Journal of Computer Science

Original Research Paper

Stock Price Classification Based on Hybrid Feature Selection Method

[1]Srivinay, [1]Manujakshi B. C., [2]Mohan Govindsa Kabadi,
[3]Nagaraj Naik and [1]Swetha Parvatha Reddy Chandrasekhara

[1]*School of CSE, Big Data Analytics Lab, Presidency University, Bengaluru, Karnataka, India*
[2]*Department of Computer Science and Engineering, GITAM School of Technology, Bengaluru, Karnataka, India*
[3]*Department of Computer Science and Engineering, Manipal Institute of Technology,*
Manipal Academy of Higher Education (MAHE), Manipal, 576104, Karnataka, India

Article history
Received: 30-09-2022
Revised: 14-12-2022
Accepted: 16-12-2022

Corresponding Author:
Srivinay
School of CSE, Big Data Analytics Lab, Presidency University, Bengaluru, Karnataka, India
Email: pressrivinay@gmail.com

And

Nagaraj Naik
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal Academy of Higher Education (MAHE), Manipal, 576104, Karnataka India
Email: nagaraj.naik@maipal.edu

Abstract: In recent years, investors and traders have used Technical Indicators (TIs) to forecast the stock market. An accurate classification model is required in the stock market to gain more profit. Selecting relevant TIs for the stock market remains a hot research topic. The proposed work aims to identify important technical indicators. Therefore, the proposed work considers a hybrid feature selection method to identify the relevant TIs. The hybrid feature selection combines two individual feature selection methods, such as Boruta and the Random Forest (RF) feature importance method. The work considers 20 TIs. The regression power of TIs is computed using the hybrid feature selection method. Using the hybrid feature selection method, the selected relevant TIs are given as input to the classification model, namely Naive Bayes (NB) and Deep Learning. The classification model aims to classify the stock price as up or down. The Hybrid Feature selection-based Deep Learning H_2O model performs better than the hybrid feature selection-based NB model in the experimental work. The accuracy of the hybrid feature selection of the Deep Learning H_2O model is around 86 to 89%. The work considers the National Stock Exchange (NSE) in India for the experimental work.

Keywords: Boruta Feature Selection, Deep Learning, Naive Bayes, Random Forest

Introduction

Stock price volatility is due to fluctuation in the stock market. The stock market is fluctuated because of Govt. policy and company earnings are reflected in stock price; hence the market is volatile. Therefore, it is challenging for investors to identify the up and down movements of stock prices. Most of the work considered Technical indicators statistics to identify the up and down movements of stock prices (Lee *et al*., 2022; Dhafer *et al*., 2022; George *et al*., 2022). Moreover, there are many technical indicators, but all are not relevant. Identifying appropriate technical indicators is a challenging task (Chandar, 2022). Most works considered a single feature selection-based approach to select the technical indicators. Combining two feature selection methods gives better results than a single feature selection (Gunduz, 2021). Therefore, this study considered hybrid feature selection for Technical indicators.

There are two methods performed on stock price analysis. First is the classification method and second is the regression method. Classification of stock price movements is one of the favorite research topics for investors to make accurate decisions (Zhang and Lou, 2021; Abdulhussein *et al*., 2022). The classification task involves classifying the stock up and down movements. This helps traders and investors predict the future movements of stocks up and down to make profits from investment. Few studies considered statistical methods (Imansyah and Mustafa, 2021), textual analysis (Amstad *et al*., 2021), data mining (Chang *et al*., 2021), and soft computing (Singh *et al*., 2021) to classify the stock prices. Financial institutions increasingly use Machine Learning (ML) techniques (Chen *et al*., 2021) for stock classification. The advantage of ML and data mining is that it can recognize stock trends from vast volumes of stock price data (Mohammed *et al*., 2022; Florencio *et al*., 2019). This study considered technical indicators for stock price classification. However, there are many technical indicators. The selection of appropriate TI for the stock remains a research topic (López Rodríguez and Zurita López, 2022). Most studies use single feature selection for stock price classification. Therefore, this study considered a hybrid feature selection-based machine learning method to classify the stock movements.

Fundamental and technical analyses are used to classify the stock prices (Puneeth *et al*., 2021). Fundamental analysis evaluates a stock's intrinsic value based on company earnings (Vora *et al*., 2023). Earning of stock depends on management quality, company earnings, and microeconomics data. Fundamental analysis uses a company balance sheet to evaluate the stock price and technical analysis where TI is considered to classify the stock prices. Technical analysis (Ayala *et al*., 2021) indicates that stock price moves are interconnected and repeat historical patterns. These technical analyses used a statistics-based method for trend identification. Accurate stock price direction predictions enable investors and traders to decide whether to sell or buy stock by reducing investment risk. Investors must understand the stock price pattern to beat the market and make money.

The investor's goal is to predict the market behavior during volatility in the market. Investors aim to purchase or sell stocks to maximize earnings. Market behavior is volatile because of the global economy, political issues, and domestic data (Xiao and Wang, 2021). According to the random walk theory, stock prices are unpredictable. AI and machine learning methods have recently forecast stock prices (Maiti, 2021).

Work focuses on selecting TI as input data to reduce forecasting dimensionality and ensure prediction performance (Gunduz, 2021). Successful stock market forecasting involves limited input features. Increasing the input feature contributes to dimensionality issues in the data. Traders use technical analysis to classify the future movement of stock price and decide to buy or sell stock (Peng *et al*., 2021). Using technical analysis, traders can identify the pattern of an undervalued stock.

Technical analysts use a simple moving average on stock price over a specific period to identify the relationship between the stock price (M'ng, 2018). For example, an 8-day moving average is the average of the last eight days' closing prices and determines the stock price behavior. The Relative Strength Index (RSI) is a technical indicator that uses recent price changes to determine if a stock is overbought or oversold (Maté, 2023). Moving Average Convergence Divergence (MACD) technical indicators displays the relationship between two price moving averages. MACD is determined by subtracting the 26-period and 12-period EMAs.

