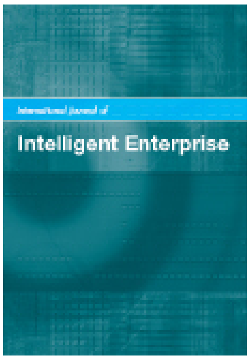


International Journal of Intelligent Enterprise

ISSN online: 1745-3240 - ISSN print: 1745-3232
https://www.inderscience.com/ijie

Feature selection for stock price prediction: a critical review

Binita Kumari, Srikanta Patnaik, Tripti Swarnkar

DOI: 10.1504/IJIE.2021.10043220

Article History:

Received:	01 May 2018
Last revised:	06 February 2020
Accepted:	17 February 2020
Published online:	30 November 2022

Feature selection for stock price prediction: a critical review

Binita Kumari* and Srikanta Patnaik

Department of Computer Science and Engineering,
Siksha 'O' Anusandhan (Deemed to be University),
Odisha, India
Email: binitakumari@soa.ac.in
Email: srikantapatnaik@soa.ac.in
*Corresponding author

Tripti Swarnkar

Department of Computer Application,
Siksha 'O' Anusandhan (Deemed to be University),
Odisha, India
Email: triptiswarnakar@soa.ac.in

Abstract: Stock price prediction has drawn huge attention due to its impact on economic stability. Accurate stock price prediction is highly essential to reduce the risk associated with it so as to decide good investment strategies. There are various factors influencing the prediction of stock indices namely gross margin, exchange rate, inflation rate, relative index and so on. Feature selection plays a vital role in effective and accurate prediction of stock indices. This paper aims to provide a clear review of widely used features affecting the stock price fluctuations, feature selection techniques and prediction models from the recent literature. The study also highlights the future directions in this domain focusing the enhancement of the prediction performance.

Keywords: features; feature selection; stock price prediction.

Reference to this paper should be made as follows: Kumari, B., Patnaik, S. and Swarnkar, T. (2023) 'Feature selection for stock price prediction: a critical review', *Int. J. Intelligent Enterprise*, Vol. 10, No. 1, pp.48–72.

Biographical notes: Binita Kumari is an Assistant Professor in the Department of Computer Science and Engineering, Faculty of Engineering and Technology, SOA University, Bhubaneswar, India. She received her BTech from the Utkal University, MTech from the SOA (Deemed to be University), India, in 2001 and 2011, respectively. She is currently a Doctorate student at the Chair for Computer Science and Engineering at the SOA (Deemed to be University), India. Her current research interests include deep learning, data mining and machine learning.

Srikanta Patnaik is a Professor in the Department of Computer Science and Engineering, Faculty of Engineering and Technology, SOA University, Bhubaneswar, India. He has published around 100 research papers in international journals and conference proceedings. He is author of two text books and 32 edited volumes and few invited book chapters, published by leading international publisher like Springer-Verlag, Kluwer Academic, etc. He is the Editor-in-Chief of *International Journal of Information and Communication Technology* and *International Journal of Computational Vision and Robotics* published from Inderscience Publishing House, England and also the Editor-in-Chief of book series on *Modeling and Optimization in Science and Technology* published from Springer, Germany.

Tripti Swarnkar is a Professor and a HOD in the Department of Computer Application, Faculty of Engineering and Technology, SOA (Deemed to be Universit)y, Bhubaneswar, India. She received her PhD in Engineering from the IIT Kharagpur, India in 2015. She is guiding PhD's in the field of Machine Learning, Deep Learning, Biomedical Engineering and Bioinformatics. She has more than 25 Master theses in the area of Machine Learning, Biomedical Applications and Bioinformatics. She has published many research papers in international journals and conference proceedings. She is an author of two books and few invited book chapters, published by leading international publishers.

1 Introduction

Stock market is where purchasing and selling of stocks (shares) happens among buyers and sellers. Many people around the globe invest in stock market. The stock market has always been highly unpredictable due to the number of reasons like political situation, global economy and so on, along with the traditional features. Successful investments are very difficult in the absence of good ability to predict the stock price. In literature, some basic factors like technical indicators, macroeconomic indicators, financial ratios, market sentiments have been used and proved as the main factors affecting the variations in stock price (Hagenau et al., 2013). The factors or an input variable varies as per the prediction models. Thus, the viewpoint of selecting principal components for stock prediction is distinct in associated works as there is no uniformity of the most key variables in stock prediction. The nature of prediction model may change when different sets of input features are used. Thus, designing the superlative prediction model for stocks is a challenging task for the researchers and investors.

This paper has been formulated as follows: Section 2 provides background details for feature selection in stock price prediction and enlists different types of features used as input variables, feature selection techniques and the prediction techniques being widely used for stock prediction. Section 3 provides a general discussion and critical review and Section 4 concludes the paper.

2 Background details

Many conventional techniques has been used to foresee the stock moving price or closing price (Tsai and Hsiao, 2010; Oztekin et al., 2016; Zhang et al., 2014; Dash et al., 2014, 2016). Technical analysis and fundamental analysis are the two traditional approaches for prediction of stock market. Stock price prediction done using technical analysis is established on historical information where charts are the prime tool (Su and Cheng, 2016; Lee, 2009). From the conventional information, relevant information is mined for pattern recognition. The study of the elements affecting supply and demand is known as the fundamental analysis. Stock price prediction done using fundamental analysis is based on information gathering and interpretation, and thus makes use of the gap between an event occurrence and response of the market to the event (Tsai and Hsiao, 2010). Economic data of companies, auditor's reports, balance sheets, and income statements are used for fundamental analysis. Since news echoes the current supply and demand chain of the market, it also plays a vital role in fundamental analysis. The statistics of the published papers on feature selection for stock prediction during 2005 to 2017 is given in Figure 1.

