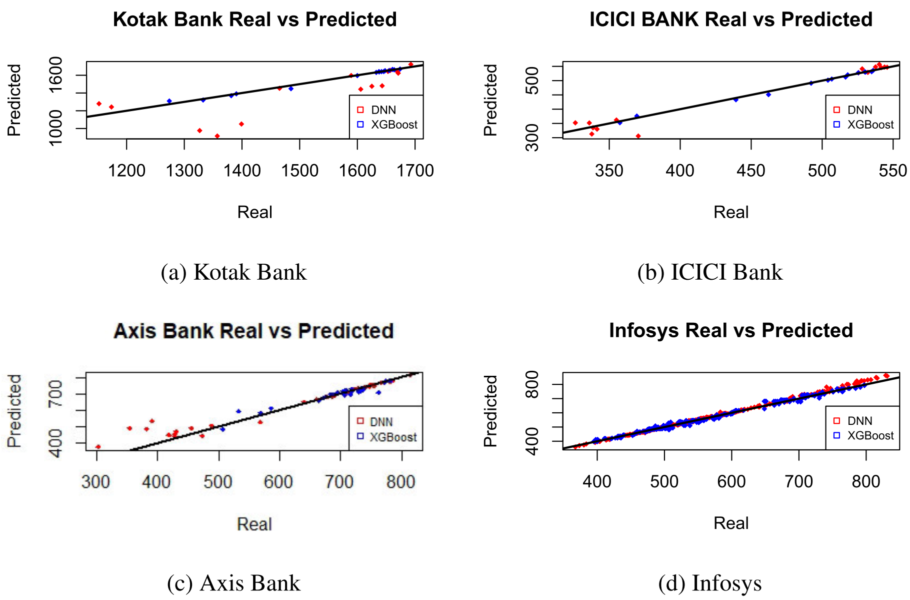
$$\frac{12}{NK(K+1)}\sum_{j=1}^{K}R_j^2 - 3N(K+1) \tag{9}$$

To check the results of DNN and XGBoost prediction model are significant or not, we have defined the null hypothesis, and alternate hypotheses are given below.

H_0: The result of the DNN and XGBoost prediction model are the same.

H_1:The results of the DNN and XGBoost prediction models are different.

For Kotak Bank stock, we validated the result using the Friedman test. We found chi-squared value is 28.6243,



(a) Kotak Bank (b) ICICI Bank

(c) Axis Bank (d) Infosys

FIGURE 13. **DNN and XGBoost prediction.**

df = 1 and p-value 0.0364. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the results of the DNN and XGBoost prediction model are the same for Kotak Bank stock. For Axis Bank stock, we found chi-squared value is 22.278, df = 1, and p-value 0.0431. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the DNN and XGBoost prediction model results are the same for Axis Bank stock. For Infosys stock, we found chi-squared value is 4.2118, df = 1, and p-value 0.04014. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the outcome DNN and XGBoost model are the same for Infosys stock. For ICICI Bank stock, we found chi-squared value is 0.0074349, df = 1, and p-value 0.01765. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the DNN and XGBoost prediction model results are the same for ICICI Bank stock. For YES Bank stock, we found chi-squared value is 0.0027855, df = 1 and p-value 0.0324. The p-value is less than 0.05, hence reject the alternative hypotheses. We concluded that the DNN and XGBoost prediction model results are the same for YES Bank stock. Based on the Friedman statistical test, we concluded that the DNN and XGBoost prediction model results are significant.

To our best knowledge, this is the first approach to HFS based stock crisis prediction model. Hence we have not compared the proposed model results with existing works.

The stock prices are affected due to many events such as company balance sheet variation, political uncertainty, bond market rate, and global market trends. Sometimes, stocks' prices react when there is a sudden change in management or share dividend and bonus announcement. Financial market stock price movements purely depend on various sources of information. It is not easy to interpret information from different sources. Aggregating and processing information from various platforms is a crucial challenge for future work.