Many machine-learning approaches exist to analyze stock trends (Mashrur *et al*., 2020). Stock trends were initially predicted using classical regression. Since stock data is nonstationary and non-linear, machine learning techniques are applied. The SVM algorithm proposed a hyperplane in a higher dimension to distinguish classes (Li and Sun, 2020). A Support Vector Machine (SVM) is a learning method that uses kernel functions and has a limited solution. Devaser and Chawla (2022) proposed an anomaly detection approach for Nifty stocks using machine learning techniques. The machine learning method has been used in financial market forecasting (Kamalov *et al*., 2021).

Boruta Feature Selection (BFS) and random forest method are used to select the important technical indicator. The reason for selecting BFS is that it creates random duplicates/shadows of the input feature. The randomness approach improves the performance of the model. Therefore, the proposed work considered the hybrid BFS and random forest method to select the important technical indicators. Naive Bayes is used to identify positive and negative emotions in text classification and sentiment analysis (Amstad *et al*., 2021). Artificial Neural Networks (ANN) and Support Vector Machines (SVM) are machine learning algorithms that identify stock and index movement. All algorithm learns patterns differently. ANN and SVM are two standard ML techniques (Kurani *et al*., 2021; Chandrasekhara and Kabadi, 2021). These are useful for long-term and short-term stock trend analysis. Deep learning is recommended for financial time series prediction (Long *et al*., 2019). With increasing computing capabilities and the ability to manage enormous stock data, there is a need for machine learning techniques.

The contribution of this study is as follows:

1) The proposed work considered a hybrid feature selection to identify technical indicators. The hybrid feature selection combines two individual feature selections like Brouta and the random forest feature importance method
2) The proposed work considered the Naive Bayes (NB) and deep learning H_2O models to classify the up-and-down movement of the stock price

Alsubaie *et al*. (2019) investigated a limited number of relevant TI to forecast the stock price. Choosing over 30 TI decreased prediction accuracy, misclassification costs, and reduced investment return. This study result indicated that at least 10 TI produced the highest accuracy, the lowest cost, and the best investment performance. High-accuracy and low-cost classifiers could have created better investment results when simulating an actual trading strategy. The proposed classifier method improves the traditional Naive Bayes (NB) classifier's value probabilities to lower classification costs. NB probability values are adjusted such that it is more likely to provide the class with the lowest prediction cost, not the precise class. Predicting stock market trends can be done using cost-sensitive learning since traders must consider the cost of misclassification.

Vargas *et al*. (2018) discussed deep learning techniques that combine a convolutional layer with an RNN to forecast the daily stock price. A trading agent uses this model's output to trade. The literature reviews on stock price classification are shown in Table 1.

Table 1: Literature reviews on stock price classification

Author	Technical Indicator	Method	Gap
Lee *et al*. (2022)	Yes	DNN	Feature selection is not considered
Dhafer *et al*. (2022)	Yes	ANN	Feature selection is not considered
Alsubaie *et al*. (2019)	Yes	Naïve Bayes	Single feature selection-based approach
Vargas *et al*. (2018)	Yes	Deep learning	Feature selection is not considered
Yun *et al*. (2021)	Yes	GA-XG boost	Single feature selection-based approach
Jing *et al*. (2021)	Yes	Hybrid deep	Learning feature selection is not considered

Yun *et al.* (2021) proposed stock price forecasting using GA-XG Boost and a 3-stage feature engineering approach. The suggested prediction model delivers better performance than some benchmark studies. This study reveals that stock price prediction success is primarily due to the proposed feature engineering method. Adding 67 TI to primary historical stock data increases accuracy.

Jing *et al.* (2021) combine deep learning with sentiment analysis to predict Chinese share stock prices. Deep learning integrates LSTM Neural Networks for stock prediction with Convolutional Neural Networks for sentiment analysis. The hybrid model comprises data preprocessing, sentiment analysis, and model creation. The proposed work first vectorizes textual input for sentiment analysis and calculates technical indicators. Second, a CNN-based investor sentiment analysis model is trained and evaluated. Finally, emotion and technical indicators were considered to predict one-day-ahead closing prices using an LSTM neural network. SVM, RNN, and LSTM are trained as controls with the same training and test data in the experiment. The proposed model beats others in F-measure. The model is then created using 30 Shanghai Stock Exchange equities (SSE). The proposed model with a MAPE of 0.0449 is better than the single and other prediction models. Combining investor feelings and technical data with the LSTM neural network can construct a more accurate stock price prediction model.

Barroso *et al*. (2021) proposed fusing Technical Analysis (AT) with portfolio Optimization (OT), modifying the coupling order ATOT. The first scenario builds an optimal investment portfolio each month and trades using technical analysis. The experimental results improve portfolio performance.

Göçken *et al.* (2016) proposed new hybrid stock price forecasting algorithms to improve accuracy. GA-ANN chose 26 relevant input variables as the best subset, while hybrid-ANN chose 23. Both GA-ANN and hybrid-ANN forecasting models have one hidden layer.

Thakkar and Chaudhari (2021) examined deep learning based neural network stock market forecasting methods. The proposed work was categorized based on their prediction approach and diversity in derivatives-based market predictions. The work considered CNN, RNN, and other DNN techniques, their variants for stock market prediction, and their usefulness in the financial industry.

Orimoloye *et al*. (2020) compared a deep feedforward Neural Network (DNN) with shallow architectures of a one-layer neural network for predicting stock price indices in developed and emerging markets. A comprehensive study uses daily, hourly, minute, and tick data from 34 financial indexes from 32 nations over six years. Our evaluation results reveal a significant benefit from training deep architectures using a Rectifier Linear (RELU) activation function across all 34 markets when employing minute data. DNN's prediction performance was inferior to shallower designs when employing tick-level data. The results show that predicted accuracy peaks when training a DNN algorithm, regardless of size. When using DNN to predict stock price indices, RELU outperforms the hyperbolic tangent Activation Function across all markets and time horizons.