Figure 1 Numbers of published articles considered on feature selection for stock prediction (2005–2017) (see online version for colours)

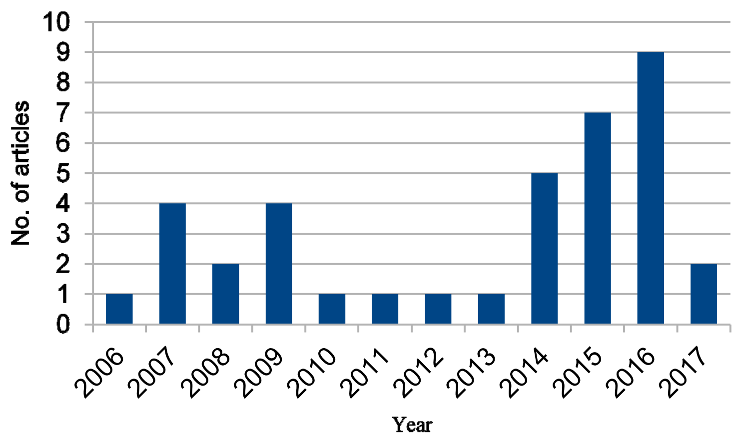


Figure 2 System architecture for stock prediction (see online version for colours)

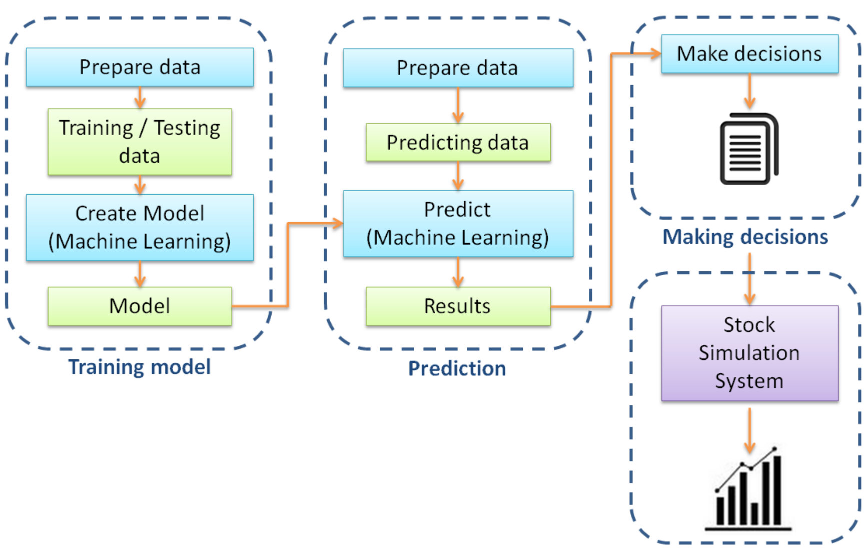


Figure 2 gives an overview of the architecture of stock prediction process. The raw stock data is taken as the initial input to the model and pre-processing steps are performed to prepare the data for training the model. The pre-processed data is prorated into training and test data to optimise the learning model. The trained model is finally used for interpreting the stock related decisions.

2.1 Feature selection

In stock prediction it is important to choose features with some prediction information. Reduction of irrelevant or redundant features results in better prediction accuracy and decreases the computational complexity. Applying feature selection can lead to simplifying data visualisation and data understanding. It reduces the measurement and storage necessities, decreases training and usage times disregarding the curse of dimensionality to enhance the prediction conduct (Schumaker and Chen, 2006; Tan et al., 2005; Huang et al., 2005). Many methods like Bayesian factor graph (Wang et al., 2015), principal component analysis (PCA) (Tsai and Hsiao, 2010; Altinbas and Biskin, 2015; Wing et al., 2014), genetic algorithm (GA) (Tsai and Hsiao, 2010; Anish and Majhi, 2016), decision trees (DT) (Tsai and Hsiao, 2010), Chi-square (Hagenau et al., 2013), sequential forward search (Ni et al., 2011), mutual information (Barak et al., 2015), factor analysis (Altinbas and Biskin, 2015) and so on has been used in literature for feature selection. GA, PCA, and DT are three well known feature selection techniques due to their ability to find near optimal feasible solution in a reduced computational time in comparison to other feature selection techniques. Next, we discuss in brief the different types of features used in literature for stock prediction.

2.2 Features

Stock prices are determined in the marketplace where selling and buying occurs. There exists no definite equation that tells about the behaviour of the stock. Literature states that there are different features that move the stock price up and down (Tsai and Hsiao, 2010). These features which affect the behaviour of the stock market, are divided into four categories: *fundamental features*, *technical features*, *macroeconomic features* and *market sentiment*. Each of these features are further discussed in brief.

2.2.1 Fundamental features

The fundamental comprises of the qualitative and quantitative information that adds to the economic well-being and the consequent financial estimation of a company, security or currency. These fundamentals are analysed by analysts and investors to establish an estimate to whether the underlying asset is treated as justifiable investment or not. For businesses, some of the fundamentals are information such as revenue, earnings, assets, liabilities and growth. In a stock market, stock prices can be determined primarily by fundamental features which is an amalgamation of an earning base for example, earnings per share (EPS) and a valuation multiple for example a P/E ratio.