V. CONCLUSION AND FUTURE WORK

Stock Crisis identification is a difficult task due to more fluctuations in the stock market. This is the first approach to address stock crisis prediction based on stock financial parameters and stock price to the best of our knowledge based on literature. We have proposed the Hybrid Feature Selection (HFS) algorithm to remove irrelevant stock financial parameters features. The NB classifier method is considered to find the fundamentally strong stock. Later, stock over price is identified by using the RSI method. Moving average statics are considered to identify the stock crisis points. The effectiveness of the model is quantified by using the XGBoost and DNN regression method. The performance of the model is evaluated based on MSE, MAE, and RMSE. HFS based XGBoost performs better than HFS based DNN method. Therefore in future work, different fundamentals stock and technical parameters can be applied to improve the model accuracy. We have explored a limited number of technical parameters of stock prices. In the future, the researchers can explore the other technical indicators to predict the crisis point. There is more scope to improve and fine-tune the XGBoost method with a different optimizer. Parameters optimization for XGBoost and DNN methods using evolutionary algorithm can be also be done as future work.

ACKNOWLEDGMENT

The task was carried out at the Department of IT, NITK Surathkal, India.

COMPETING INTERESTS

The authors declare that they have no competing interests

SIGNIFICANCE OF PROPOSED WORK

Stock Crisis identification is a difficult task due to volatility in stock market. There are many financial parameters such as price to earnings, company returns, company dept, etc. Identification of relevant stock financial parameters is a challenging task. Therefore we proposed Hybrid Feature Selection technique to select an essential financial parameters feature. HFS technique combines two individual algorithms, namely Recursive Feature Elimination and Boruta Feature Selection. The XGBoost and DNN regression method was used to predict the stock crisis event. HFS based XGBoost performs better than HFS based DNN method.

REFERENCES

[1] O. O. Aalen, "A linear regression model for the analysis of life times," *Statist. Med.*, vol. 8, no. 8, pp. 907–925, Aug. 1989.

[2] M. Aditya, A. Helen, and I. Suryana, "Naïve Bayes and maximum entropy comparison for translated novel's genre classification," in *Proc. J. Phys., Conf. Ser.*, vol. 1722. Bristol, U.K.: IOP Publishing, Jan. 2021, Art. no. 012007.

[3] S. S. Alexander, "Price movements in speculative markets: Trends or random walks," *Ind. Manage. Rev.)*, vol. 2, no. 2, p. 7, 1961.

[4] T. G. Andersen, T. Bollerslev, F. X. Diebold, and P. Labys, "Modeling and forecasting realized volatility," *Econometrica*, vol. 71, no. 2, pp. 579–625, Mar. 2003.

[5] V. Bag and U. V. Kulkarni, "Stock price trend prediction and recommendation using cognitive process," *Int. J. Rough Sets Data Anal.*, vol. 4, no. 2, pp. 36–48, Apr. 2017.

[6] B. J. Blair, S. H. Poon, and S. J. Taylor, "Forecasting S&P 100 volatility: The incremental information content of implied volatilities and high-frequency index returns," in *Handbook of Quantitative Finance and Risk Management*, C. F. Lee, A. C. Lee, and J. Lee, Eds. Boston, MA, USA: Springer, 2010, doi: 10.1007/978-0-387-77117-5_88.

[7] J. Bollen, H. Mao, and X. Zeng, "Twitter mood predicts the stock market," *J. Comput. Sci.*, vol. 2, no. 1, pp. 1–8, Mar. 2011.

[8] I. I. L. Bougoudis and A. Papaleonidas, "Fuzzy inference ANN ensembles for air pollutants modeling in a major urban area: The case of athens," in *Proc. Int. Conf. Eng. Appl. Neural Netw.* Cham, Switzerland: Springer, 2014, pp. 1–14.

[9] BSE. (2021). *Bombay Stock Exchange*. Accessed: Jul. 19, 2008. [Online]. Available: https://www.bseindia.com/

[10] S. K. Chandar, "Stock price prediction based on technical indicators with soft computing models," in *Proc. Int. Conf. Image Process. Capsule Netw.* Cham, Switzerland: Springer, 2020, pp. 685–699.