Alonso-Monsalve *et al.* (2020) used high-frequency technical analysis to examine convolutional neural networks as an alternative to multilayer perceptrons for trend classification of Bitcoin exchange rates. Convolutional neural networks, hybrid CNN-LSTM networks, multilayer perceptron, and radial basis function neural networks were compared to forecast whether Bitcoin, Dash, Ether, Litecoin, Monero, and Ripple will increase value vs. USD in the next minute. Based on 18 technical indicators generated from one-minute exchange rates over a year, the results suggest that all series were predictable. Convolutional LSTM neural networks outperformed the rest, but CNN neural networks did well in Bitcoin, Ether, and Litecoin. This study on the appropriateness of AI to anticipate trends in high-frequency cryptocurrency data using technical indicators gives additional evidence of the usefulness of convolution and deep learning in this field.

Zhao *et al*. (2021) introduced a neural network-based stock trend process. The work considered six models created using RNN and LSTM. The experimental results showed that RNN and LSTM prediction models were outperformed.

Kumar Chandar (2021) proposed a model for predicting stock market prices using historical data and natural-inspired algorithms. Elman Neural Network (ENN) was used because it can remember past knowledge and solve stock problems. ENN settings are determined through trial and error. This study used a grey wolf optimizer to optimize ENN settings. For model evaluation, NYSE and NASDAQ stock data are used. Experimental results and statistical metrics show that GWO-optimized ENN performs better. Results show that GWO-ENN outperforms other bio-inspired ANN models.

Predicting and classifying stock prices is difficult due to more ups and downs in the financial market. Therefore, there is a need for a more robust predictive model to classify and predict stock prices. Most of the current literature is based on machine learning techniques. It considers very few technical indicators to predict and classify stock price movements and most work has yet to consider the technical indicator selection. This study addresses two tasks (a) Technical indicators selection and (b) Stock price classification.

Proposed Work

Due to more technical indicators, identifying the relevant TI is one important task in stock price classification. The goal is to reduce the number of TI which helps to improve the stock classification task. Kara *et al*. (2011); Patel *et al*. (2015) was proposed stock price classification and it described in Fig. 1 and 2. However, they have not considered the feature selection approaches. Therefore, this study considered a hybrid feature selection-based stock price classification method. To our knowledge, this is the first approach to stock price classification based on boruta and the random forest feature importance method. The overall proposed work is shown in Fig. 3. The hybrid feature selection combines two individual feature selections like Brouta and the random forest feature importance method. This study considered the Naive Bayes (NB) and deep learning H_2O models to classify the up and down movement of the stock price.

Brief Description of Technical Indicators

Most investors use technical indicators to identify the stock price entry and exit level. In the majority of work, technical indicators selection is based on an expert's opinion or random selection (Kara *et al*., 2011). Below a few technical indicators are listed:

1. Simple Moving Average (SMA): SMA is a type of technical analysis used to find the average mean value of the stock price. The SMA analysis is used to know how the stock price changes over a specified number of days. The moving average is calculated based on the closing price of the stock. In SMA, if stock price trading is above the moving average technical indicator, then it indicates that the stock is in an uptrend. If the stock price is trading below the moving average of the technical indicator, it indicates that the stock is in a downtrend
2. Exponential Moving Average (EMA): EMA technical analysis gives more weight to the recent data and less weight to the older data. In SMA, weights are given equally to all the data. The EMA gives more importance to the most recent data points; this helps the trader to make quick trading decisions based on recent data
3. Momentum (MOM): It refers to the rate of acceleration of stock prices. It is the difference between the current closing price and the closing price of N (number of days) days ago. Based on the MOM trend, the investor will take a long or short position on a stock by looking at the acceleration in stock prices
4. Stochastic K % (STCK): George Lane developed a stochastic indicator. It analyzes the stock based on the closing price over a specified number of days. Most of the analysts use the default three days moving average. The range of stochastic indicators is between 0 to 100. The range values between 1 to 20 indicate oversold levels and 80 to 100 technical indicator ranges indicate overbought levels
5. Relative Strength Index (RSI): RSI is one kind of momentum indicator. 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

Hybrid Feature Selection

Boruta Feature Selection (BFS) and random forest method are used to select the important technical indicator. The reason for selecting BFS is that it creates random duplicates/shadows of the input feature. The randomness approach improves the performance of the model. Combining two feature selection methods gives better results than a single feature selection. Therefore, this study considered the Hybrid Feature Selection method. The hybrid feature selection combines two individual feature selections method, like Boruta and the Random Forest (RF) feature importance method.

Boruta Feature Selection

Initially, this study considered 20 technical indicators listed in Table 2. First, it randomizes the data by shuffling all features called shadow features. Then, it trains a random forest classifier on the data set and applies a feature importance metric to evaluate the value of each feature:

$$Boruta\ Z\ score = \frac{(Feature - \mu)}{\sigma} \tag{1}$$

At each iteration, it evaluates whether the feature is more important than the best of its shadow features. To evaluate the feature, we considered the Boruta *Z* score which is defined in Eq. 1. The lower *Z* score feature is removed from the feature list.