The stock price has correlation with the intrinsic values like sales, profitability, EPS, management, corporate image, product quality, employee layoffs, accounting errors or scandals and so on. Table 1 lists out some of the fundamental factors (Tsai and Hsiao, 2010; Zhang et al., 2014) known from the literature. The fundamental factors may vary from region to region. Till date there exists no in general agreement on fundamental variables.

Table 1 List of fundamental, technical features and macroeconomic features

Sl. no.	*Fundamental features*	*Technical features*	*Macroeconomic features*
1	Gross margin growth	20-day bias	Exchange rate
2	Net income	Rate of change	Inflation rate
3	Operating income	Stochastic indicator	Oil price
4	Cash flow ratio	Relative index	US lagging indicator
5	Current ratio	10-day moving average	Foreign exchange reserves
6	Quick ratio	Moving average convergence/divergence (MACD)	Interest rate
7	Sales growth ratio	Commodity channel index (CCI)	Monitoring indicator
8	Liabilities ratio	Buying/selling willingness indicator	Merchandise trade volume
9	Operation income growth	Moving average oscillators (MAO)	Export foreign exchange volume
10	Ordinary income growth	Buying/selling momentum indicator	Merchandise export
11	Continued income growth	Psychological line	Industry production index
12	Fixed asset turnover	Relative strength index (RSI)	Import price index increase rate
13	Inventory turnover	Rate of change (ROC)	Export price index increase rate
14	Net income growth	Stochastic slow	Foreign investment approval
15	Total asset growth	Disparity 5	Merchandise import
16	Operating Income	Momentum	
17	Return on total asset	Disparity 10	
18	Total asset growth		
19	Inventory turnover		
20	Days payables outstanding		

2.2.2 Technical features

Technical features are also referred as technical indicators as it indicates future price levels or direction from past patterns. It is a class of measurement, in which value is obtained from generic price activity in a particular stock. These are primarily used for analysing short term price movements. Some of the technical features being widely used are listed in Table 1. As per, Huang and Tsai (Lee, 2009), technical indexes are practiced on daily price variation in the stock. More technical features along with details can be referred from Tsai and Hsiao (2010) and Dash et al. (2014).

2.2.3 Macroeconomic features

Macroeconomic features are the external factors affecting the stock price. Literature states that factors like gross domestic product (GDP), inflation, exchange rate, etc. affect stock price (Tsai and Hsiao, 2010; Zhang et al., 2014). They play an important role in the stock price movement. Table 1 lists out some of the macroeconomic features widely being used.

The macroeconomic features vary from region to region. Thus, there exists no globally agreed upon list of macroeconomic features. The economic and political condition of one region may affect other region too. Tsai and Hsiao (2010) provides the list of few more macroeconomic features.

2.2.4 Market sentiments

Market sentiment is the market participants psychology. It may be the opinion of a single person or a group of persons and it may be often biased (Barak et al., 2015). It is considered as one of the important factor in stock price prediction. People give their opinions in social media which are accessible worldwide. The news also affects the stock market. Dynamics of market sentiment is being explored by Hagenau et al. (2013), de Fortuny et al. (2014), Barak et al. (2015) and Wing et al. (2014).

3 Prediction techniques

Choosing a prediction model is another key feature in obtaining a good accuracy for stock price prediction. Many techniques like artificial neural network (ANN) (Tsai and Hsiao, 2010; Zhang et al., 2014; Su and Cheng, 2016; Anish and Majhi, 2016; Das and Padhy, 2012), support vector machine (SVM) (Hagenau and Liebmann, 2013; Oztekin et al., 2016; Zhang et al., 2014; Su and Cheng, 2016; Huang and Tsai, 2009; Ni et al., 2011; Nguyen et al., 2015; Das and Padhy, 2018; Dash et al., 2014), fuzzy inference systems (Oztekin et al., 2016; Barak et al., 2015) are widely accepted prediction models. Neural networks can implicitly detect complex non-linear relationships between different features (Zhang et al., 2014). Opinion mining also has been used for sentimental analysis of stock related news or opinions.

Prediction techniques may vary in accordance with the feature set. The fusion of various prediction techniques may be used for better accuracy in stock price prediction. Next, we discuss in brief the standard widely used prediction techniques in literature for stock prediction.

3.1 Support vector machine

SVM from the ground of machine learning, is relevant to classification and regression among many supervised learning techniques. As per Vladimir Vapnik's statistical theory of learning algorithm, SVM is a supervised methodwhichcan classify as per prior prescribed classes or do regression. Because of this, SVM Algorithm is sometimes known as a support vector regression. SVM algorithms have been gaining more popularity due to its features and promising empirical conduct. As compared to the traditional empirical risk minimisation methods, SVM's structural risk minimisation features show much better and superior results. Expected risk of an upper bound is minimised by structural risk minimisation, whereas empirical risk minimisation lowers the miscalculation of training data. Thus, SVMs give a decent generalisation conduct with an efficiency of computation as far as the speed and complexity is concerned. It deals with multi-dimensional data easily. SVMs also work very well under different situations even when the sample dataset is small (Cristianini and Shawe-Taylor, 2000; Zhang et al., 2014).