[11] S. K. Chandar, "Grey wolf optimization-elman neural network model for stock price prediction," *Soft Comput.*, vol. 25, no. 1, pp. 649–658, Jan. 2021.

[12] S. P. Chatzis, V. Siakoulis, A. Petropoulos, E. Stavroulakis, and N. Vlachogiannakis, "Forecasting stock market crisis events using deep and statistical machine learning techniques," *Expert Syst. Appl.*, vol. 112, pp. 353–371, Dec. 2018.

[13] T. Chen, T. He, M. Benesty, V. Khotilovich, Y. Tang, and H. Cho, "Xgboost: Extreme gradient boosting," *R Package Version* vol. 1, no. 4, pp. 1–4, Aug. 2015.

[14] K. Demertzis, L. S. Iliadis, and V.-D. Anezakis, "Extreme deep learning in biosecurity: The case of machine hearing for marine species identification," *J. Inf. Telecommun.*, vol. 2, no. 4, pp. 492–510, Oct. 2018.

[15] D. Enke and S. Thawornwong, "The use of data mining and neural networks for forecasting stock market returns," *Expert Syst. Appl.*, vol. 29, no. 4, pp. 927–940, Nov. 2005.

[16] S. Fleitas, P. Fishback, and K. Snowden, "Economic crisis and the demise of a popular contractual form: Building & loans in the 1930s," *J. Financial Intermediation*, vol. 36, pp. 28–44, Oct. 2018.

[17] T. J. Glezakos, T. A. Tsiligiridis, L. S. Iliadis, C. P. Yialouris, F. P. Maris, and K. P. Ferentinos, "Feature extraction for time-series data: An artificial neural network evolutionary training model for the management of mountainous watersheds," *Neurocomputing*, vol. 73, nos. 1–3, pp. 49–59, Dec. 2009.

[18] M. Göçken, M. Özçalıcı, A. Boru, and A. T. Dosdoğru, "Stock price prediction using hybrid soft computing models incorporating parameter tuning and input variable selection," *Neural Comput. Appl.*, vol. 31, no. 2, pp. 577–592, Feb. 2019.

[19] B. Graham, *The Intelligent Investor*. New Delhi, India: Prabhat Prakashan, 1965.

[20] B. Graham and C. McGolrick, *The Interpretation Financial Statements*. New York, NY, USA: HarperCollins Publishers, 1975.

[21] B. M. Henrique, V. A. Sobreiro, and H. Kimura, "Stock price prediction using support vector regression on daily and up to the minute prices," *J. Finance Data Sci.*, vol. 4, no. 3, pp. 183–201, Sep. 2018.

[22] E. H. Houssein, M. Dirar, K. Hussain, and W. M. Mohamed, "Assess deep learning models for Egyptian exchange prediction using nonlinear artificial neural networks," *Neural Comput. Appl.*, vol. 33, pp. 5965–5987, Sep. 2020.

[23] C. Huang, D. Yang, and Y. Chuang, "Application of wrapper approach and composite classifier to the stock trend prediction," *Expert Syst. Appl.*, vol. 34, no. 4, pp. 2870–2878, May 2008.

[24] L. Iliadis, S. Spartalis, A. Paschalidou, and P. Kassomenos, "Artificial neural network modelling of the surface ozone concentration," *Int. J. Comput. Appl. Math.*, vol. 2, no. 2, pp. 125–139, 2007.

[25] Z. Jin, Y. Yang, and Y. Liu, "Stock closing price prediction based on sentiment analysis and LSTM," *Neural Comput. Appl.*, vol. 32, pp. 9713–9729, Sep. 2019.

[26] Y. Kara, M. A. Boyacioglu, and Ö. K. Baykan, "Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul stock exchange," *Expert Syst. Appl.*, vol. 38, no. 5, pp. 5311–5319, May 2011.

[27] K.-J. Kim, "Financial time series forecasting using support vector machines," *Neurocomputing*, vol. 55, nos. 1–2, pp. 307–319, Sep. 2003.