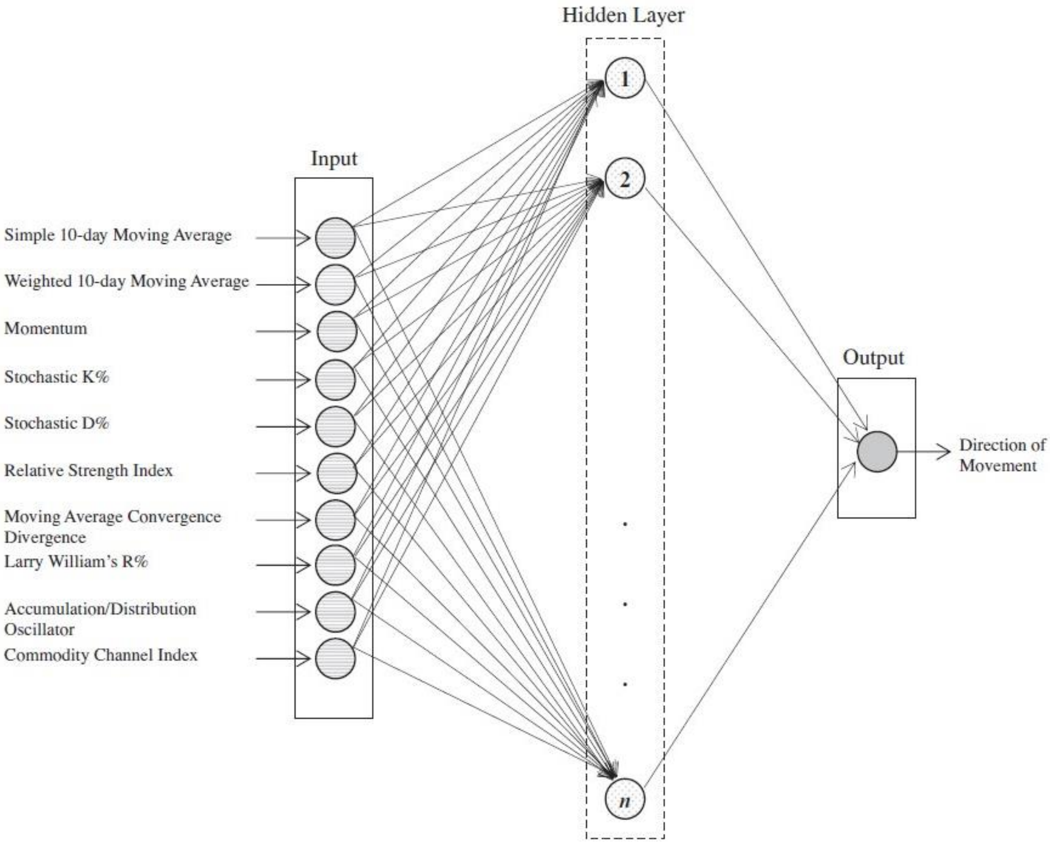

Fig. 1: Kara *et al.* (2011) ANN model for stock price classification

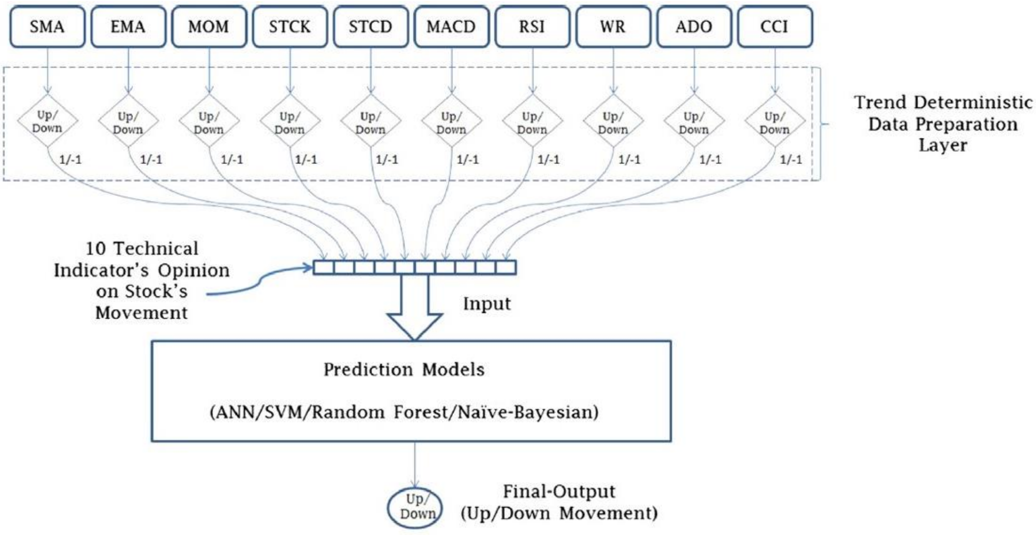

Fig. 2: Patel *et al*. (2015) stock price classification using trend deterministic data

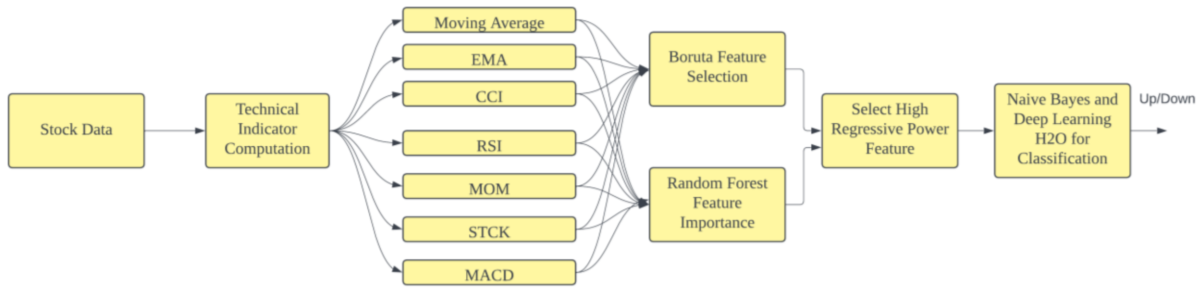

Fig. 3: Hybrid feature selection

Random Forest Feature Selection

Classification using decision trees is one of the popular techniques due to being accurate and efficient compared to other categorization methods. In this study, Random forests are considered to identify important technical indicators. Each sample fits a decision tree with a random subset of the original columns. Initially, this study considered 20 technical indicators listed in Table 2. RF selects the first v most essential attributes. In feature set $v = v_1, v_2, v_3,..., v_N$ (N is the total number of features) and it is defined in Eq. 2. The loss function is defined in Eq. 3, where, L indicates the loss function, Y indicates prediction and $f(v)$ is the function with different v attributes sets. $f(v)$ function aim is to identify the important technical indicators from base learner function $h_j(v)$ and it is defined in Eq. 4:

$$v = (v_1, v_2, v_3 vn) \tag{2}$$

$$L(Y, f(v)) \tag{3}$$

$$f(v) = argmax \sum_{i=1}^{i} (y = h_j(v)) \tag{4}$$

Finally, all of the predictions made by the trees are combined and the average mean value is considered for regression. This study feature with the highest regressive power is considered an important feature. The hybrid feature selection approach obtained the common feature from Boruta and random forest methods. Removed the features that do not exist in both feature lists. The hybrid feature selection algorithm is described in Algorithm 1. In experimental work, 70% of the data is used for training and 30% is used for testing. These selected features are given in the classification model.