From literature, it is found that several authors use SVR to match with their projected model. SVR is that the SVMs comprehend to unravel assessment issues of nonlinear regression. The concept of SVR depends on the calculation of a linear regression which performs in an exceedingly high dimension feature area wherever the in take information are mapped via a nonlinear function (Schumaker and Chen, 2006).

The SVM has been chosen by Oztekin and Kizilaslan (2016) as the main driver for miniature formation for text mining-based stock price prediction. For classification as buy/sell, the model has been measured with adaptive neuro fuzzy inference system and neural network based on MLP. It is found that SVM outperforms ANFIS and ANN in terms of its accuracy results. SVM is utilised on different mix of messages, which are expressed in feature vectors, and their subsequent price effect of stock (Hagenau et al., 2013). Further, the performance of ANN, Naive Bayes and SVMs have been compared and resulted in SVMs to be far better as far as performance is concerned as part of a pilot study.

A forecasting model which is based on SVM with a hybrid method to select features (F_SSFS) has been proposed as a stock trend predictor (Ni et al., 2011) reducing the number of training for SVM. With the combination of different prediction techniques, a data analytic approach has been developed which shows that stock price forecasting performance can be dramatically executed by utilising ANFIS, ANN, and SVM (Zhang et al., 2014). Examples of financial prediction using SVM applications. Least squares SVM (LS-SVM) utilises linear rather than quadratic programming, thus reducing computational complexity of the actual SVM technique (Gestel et al., 2003).Mapping the data to a features space is included by LS-SVM, which has a function formulated that can be utilised for TS prediction (Huang and Shyu, 2010). SVM has been utilised as a part of writing because of its capacity to deal with risk in an extremely successful way.

3.2 Artificial neural network

An ANN is an interconnected assembly of straightforward multiprocessing components, whose practicality is loosely supported animal neuron. The processing capability of the network is hold on within the inter unit affiliation strengths or weights, attained by a method of adaptation to or learning from a collection of training patterns (Hassoun, 1995). An ANN incorporates an oversized range of parameters that permits learning the intrinsic nonlinear link presented in time-series, expanding their forecasting potential (Arasaratnam and Haykin, 2008; Specht, 1990; Zhang et al., 2014). Main power of ANNs is that the capacity of versatile nonlinear operate approximation with a name deficiency (Cybenko, 1989). As a nonparametric and information driven model, ANNs do not need further presumptionsprior to the model formation. Varied issues and threats are related to ANNs. The chosen weights and thresholds have major hit on the prediction outcomes. During training the ANN, local optima are often found in – lieu of the global optimum. ANN continues to be thought-about as a 'black-box' and does not give intuitive narration of the prediction method.

Göçken et al. (2016) planned a hybrid ANN model for 45 technical indicators (features) that were reduced to 26 and 23 non-redundant options by GA and harmony search accordingly. It absolutely was found that stock price prediction efficiency of the HS-ANN is considerably higher than that of GA-ANN model and therefore the regular model of ANN. Higher prediction performance is found by combining multiple feature selection methods over single feature selection technique by Tsai and Hsiao (2010). They indeed combined PCA, GA and DT approach on union, intersection and multi-intersection level along side MLP for stock's fall/rise prediction so as to recommend for buy/sell severally.

3.3 Fuzzy inference system

An ANFIS will construct associate in nursing input-output mapping supported each human information within the kind of fuzzy if–then rules with membership of appropriate methods and provided input-output knowledge combinations. It implements a neural network in assurance of the form of membership functions and of the extraction of rule. ANFIS design adopts a composite processs of learning within the structure of adaptative networks (Jang, 1993). This illation system integrates the most effective options of formal logic and ANNs which handles the nonlinearity and unpredictability in the real-world systems (Lee and Ouyang, 2003). Fuzzy sets adaptation (Chaudhuri and De, 2011) is competent to handle unpredictability and impreciseness in forecasting of company knowledge. It is effective to find a set of optimum options and different parameters. Barak et al. (2015) gift a completely unique prognostication miniature for stock markets on the premise of the wrapper adaptive neural fuzzy inference system (ANFIS) – imperialist competitive algorithm (ICA) and technical analysis of Japanese holder. Imperialist competitive formula has been used to set choice.

C.H. Su and Cheng (2016) planned a time-series model that implemented planned ANFS methodology to pick the cheap variables into the ANFIS model, and used a model of adaptative expectation to enhance prognostication ability. The study in Su and Cheng (2016) planned associate in nursing INFS methodology to pick three vital technical indicators, and used the chosen technical indicators as intake variables for ANFIS model of prognostication to get the preliminary estimates.