[28] W. Kristjanpoller and M. C. Minutolo, "A hybrid volatility forecasting framework integrating GARCH, artificial neural network, technical analysis and principal components analysis," *Expert Syst. Appl.*, vol. 109, pp. 1–11, Nov. 2018.

[29] S. Lahmiri, "Minute-ahead stock price forecasting based on singular spectrum analysis and support vector regression," *Appl. Math. Comput.*, vol. 320, pp. 444–451, Mar. 2018.

[30] C. Li, "Log-periodic view on critical dates of the Chinese stock market bubbles," *Phys. A, Stat. Mech. Appl.*, vol. 465, pp. 305–311, Jan. 2017.

[31] X. Li, H. Xie, R. Wang, Y. Cai, J. Cao, F. Wang, H. Min, and X. Deng, "Empirical analysis: Stock market prediction via extreme learning machine," *Neural Comput. Appl.*, vol. 27, no. 1, pp. 67–78, Jan. 2016.

[32] T. S. Lim, K. G. Tay, A. Huong, and X. Y. Lim, "Breast cancer diagnosis system using hybrid support vector machine-artificial neural network," *Int. J. Electr. Comput. Eng. (IJECE)*, vol. 11, no. 4, p. 3059, Aug. 2021.

[33] G. Maji, D. Mondal, N. Dey, N. C. Debnath, and S. Sen, "Stock prediction and mutual fund portfolio management using curve fitting techniques," *J. Ambient Intell. Hum. Comput.*, pp. 1–14, 2021.

[34] B. G. Malkiel, "The efficient market hypothesis and its critics," *J. Econ. Perspect.*, vol. 17, no. 1, pp. 59–82, Feb. 2003.

[35] E. F. Fama, "Efficient capital markets: A review of theory and empirical work," *J. Finance*, vol. 25, no. 2, p. 383, May 1970.

[36] E. Najjar and S. Al-Augby, "Sentiment analysis combination in terrorist detection on Twitter: A brief survey of approaches and techniques," in *Research in Intelligent and Computing in Engineering*. Singapore: Springer, 2021, pp. 231–240.

[37] N. Y. Nikolaev, G. N. Boshnakov, and R. Zimmer, "Heavy-tailed mixture GARCH volatility modeling and Value-at-Risk estimation," *Expert Syst. Appl.*, vol. 40, no. 6, pp. 2233–2243, May 2013.

[38] T. W. O'Gorman, "A comparison of the F-test, Friedman's test, and several aligned rank tests for the analysis of randomized complete blocks," *J. Agricult., Biol., Environ. Statist.*, vol. 6, no. 3, pp. 367–378, Sep. 2001.

[39] M. Orhan and B. Köksal, "A comparison of GARCH models for VaR estimation," *Expert Syst. Appl.*, vol. 39, no. 3, pp. 3582–3592, Feb. 2012.

[40] P.-F. Pai and C.-S. Lin, "A hybrid ARIMA and support vector machines model in stock price forecasting," *Omega*, vol. 33, no. 6, pp. 497–505, Dec. 2005.

[41] J. Patel, S. Shah, P. Thakkar, and K. Kotecha, "Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques," *Expert Syst. Appl.*, vol. 42, no. 1, pp. 259–268, Jan. 2015.

[42] R. R. Prechter and W. D. Parker, "The Financial/Economic dichotomy in social behavioral dynamics: The socionomic perspective," *J. Behav. Finance*, vol. 8, no. 2, pp. 84–108, May 2007.

[43] R. Roll, "The international crash of October 1987," *Financial Analysts J.*, vol. 44, no. 5, pp. 19–35, 1988.

[44] F. Ronaghi, M. Salimibeni, F. Naderkhani, and A. Mohammadi, "ND-SMPF: A noisy deep neural network fusion framework for stock price movement prediction," in *Proc. IEEE 23rd Int. Conf. Inf. Fusion (FUSION)*, Jul. 2020, pp. 1–7.