Algorithm 1: Hybrid feature selection

1: Input TI.
2: Boruta feature selection steps:
3: Duplicate the TI or create a shadow of the original TI.
4: Random forest method to assess the importance of shadow or original feature.
5: Calculate $Feature - Score = \frac{(Feature - \mu)}{\sigma}$
6: Selected the confirmed feature and stored it in list S.
7: Random forest feature importance steps:
8: Input TI.
9: Create multiple tree T.
10: Use the information gained to split tree D.
11: Random forest method to assess the importance of each feature.
12: Selected the feature with high regressive power and store list K.
13: Output: Select the common feature from List S and K.

Naive Bayes

Naïve Bayes assumes class independence. Bayesian classifier predicts data class given test data. Bayes' theorem predicts probability. Bayes' theorem computes the posterior probability, defined in the equation below:

$$P(c \mid x) = \frac{P(x \mid c) P(c)}{P(x)} \tag{5}$$

$P(c|x)$ is the posterior probability of hypothesis c given occurrence x. In our scenario, hypothesis c is class Up/Down probability and event x is test data. $P(x|c)$ is the conditional probability of event x given c. A simple Bayesian classifier works as follows: Here are two classes up and down and data x occurrence. Bayesian classifiers classify the data using the highest probability. In experimental work, 70% of the data is used for training and 30% is used for testing.

Deep Learning for Classification

Artificial neural networks are dense networks of interconnected neurons activated by inputs. The proposed work considered a four-layer deep learning H_2O model for stock price classification. It consists of two hidden layers. The summation of the weighted input of technical indicators and bias is computed using Eq. 6. Twenty technical indicators are considered as inputs to the input layer. Due to the sigmoid function, the output layer produces a continuous 0-1 output. The up or down prediction threshold is set to 0.5 for proposed experiments. For output values larger than 0.5, the forecast is up; otherwise, down. At each epoch, weights are modified using gradient descent with momentum to reach a global minimum. In the experimental setup, we fine-tuned the parameter of predictive models, such as the number of hidden layer neurons, learning rate, momentum constant, and epochs:

$$\alpha = \left(\sum_{i=1}^{n} W_i \; t_i \; Bias \right) \tag{6}$$

Experimental Results

This study considered 12-year data from CNX Nifty, stock exchange India. Initially, it considered 20 technical indicators. The datasets sample are described in Table 5. It consists of open stock price, low price, high price, close price, and technical indicators. These technical indicators are used for estimating the future value of stock prices. Due to more technical indicators, identifying the relevant TI is critical in stock price classification. The goal is to reduce the number of TI which helps to improve the stock classification. In the proposed work hybrid feature selection combines two individual feature selections like Boruta and the Random Forest Feature importance method. Using Boruta and RF-based, selected features are shown in Tables 3 and 4. Here each feature score is computed. The feature score lies between 0 to 1. If the feature score is near 1, it indicates the relevant feature. This study considered the Naive Bayes (NB) and Deep Learning H_2O models to classify the up and down movement of the stock price. Accuracy metrics are used in the performance evaluation process, defined in Eq. 7. Accuracy measures the percentage of accurate classifications. We divide correct predictions by total predictions. The loss function of Apollotyre and DLF stock are described in Fig. 4 and 5. The green line indicates the training error and the blue line indicates the testing error. Initially, we considered 1000 epochs. After running each iteration, it reduces the error and converges the error rate. The accuracy of the hybrid feature selection of the deep learning model is around 86 to 89% and it is shown in Table 6. The accuracy before and after feature selection is shown in Table 6. Using the deep learning method, the following stock, namely APOLLO TYRE, CUB, DLF, RAIN IND, and SONATA SOFTWARE model classification results are 81.35, 84.50, 89, 80.50,74.50%:

$$Accuracy = \frac{TruePos + TrueNeg}{TruePos + TrueNeg + FalsePos + FalseNeg} \tag{7}$$

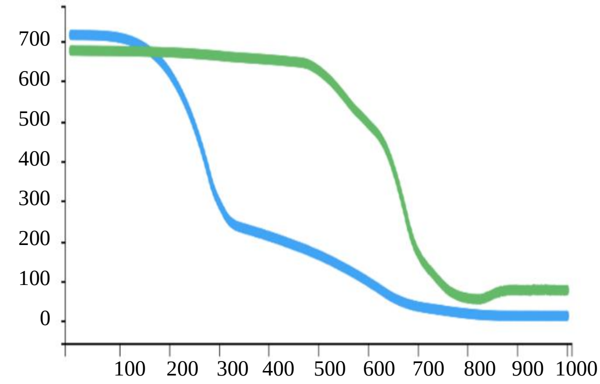

Fig. 4: Loss function of apollotyre stock

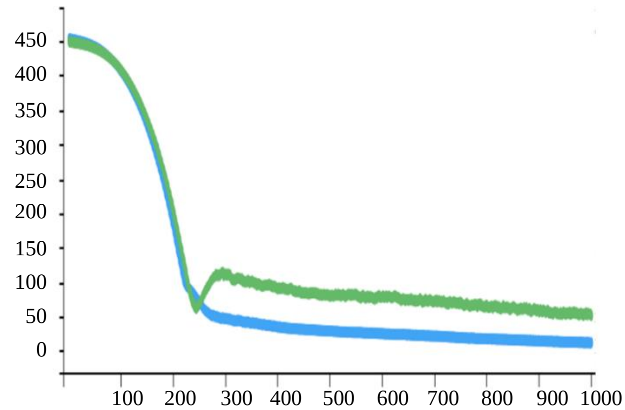

Fig. 5: Loss function of DLF stock

Table 2: Number of features

Technical indicator
Open
High
Low
Close
EMA
Smoothing line
Parabolic SAR
MA
Smoothing line
RSI
MACD histogram
MACD
Signal
% R
WMA
CCI
Smoothing line
MOM
% K
% D