4 Discussion and critical review

Stock market data are considered to be highly unpredictable. At any given time, there can be trends, random walks, cycles or combination of all which may result in unpredictable situation of stock market. Buying or selling a stock is a very complicated decision as there are several aspects that might influence the stock price values. Many external factors affect the stock values which in turn results in variation in stock market. It is being observed that the dynamic political scenarios have their own impacts along with fundamental and macroeconomic features on stock market trends. Identifying the influential factors is a key task in the field of stock market.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
			Macroeconomic feature			
Huang et al. (2005)	Combining model	9	NIKKEI 225 Index	Classification	The combination of models outperform the SVM, LDA, EBNN.	Combinations of other models may be explored with a larger input feature set.
Wang et al. (2015)	Dynamic Bayesian factor graph	9	Shenzhen Stock Exchange component index and S&P 500	Forecasting market trend	Change in market trends are captured effectively.	High computational complexity exists for large factor set and thus proper subset selection algorithm may be used to bring the improvement.
Altinbas et al. (2015)	Sequential forward selection with ANN	17	BIST 100	Prediction	One month lagged stock market indicator index values are sufficient for future prediction.	Larger feature set by integrating different types of features may be explored for better prediction performance.
Oztekin et al (2016)	Adaptive neuro fuzzy inference system, MLP-based neural network and SVM are deployed and compared	6	BIST 100 Index	Classification as buy/sell based on market rise/fall	SVM outperforms ANFIS and ANN in terms of its accuracy results.	Integration of different types of features may be explored for better prediction accuracy. Other data analytical models may be incorporated and compared with the existing model.
			Sentimental features			
Schumaker and Chen (2006)	SVM	Financial news	S&P 500	Prediction	The model using stock price and article terms give good accuracy. Also a proper noun scheme gives improved performance than bag of words.	Approach to test a model based on article terms and stock price change percentage with other machine learning techniques for various other stocks may be explored.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
			Sentimental features			
Sehgal and Song (2007)	Bayesian network	Textual news	Apple, ExxonMobile and Starbucks	Classification	Strong interrelationship between stock prices and web sentiment is found and sentiments can be used to make predictions about stock prices over a short-term (a day).	Model can be extended to make long-term predictions.
Hagenau et al. (2013)	Chi-square, bi-normal separation and frequency-based feature selection methods are compared along with SVM with linear kernel	Textual news	Textual news consisting of corporate announcements from Germany and the UK publications	Classification	Improved classification accuracy is found by Chi-2-based and BNS-based feature selection over frequency-based feature selection method.	The text base may consist of wide range of single text messages with less number of words. The study suggests that more texts may be allowed for good training and validation along with improved sentiment analysis.
Gunduz and Cataltepe (2015)	Balanced mutual information (BMI) with Naive bayes	Financial news	BIST 100 Index	Closing price prediction	Better macro averaged F-measure value is achieved by BMI in comparison to mutual information and Chi-square. The study reveals that incorporation of internet news and social media improves the prediction accuracy.	Combination of word occurrences along with semantic relations may be considered for improved outcomes. Intraday prediction may be considered as future research.
Zhu and Niu (2016)	Sentiment index based on PCA	Investor sentiment	Shanghai and Shenzhen A-share	Prediction	Investor sentiment is found to play a vital role in stock price prediction.	Other types of factors especially macroeconomic factors may be considered and explored along with sentimental features for improved prediction accuracy.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
			Technical features			
Grogan et al. (2005)	Multi-expression programming	5	Nasdaq-100, S&P CNX NIFTY	Prediction	The model gives lowest MAP value when compared to ANN, SVM, neuro fuzzy model and difference boosting neural network. The main task was to optimise different error measures.	It is difficult for one model to perform well for various stock indices.
Tan et al. (2005)	Genetic complementary learning (GCL) fuzzy neural	Closing price	Stock Exchange of Singapore (SES)	Classification	GCL does not need presumptions to be made on data. GCL is autonomous. Thus superior performance is shown by GCL over SVM, LR and MLP.	Other features affecting stock price may be considered for an increased GCL performance.
Qian and Rasheed (2007)	Ensemble of ANN, kNN, DT	6	DJIA Index	Prediction	Hurst exponent and chaos theory heuristics achieved good accuracy.	Hurst exponent may be used for guidance of data selection before forecasting.
Afolabi and Olude (2007)	Hybrid Kohonen self-organising map (KSOM)	64	Airline industries	Prediction	Hybrid Kohonen SOM is rapid and powerful with limited classification mismatch, hence, a finer predictor in comparison to Kohonen SOM and back propagation networks.	Hybrid Kohonen SOM is data dependent which needs to be removed. The tuning of the networks is a challenging task. Different fuzzy logic methods may be applied toward all the networks.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
				Technical features		
Kwon and Moon (2007)	GA combined with recurrent NN	75	NYSE and NASDAQ	Classification	Parallelised GA is implemented which shortens the response time without loss of performance. The model works efficiently with the transaction costs.	Portfolio optimisation may be done using the proposed prediction model. Future explorations may be like the number of stocks to be chosen and the number of shares of each stock to be purchased.
Huang et al. (2008)	Combination of wrapper with SVM, KNN, BP, DT	23	Korea and Taiwan stock	Classification	Better prediction is achieved by voting plus wrapper approach. The model shows improved performance when distinct classifiers are merged into the voting scheme.	Different classifiers combination may be explored along with other useful features for better stock prediction performance.
Powell et al. (2008)	K-means, SVM	9	S&P 500	Prediction	SVM gives better accuracy than K-means.	Model may be applied to larger and evenly distributed datasets with larger feature set.
Lin et al. (2009)	Echo state network (ESN)	5	S&P 500	Prediction	ESN is an efficient model for time series prediction with its capability of short-term memory. It gives improved performance over conventional neural networks. ESN is more advisable for stocks that oscillate a bit.	Common characteristics of dataset series need to be explored so that ESN can be applied. The long-term stock data mining ability of ESN may be explored.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
				Technical features		
Lee et al. (2009)	Integration of F_score and supported sequential forward search (F_SSFS) along with SVM	29	NASDAQ Index	Classification as +/– for index direction	Reduction in high computational cost and over-fitting is resulted. The model gives higher accuracy and generalisation performance over BPNN along with information gain, symmetrical uncertainty and correlation-based feature selection.	Selection of optimal parameter values in SVM may be done to achieve better prediction performance. Generalisation of SVM as per the training set size may be considered as a study.