[45] A. Safari and A. A. Ghavifekr, "International stock index prediction using artificial neural network (ANN) and Python programming," in *Proc. 7th Int. Conf. Control, Instrum. Autom. (ICCIA)*, Feb. 2021, pp. 1–7.

[46] S. Singh, K. S. Parmar, and J. Kumar, "Soft computing model coupled with statistical models to estimate future of stock market," *Neural Comput. Appl.*, pp. 1–19, 2021.

[47] V. L. Smith, "Constructivist and ecological rationality in economics," *Amer. Econ. Rev.*, vol. 93, no. 3, pp. 465–508, 2003.

[48] K. Srikanth, N. Murthy, and P. P. Reddy, "Sentiment classification on online retailer reviews," in *Proc. ICCCE*. Cham, Switzerland: Springer, 2021, pp. 1557–1563.

[49] I.-C. Tsai, "The relationship between stock price index and exchange rate in Asian markets: A quantile regression approach," *J. Int. Financial Markets, Inst. Money*, vol. 22, no. 3, pp. 609–621, Jul. 2012.

[50] R. Tsaih, Y. Hsu, and C. C. Lai, "Forecasting S&P 500 stock index futures with a hybrid ai system," *Decis. Support Syst.*, vol. 23, no. 2, pp. 161–174, 1998.

[51] J.-Z. Wang, J.-J. Wang, Z.-G. Zhang, and S.-P. Guo, "Forecasting stock indices with back propagation neural network," *Expert Syst. Appl.*, vol. 38, no. 11, pp. 14346–14355, May 2011.

[52] J. H. Wosnitza and C. Denz, "Liquidity crisis detection: An application of log-periodic power law structures to default prediction," *Phys. A, Stat. Mech. Appl.*, vol. 392, no. 17, pp. 3666–3681, Sep. 2013.

[53] C.-Y. Yeh, C.-W. Huang, and S.-J. Lee, "A multiple-kernel support vector regression approach for stock market price forecasting," *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2177–2186, Mar. 2011.

[54] P. Yu and X. Yan, "Stock price prediction based on deep neural networks," *Neural Comput. Appl.*, vol. 32, no. 6, pp. 1609–1628, Mar. 2020.

[55] J. Zhang, Y.-F. Teng, and W. Chen, "Support vector regression with modified firefly algorithm for stock price forecasting," *Int. J. Speech Technol.*, vol. 49, no. 5, pp. 1658–1674, May 2019.

[56] L. Zhang, F. Wang, B. Xu, W. Chi, Q. Wang, and T. Sun, "Prediction of stock prices based on LM-BP neural network and the estimation of overfitting point by RDCI," *Neural Comput. Appl.*, vol. 30, no. 5, pp. 1425–1444, Sep. 2018.

[57] Q. Zhang, D. Sornette, M. Balcilar, R. Gupta, Z. A. Ozdemir, and H. Yetkiner, "LPPLS bubble indicators over two centuries of the S&P 500 index," *Phys. A, Stat. Mech. Appl.*, vol. 458, pp. 126–139, Sep. 2016.

[58] Q. Zhang, Q. Zhang, and D. Sornette, "Early warning signals of financial crises with multi-scale quantile regressions of log-periodic power law singularities," *PLoS ONE*, vol. 11, no. 11, Nov. 2016, Art. no. e0165819.

NAGARAJ NAIK (Graduate Student Member, IEEE) received the B.E. degree in computer science engineering from VTU, India, in 2006, and the M.Tech. degree, in 2012. He is currently pursuing the Ph.D. degree with the Department of Information Technology, National Institute of Technology Karnataka, India. His research interests include financial data analysis, stock price analysis, and statistical techniques.

BIJU R. MOHAN received the Ph.D. degree in information technology from the National Institute of Technology Karnataka, India. He is currently working as an Assistant Professor and the Head of Information Technology Department, National Institute of Technology Karnataka. He has more than 15 years of experience in teaching and research. His research interests include time series analysis, statistical techniques, financial data analysis, software aging, and visualization.