Table 3: Selected feature using Boruta

Technical indicator	Norm hits	Decision
Open	0.989898990	Confirmed
High	0.595959596	Confirmed
Low	0.878787879	Confirmed
Close	0.959595960	Confirmed
EMA	0.787878788	Confirmed
Smoothing line	0.767676768	Confirmed
Parabolic SAR	0.222222222	Rejected
MA	0.888888889	Confirmed
Smoothing line	0.767676768	Confirmed
RSI	1.000000000	Confirmed
MACD histogram	0.040404040	Rejected
MACD	0.909090909	Confirmed
Signal	0.050505051	Rejected
% R	1.000000000	Confirmed
EMA	0.939393939	Confirmed
CCI	1.000000000	Confirmed
Smoothing line	0.979797980	Confirmed
MOM	0.868686869	Confirmed
% K	1.000000000	Confirmed
% D	0.989898990	Confirmed

Table 4: Selected feature using random forest

Technical indicator	RF importance score
Low	5.100000000
Close	1.100000000
High	6.500000000
Open	6.000000000
EMA	5.000000000
MA	4.000000000
Smoothing line	3.300000000
Smoothing line.1	2.900000000
Parabolic SAR	1.200000000
MACD	9.706880982
RSI	9.381821591
MOM	7.841674935
Signal	7.776633270
Smoothing Line.2	6.838753567
X.R	6.771598009
CCI	6.607176521
X.D	6.518160990
X.K	6.269664775
EMA.1	6.080791149
MACD histogram	5.617174525

Table 5: Sample data-sets

Open	High	Low	Close	EMA	Smoothing line	Parabolic SAR	MA	Smoothing line	RSI
175.74	186.90	175.74	181.74	178.51	178.65	172.10	178.33	179.04	58.88
178.74	188.60	178.74	185.44	179.90	178.66	172.40	178.62	178.78	62.90
182.09	189.80	182.09	188.29	181.57	179.18	173.04	179.50	178.84	65.68
185.19	189.50	183.00	186.45	182.55	180.05	174.05	180.23	179.06	62.43
185.82	185.82	179.00	182.41	182.52	181.01	174.99	180.69	179.47	55.89
184.12	184.12	174.00	177.01	181.42	181.59	189.80	180.84	179.98	48.56
180.56	180.56	168.20	172.40	179.62	181.54	189.48	180.52	180.36	43.33
176.48	185.80	172.05	177.84	179.26	181.07	188.63	180.80	180.62	50.14
177.16	182.95	175.75	179.75	179.36	180.44	187.82	181.26	180.82	52.32
178.45	182.85	177.15	180.42	179.57	179.85	187.03	181.11	180.91	53.09
179.44	183.65	178.40	181.49	179.95	179.55	186.28	180.67	180.87	54.35
180.46	180.46	177.00	178.76	179.72	179.57	185.55	179.62	180.69	50.60
179.61	181.00	177.00	179.09	179.59	179.64	184.86	178.80	180.29	51.03
179.35	181.60	179.10	180.31	179.73	179.71	184.19	178.56	179.75	52.72
179.83	182.30	178.20	180.19	179.83	179.76	183.55	178.92	179.31	52.52
180.01	183.45	175.10	180.00	179.86	179.75	168.20	179.76	179.13	52.20
180.00	185.95	180.00	183.44	180.58	179.92	168.51	180.38	179.28	57.31
181.72	186.00	181.72	184.69	181.40	180.28	169.20	180.93	179.71	59.02
183.20	187.65	182.30	185.02	182.12	180.76	170.21	181.44	180.29	59.50
184.11	186.50	181.95	184.22	182.54	181.30	171.61	181.75	180.85	57.79

Table 6: Experimental result

Stock	Method	Accuracy before feature selection	Accuracy using random feature selection	Accuracy using Boruta feature selection	Accuracy using hybrid feature selection
Apollotyre	Naive Bayes	67.78	71.35	69.56	76.67
Apollotyre	Deep Learning H_2O	70.35	74.67	75.98	81.35
CUB	Naive Bayes	72.50	74.65	76.78	78.45
CUB	Deep Learning H_2O	73.17	76.20	80.70	84.50
DLF	Naive Bayes	76.50	78.50	81.50	86.90

Table 6: Continue

DLF	Deep learning H_2O	72.50	71.50	88.00	89.00
RAIN IND	Naive bayes	64.50	60.70	73.70	77.50
RAIN IND	Deep learning H_2O	67.50	68.60	76.50	80.50
Sonata software	Naive Bayes	64.70	66.50	71.50	74.60
Sonata software	Deep learning H_2O	67.70	68.50	69.60	74.50

Conclusion

Due to more technical indicators, identifying the relevant TI is critical in stock price classification. This study addresses two tasks (a) Technical indicators selection and (b) Stock price classification. In the proposed work the hybrid feature selection method is used to identify the important technical indicators. Sixteen technical indicators in the experiment found the highest regressive power and were considered important technical indicators 20. These 16 technical indicators are given input to the classification model, namely Naive Bayes (NB) and deep learning. The classification model aims to classify the stock price up and down. The accuracy of the hybrid feature selection of the deep learning model is around 86 to 89%. In the proposed work Combining two feature selection methods gives better results than a single feature selection method. GARCH parameters can be optimized in future work using meta-heuristic methods and machine learning techniques.

Acknowledgment

Thank you to the publisher for their support in the publication of this research article. We are grateful for the resources and platform provided by the publisher, which have enabled us to share our findings with a wider audience. We appreciate the efforts of the editorial team in reviewing and editing our work, and we are thankful for the opportunity to contribute to the field of research through this publication.

Funding Information

The authors have not received any financial support or funding to report.

Author’s Contributions

Srivinay: Designed the research plan and methodology.

Manujakshi B. C. and Mohan Govindsa Kabadi: Conceptualization and Supervision.

Nagaraj Naik and Swetha Parvatha Reddy Chandrasekhara: Methodology and validation.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Abdulhussein, S. H., Al-Anber, N. J., & Atee, H. A. (2022). “Iraqi stock market prediction using proposed model of convolution neural network,” *Journal of Computer Science, vol. 18*, pp: 350-358.