Huang and Tsai (2009)	Hybridisation of support vector regression (SVR) with self-organising feature map (SOFM) along with coefficient of determination	13	Taiwan Index Futures (FITX)	Prediction of daily stock index	Improvement in prediction accuracy and reduction in training time is observed.	Optimisation of SVR parameters may be done by using different optimisation algorithms. Wrapper-based feature selection techniques may be used for further improvement of the outcome.
Ni et al. (2011)	Hybridisation of fractal feature selection and SVM	19	Sanghai Stock Exchange Composite Index (SSECI)	Prediction of stock price	Higher prediction accuracy is found in comparision to information gain, symmetrical uncertain, relief F and Cfs techniques.	Integration of different types of features may be done for improvement in prediction performance.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
			Technical features			
Das and Padhy (2012)	Back propagation technique (BP) and SVM	5	S&P CNX NIFTY, BANK NIFTY, S&P CNX 500, CNX INFRA, and CNX 100	Prediction of future price index	SVM gives better performance results than BP.	SVM performance may be compared with other techniques along with integrating different types of features like macroeconomic and sentimental features along with selected technical features for improvement in prediction.
Das and Padhy (2018)	Teaching learning-based optimisation hybridised with SVM	17	MCX COMDEX	Prediction	The optimal free parameters for SVM regression model is found and the predictions are better than the standard SVM.	The accuracy may be enhanced by integrating efficient macroeconomic features along with selected technical features.
Ng et al. (2014)	LG-trader using NSGA-II and MLPNN	17	TESCO and DJIA	Classification as buy/hold	Features are optimised and the model outperforms buy and hold strategy.	The model may be explored with other multi-objective optimisation algorithms along with a large feature set. MLPNN may be replaced with support vector machines and compared with the model.
Sukhija (2014)	Fixed effect model and random effect model along with Hausman test	8	BSE 200	Prediction	Dividend per share were found to be the important determinants of share prices.	Macroeconomic and sentimental factors may be considered for further enhancing the prediction accuracy.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
			Technical features			
Zbikowski (2015)	Fishers method combined with SVM	6	QCOM, EBAY, NWL, ACT, CCI, JPM, MCD, USB, BMS, LNC, MAC, NWL, SYK, YHOO	Prediction	The study reveals that input vector delays in technical indicators enhance the prediction performance.	Other feature selection techniques may be considered along with different types of features (macroeconomic and sentimental) for better prediction accuracy.
Barak et al. (2015)	Imperialist competitive algorithm (ICA) combined with ANFIS (ANFIS-ICA)	24	General motors company at New York Stock Exchange	Prediction of stock trading	In comparision to GA, feature selection by ICA has a positive effect on the ANFIS's efficiency.	Use of other metaheuristic approaches may be explored along with filter and wrapper methods on fundamental and textual features to increase prediction accuracy.
Patel et al. (2015)	SVR, ANN, RF, SVR	10	CNX and BSE	Prediction	Fusion of models gives a decreased error rate.	Macroeconomic and sentimental factors may be considered for further enhancing the prediction accuracy.
Kumar et al. (2016)	Linear correlation (LC), rank correlation (RC), regression relief (RR), random forest (RF) are combined with proximal support vector machines (PSVM) respectively	55	BSE Sensex, CNX Nifty, GDAXI, HSI	Classification	Highest accuracy is achieved by random forest clubbed with PSVM (RC-PSVM) in comparison to LC-PSVM, RC-PSVM and RR-PSVM.	Integration of fundamental and macroeconomic features along with technical features may be explored for improved results.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
			Technical features			
Su and Cheng (2016)	Integrated nonlinear feature selection (INFS) integrated with adaptive neuro fuzzy inference system (ANFIS)	13	Taiwan Stock Exchange and HSI Index	Prediction of next day stock index	The proposed model results in strengthening of forecasting performance due to utilisation of adaptive expectation model over weighted fuzzy time series models.	Integration of different types of features may be explored for further improvement in prediction performance. The model may be explored with different optimisation techniques.
Bhattacharya et al. (2016)	Interval Type-2 fuzzy reasoning with triangular and Gaussian membership function (MF) along with t-test for statistical analysis	6	TAIEX, DOWJONES and NASDAQ	Stock index prediction	Stock index prediction is improved by employing triangular MF as compared to Gaussian MF with root mean square error taken as the metric.	Integration of different types of features along with general type-2 fuzzy sets may be explored for better performance.
Göçken et al. (2016)	GA-ANN and harmony search (HS) – ANN	45	Turkish stock market	Prediction	Feature selection complexity is reduced to half and optimum number of neurons in hidden layer is found resulting in elimination of overfitting/ underfitting of ANN.	Parameters of ANN may be changed and explored along with variants of GA and HS to increase the prediction accuracy.
Anish and Majhi (2016)	Feedback type of FLANN (FFLANN) with recursive least square training and factor analysis	10	DJIA AND S&P 500	Prediction	Prposed model is found to be computationally efficient in comparison to SVM, RBFNN and MLANN and reduction in training time is observed.	GA and PSO may be implemented to train the weights of the model so as to solve the local minima problem in order to attain better prediction efficiency.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
				Technical features		
Das et al. (2014)	Self adaptive differential harmony search technique along with extreme learning machine (ELM)	17	CNX Nifty and BSE Sensex	Prediction	Forecasting performance is improved in comparison to ELM, differential evolution (DE) and DE-OELM.	This model may be applied to handle various problems of financial time series forecasting. Systematic ways for choosing optimal network size for both RBF and FLANN models may be explored. Fuzzy logic may be incorporated to handle uncertainty and outliers in dataset.
Dash and Dash (2016)	Computational efficient functional link artificial neural network (CEFLANN) integrated with ELM	6	BSE SENSEX and S&P 500	Prediction	The model provides superior prediction accuracy results in comparison to SVM, NB, KNN and DT.	Structure optimisation of the model by integrating different optimisation techniques along with more technical indicators may be considered for better result.
Zhong and Enke (2017)	PCA, FRPCA, KPCA and ANN	60	US SPDR S&P 500 ETF	Classification	ANN-PCA gives better prediction efficiency and higher risk-adjusted profits as measured against FRPCA and KPCA.	Other kernel functions with relevane parameters may be suggested.
Chen and Hao (2017)	Information gain, Feature weighted support vector machine and feature weighted K-nearest neighbour	9	SSE Composite Index, SZSE COMP SUB IND	Classification and prediction	Relative significance of each feature were considered for the classification in SVM and prediction in KNN. The model provides better accuracy than SVM and KNN.	Other correlation weighted techniques may be examined and explored along with a larger feature set.