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. https://doi.org/10.1016/j.eswa.2020.113250

Alsubaie, Y., El Hindi, K., & Alsalman, H. (2019). Cost-sensitive prediction of stock price direction: Selection of technical indicators. *IEEE Access*, *7*, 146876-146892. https://ieeexplore.ieee.org/abstract/document/8861031

Amstad, M., Gambacorta, L., He, C., & Xia, F. D. (2021). Trade sentiment and the stock market: New evidence based on big data textual analysis of Chinese media. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3783897

Ayala, J., García-Torres, M., Noguera, J. L. V., Gómez-Vela, F., & Divina, F. (2021). Technical analysis strategy optimization using a machine learning approach in stock market indices. *Knowledge Based Systems*, *225*, 107119. https://doi.org/10.1016/j.knosys.2021.107119

Barroso, B. C., Cardoso, R. T., & Melo, M. K. (2021). Performance analysis of the integration between Portfolio Optimization and Technical Analysis strategies in the Brazilian stock market. *Expert Systems with Applications*, *186*, 115687. https://doi.org/10.1016/j.eswa.2021.115687

Chandar, S. K. (2022). Convolutional neural network for stock trading using technical indicators. *Automated Software Engineering, 29*(1), 1-14. https://doi.org/10.1007/s10515-021-00303-z

Chandrasekhara, S. P. R., & Kabadi, M. G. (2021). Wearable IoT-based diagnosis of prostate cancer using GLCM-multiclass SVM and SIFT-multiclass SVM feature extraction strategies. *International Journal of Pervasive Computing and Communications.* https://doi.org/10.1108/IJPCC-07-2021-0167

Chang, T. H., Wang, N., & Chuang, W. B. (2021). Stock Price Prediction Based on Data Mining Combination Model. *Journal of Global Information Management (JGIM)*, *30*(7), 1-19. https://doi.org/10.4018/JGIM.296707

Chen, W., Zhang, H., Mehlawat, M. K., & Jia, L. (2021). Mean-variance portfolio optimization using machine learning-based stock price prediction. *Applied Soft Computing*, *100*, 106943. https://doi.org/10.1016/j.asoc.2020.106943

Devaser, V., & Chawla, P. (2022). "A novel anomaly detection approach for nifty stocks using machine learning for construction of efficient portfolio to reduce losses and protectnbsp; gains," *Journal of Computer Science*, *vol. 18*, pp. 441-452.

Dhafer, A. H., Mat Nor, F., Alkawsi, G., Al-Othmani, A. Z., Ridzwan Shah, N., Alshanbari, H. M., ... & Baashar, Y. (2022). Empirical Analysis for Stock Price Prediction Using NARX Model with Exogenous Technical Indicators. *Computational Intelligence and Neuroscience*, *2022*. https://doi.org/10.1155/2022/9208640

Florencio, F., Valenca, T., Moreno, E. D., & Colaço Jr, M. (2019). Performance analysis of deep learning libraries: TensorFlow and PyTorch. *Journal of Computer Science*, *15*(10.3844).

George, J., Nair, A. M., & Yathish, S. (2022). Analysis of Market Behavior Using Popular Digital Design Technical Indicators and Neural Network. In *Expert Clouds and Applications* (pp: 445-458). Springer, Singapore. https://doi.org/10.1007/978-981-16-2126-0_37

Göçken, M., Özçalıcı, M., Boru, A., & Dosdoğru, A. T. (2016). Integrating metaheuristics and artificial neural networks for improved stock price prediction. *Expert Systems with Applications*, *44*, 320-331. https://doi.org/10.1016/j.eswa.2015.09.029

Gunduz, H. (2021). An efficient stock market prediction model using a hybrid feature reduction method based on variational autoencoders and recursive feature elimination. *Financial Innovation*, *7*(1), 1-24. https://doi.org/10.1186/s40854-021-00243-3

Imansyah, S., & Mustafa, M. H. (2021). The analysis of financial ratios effect the stock price of consumer goods sector companies listed in kompas100 index. *Dinasti International Journal of Digital Business Management*, *2*(2), 371-384. https://doi.org/10.31933/dijdbm.v2i2.779

Jing, N., Wu, Z., & Wang, H. (2021). A hybrid model integrating deep learning with investor sentiment analysis for stock price prediction. *Expert Systems with Applications*, *178*, 115019. https://doi.org/10.1016/j.eswa.2021.115019

Kamalov, F., Gurrib, I., & Rajab, K. (2021). Financial forecasting with machine learning: Price vs return. Kamalov, F., Gurrib, I. & Rajab, K. (2021). *Financial Forecasting with Machine Learning: Price Vs Return. Journal of Computer Science*, *17*(3), 251-264. https://doi.org/10.3844/jcssp.2021.251.264

Kara, Y., Boyacioglu, M. A., & Baykan, Ö. K. (2011). Predicting direction of stock price index movement using artificial neural networks and support vector machines: The sample of the Istanbul Stock Exchange. *Expert Systems with Applications*, *38*(5), 5311-5319. https://doi.org/10.1016/j.eswa.2010.10.027

Kumar Chandar, S. (2021). Grey Wolf optimization-Elman neural network model for stock price prediction. *Soft Computing*, *25*(1), 649-658. https://doi.org/10.1007/s00500-020-05174-2

Kurani, A., Doshi, P., Vakharia, A., & Shah, M. (2021). A comprehensive comparative study of Artificial Neural Network (ANN) and Support Vector Machines (SVM) on stock forecasting. *Annals of Data Science*, 1-26. https://doi.org/10.1007/s40745-021-00344-x

Lee, M. C., Chang, J. W., Yeh, S. C., Chia, T. L., Liao, J. S., & Chen, X. M. (2022). Applying attention-based BiLSTM and technical indicators in the design and performance analysis of stock trading strategies. *Neural Computing and Applications*, 1-13. https://doi.org/10.1007/s00521-021-06828-4