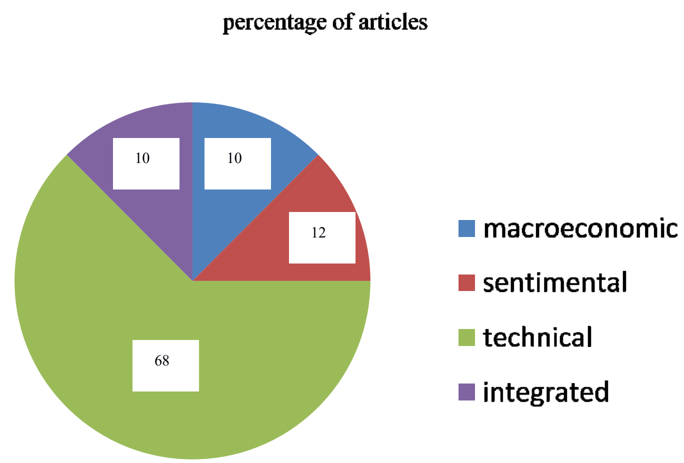
Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
			Technical features			
Chong et al. (2017)	PCA, autoencoder (AE), restricted Boltzmann machine (RBM), deep neural network	7	KOSPI	Classification	DNNs perform better than a linear autoregressive model in the training set, but the advantage mostly fades in the test set. model to covariance-based market structure analysis and find that their model enhances covariance estimation dramatically.	Exploiting other factors may give better performance accuracy. Also the relationship between the stock prices and other features may be explored using deep neural network. Data representation methods and network functions other than those examined may also offer improved performance.
			Integrated features			
Tsai and Hsiao (2010)	PCA, GA and DT approach has been combined on union, intersection and multi-intersection level along with MLP	85 features including fundamental and macroeconomic features	Taiwan Economic Journal (TEJ), Taiwan Stock Exchange (TSE)	Prediction of stock's fall/rise and suggest for buy (rising stocks) and sell (falling stock) respectively	Better prediction performance is found by combining multiple feature selection methods over single feature selection method. It is observed that multi-intersection method performs the best over union and intersection methods, providing highest accuracy with lowest error rate.	Different variants of PCA may be applied and explored along with other feature selection techniques apart from the ones used in the model.
de Fortuny et al. (2014)	Support vector machine	4 technical indicators and news article	Online articles of all major flemish newspaper, Euronext Brussels Stock Exchange	Classification/ movement direction (up/down)	Efficient market hypothesis is validated.	More features may be included and working with sentiment of verbs and nouns based on sentiment models may be explored along with the suggested features.

Table 2 Summarisation of analysis of feature selection techniques in stock price prediction (continued)

Author (year)	*Technique used*	*No. of input features*	*Dataset(s)*	*Classification/ prediction strategy*	*Results obtained*	*Observation(s)*
				Integrated features		
Zhang et al. (2014)	Causal feature selection (CFS) combined with SVM, NN and NB	50 fundamental and momentum variables	Shanghai Stock Exchange	Classification as 0/1 for positive or negative return respectively	CFS outperforms PCA and DT in terms of prediction accuracy. It achieved stable prediction accuracy along with SVM, NN and NB models.	Further reduction in features may be explored by integrating other types of features along with the existing ones without compromising the prediction performance. Further study may be done to reduce the complexity of the model.
Nguyen et al. (2015)	Sentiment analysis and opinion mining	Price and sentiments	18 stocks from Yahoo Finance Message Board	Prediction	Effectiveness of incorporation of the sentiment analysis is observed.	A non-parametric topic model capable of automatically extracting the number of topics and sentiments for stock prediction may be incorporated. More number of features may be considered for better accuracy.

Figure 3 Numbers of published articles considered on feature selection for stock prediction for different types of features (2005–2017) (see online version for colours)



Articles surveyed mainly focus on feature selection for the stock price prediction of either single stock index or multiple stock market indexes. But many other studies focus in forecasting the return of single or multiple stocks. Articles in Table 2 may be divided in four sections: category 1 (macroeconomic feature), category 2 (sentimental feature), category 3 (technical features), and category 4 (integrated features).