Li, X., & Sun, Y. (2020). Stock intelligent investment strategy based on support vector machine parameter optimization algorithm. *Neural Computing and Applications*, *32*(6), 1765-1775. https://doi.org/10.1007/s00521-019-04566-2

Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, *164*, 163-173. https://doi.org/10.1016/j.knosys.2018.10.034

López Rodríguez, F. S., & Zurita López, J. M. (2022). Detection of buy and sell signals using technical indicators with a prediction model based on neural networks. In *International Conference of the Thailand Econometrics Society* (pp. 721-737). Springer, Cham. https://doi.org/10.1007/978-3-030-97273-8_48

Maiti, M. (2021). Random Walk Hypothesis. In *Applied Financial Econometrics* (pp. 47-65). Palgrave Macmillan, Singapore. https://doi.org/10.1007/978-981-16-4063-6_2

Mashrur, A., Luo, W., Zaidi, N. A., & Robles-Kelly, A. (2020). Machine learning for financial risk management: A survey. *IEEE Access*, 8, 203203-203223. https://doi.org/10.1109/ACCESS.2020.3036322

Maté, C. (2023). The Relative Strength Index (RSI) to Monitor GDP Variations. Comparing Regions or Countries from a New Perspective. *Trends in Mathematical, Information and Data Sciences*, 83-91. https://doi.org/10.1007/978-3-031-04137-2_9

M'ng, J. C. P. (2018). Dynamically Adjustable Moving Average (AMA') technical analysis indicator to forecast Asian Tigers' futures markets. *Physica A: Statistical Mechanics and its Applications*, *509*, 336-345. https://doi.org/10.1016/j.physa.2018.06.010

Mohammed, F. E., Zghal, N. S., Aissa, D. B., & El-Gayar, M. M. (2022). "Classify breast cancer patients using hybrid data-mining techniques," *Journal of Computer Science*, *18*, pp: 316-321. https://thescipub.com/abstract/jcssp.2022.316.321

Orimoloye, L. O., Sung, M. C., Ma, T., & Johnson, J. E. (2020). Comparing the effectiveness of deep feedforward neural networks and shallow architectures for predicting stock price indices. *Expert Systems with Applications*, *139*, 112828. https://doi.org/10.1016/j.eswa.2019.112828

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems with Applications*, *42*(1), 259-268. https://doi.org/10.1016/j.eswa.2014.07.040

Peng, Y., Albuquerque, P. H. M., Kimura, H., & Saavedra, C. A. P. B. (2021). Feature selection and deep neural networks for stock price direction forecasting using technical analysis indicators. *Machine Learning with Applications*, *5*, 100060. https://doi.org/10.1016/j.mlwa.2021.100060

Puneeth, K., Rudagi, S., Namratha, M., Patil, R., & Wadi, R. (2021, December). Comparative Study: Stock Prediction Using Fundamental and Technical Analysis. In *2021 IEEE International Conference on Mobile Networks and Wireless Communications (ICMNWC)* (pp. 1-4). IEEE. https://doi.org/10.1109/ICMNWC52512.2021.9688449

Singh, S., Parmar, K. S., & Kumar, J. (2021). Soft computing model coupled with statistical models to estimate future of stock market. *Neural Computing and Applications*, *33*(13), 7629-7647. https://doi.org/10.1007/s00521-020-05506-1

Thakkar, A., & Chaudhari, K. (2021). A comprehensive survey on deep neural networks for stock market: The need, challenges and future directions. *Expert Systems with Applications*, *177*, 114800. https://doi.org/10.1016/j.eswa.2021.114800

Vargas, M. R., Dos Anjos, C. E., Bichara, G. L., & Evsukoff, A. G. (2018, July). Deep leaming for stock market prediction using technical indicators and financial news articles. In *2018 international joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE. https://doi.org/10.1109/IJCNN.2018.8489208

Vora, V., Shah, M., Chouhan, A., & Tawde, P. (2023). Stock Market Prices and Returns Forecasting Using Deep Learning Based on Technical and Fundamental Analysis. In *Information and Communication Technology for Competitive Strategies (ICTCS 2021)* (pp. 717-728). Springer, Singapore. https://doi.org/10.1007/978-981-19-0098-3_68

Xiao, J., & Wang, Y. (2021). Investor attention and oil market volatility: Does economic policy uncertainty matter? *Energy Economics*, *97*, 105180. https://doi.org/10.1016/j.eneco.2021.105180

Yun, K. K., Yoon, S. W., & Won, D. (2021). Prediction of stock price direction using a hybrid GA-XGBoost algorithm with a three-stage feature engineering process. *Expert Systems with Applications*, *186*, 115716. https://doi.org/10.1016/j.eswa.2021.115716

Zhang, D., & Lou, S. (2021). The application research of neural network and BP algorithm in stock price pattern classification and prediction. *Future Generation Computer Systems*, *115*, 872-879. https://doi.org/10.1016/j.future.2020.10.009

Zhao, J., Zeng, D., Liang, S., Kang, H., & Liu, Q. (2021). Prediction model for stock price trend based on recurrent neural network. *Journal of Ambient Intelligence and Humanized Computing*, *12*(1), 745-753. https://doi.org/10.1007/s12652-020-02057-0

Abbreviation Meaning

TI:	Technical Indicators
ML:	Machine Learning
AI:	Artificial Intelligence
MACD:	Moving Average Convergence Divergence
SVM:	Support Vector Machine

DOI: 10.3844/jcssp.2023.274.285

RNN: Recurrent Neural Network
GA: Genetic Algorithm
LSTM: Long Short-Term Memory
CNN: Convolutional Neural Network
MAPE: Mean Absolute Percentage Error
NP: Non-deterministic Polynomial-time
GA: Genetic Algorithm
BPNN: Backpropagation feed forward Neural Network
ENN: Elman Neural Network model
BSE: Bombay Stock Exchange
NYSE: New York Stock Exchange
EMA: Expontial Moving Average
GARCH: Generalized Autwo Regressive Conditional Heteroskedasticity
RELU: Rectified Linear Activation Function
NASDAQ: National Association of Securities Dealers Automated Quotations
WMA: Weighted Moving Average
MA: Moving Average