A very significant part of decision making is based on data collection and consolidation via data analytic modelling, so this kind of a Parsimonious model definitely will be very interesting to the decision makers.

To simplify the intricacy of the model and achieve frugality, decreasing the number of input variables required for estimating is very critical which is a well-known factor, and hence to quicken the decision making operation and reduce the computation time is also critical. In turn expanding the number of variables becomes harder to access and/or collect as well resulting in high computational complexity (Wang et al., 2015).

The first category in Table 2 includes articles that use macroeconomic features as input data features. Macroeconomic feature of well-developed North America, Western Europe and other solid economy countries has been considered under this category. It is evident from Table 2 that generally, 10 or more input variables are used in most of the studies under category 1. There are also examples like Olson and Mossman (2003) who used 59 input variables and Zorin and Borisov (2002) who used 61 variables. Exchange rate, price of oil, rate of interest, inflation rate, foreign exchange, GDP and import/export price are the most common macroeconomic features used as input variables (Wang et al., 2015; Oztekin et al., 2016; Altinbas and Biskin, 2015). Internal and external political uncertainty mostly causes the rise and fall of the GDP. In addition to the common macroeconomic features many other features are also used in other papers, which varies from country to country. GDP usually impacts one country/location to other country/location. Hence countries affected by macroeconomic features are also considered in many papers.

It is evident that financial assimilation exists across the globe from the experimentation and writings (Fischer and Palasvirta, 1990). Fluctuation of stock market usually flows among the developed countries and gets transmitted to emerging markets.

Fischer and Palasvirta (1990) studied that there exists an immense statistical relationship among 23 stock indices. Oztekin et al. (2016) represented a collection of developed and emerging markets, with US stock prices dominating the rest of the world. Thus, for the forecasting of the BIST 100, it is sensible to use a US index. Directional movements in the US indices, like the DJIA, the Nasdaq Composite, or the S&P 500 mirror each other approximately. From Table 2, it is evident that a combination of macroeconomic features along with other features for stock prediction may give better prediction accuracy.

The second category in Table 2 includes articles that use sentimental features as input data features. Textual news and textual opinion affect the stock market in a very effective manner. Thus, we may say that market sentiments have greater impact on the stock prediction results. Often, the buy and sell of stock is biased by the user sentiments. Different techniques have been used to evaluate the sentiments for prediction of stock. Balanced mutual information along with Naive Bayes has been used in Gunduz and Cataltepe (2015) for closing price prediction taking into consideration the financial news. Chi-square-based feature selection gives good classification accuracy in Hagenau et al. (2013) for the textual news from Germany and UK publications. Chi-square, bi-normal separation and frequency-based feature selection methods are compared along with SVM with linear kernel for sentimental features. The study reveals the improvement of stock price prediction due to the sentimental features. Many techniques have been used for sentimental features and it is found that more noun, verbs and words should be allowed for good training and validation purpose.

The third category in Table 2 summarises articles that use technical features as input data features. From literature, it is found that for stock price prediction, usually technical features are used as input feature sets (Oztekin et al., 2016; Huang and Tsai, 2009; Lee, 2009; Ni et al., 2011; Nguyen et al., 2015; Göçken et al., 2016; Anish and Majhi, 2016; Altinbas and Biskin, 2015; Das and Padhy, 2018; Dash et al., 2014; Dash and Dash, 2016). Usually, 7 to 10 technical indicators are used for stock price prediction. But, in some papers more than 25 technical features have been used as seen from Table 2. The most common technical features used are Simple moving average (SMA), Exponential moving average (EMA), Stochastic oscillator (STOC %K and %D), Relative strength index (RSI), Price rate of change (PROC), Closing price acceleration (CPACC), High price acceleration (HPACC), moving average convergence/divergence (MACD). Historical stock prices are used by technical analysts for calculating indicators which are plotted together with price of stocks on the same chart. These charts are used by investors to search for particular pattern that indicates movements of the market in future and in turn adds basis and trend for their decision making. For technical analysis, there are multitude of indicators which are proposed and each of these indicators provides different information about the market. Examples are: trends of the market are indicated by *moving averages*, whereas *relative strength* and *momentum* gives a perspective how a given stock is *overbought* or *oversold*.

The fourth category in Table 2 includes articles that uses an integrated set of features as input data features. A combination of technical features along with sentimental features gives a better accuracy as seen from work done by de Fortuny et al. (2014) and Nguyen et al. (2015). 85 features including fundamental and macroeconomic features has been used for stock prediction in Tsai and Hsiao (2010). The study reveals that helping to improve the prediction accuracy a combination of different types of features would be even much better for stock price prediction.

From Table 2, it is evident that the count of input variables practiced, differs from model to model. Average count of input variables adopted in most of the models are between four to ten, however, there are models where only two input variables are used. There are also example of Olson and Mossman (2003) where 59 input variables used and Zorin and Borisov (2002) where 61 input variables used. The statistics of the published papers on feature selection for stock prediction for different types of features during 2009 to 2016 is given in Figure 3.

Different methods are also applied to find the most significant input variables for the forecasting procedure amid a large number of candidate ones, depending on how particular input influences the attained outcomes. A large number of observations are covered by some studies over a period of years. Whenever there is a missing observation, the mean value or the latest stock value observed is used to fill. The opening or closing price of the stock index is most commonly used inputs, also the daily lowest and highest values are commonly used. Most of the surveyed articles use the daily closing price or some indicator relying only on it as input.

5 Conclusions

Stock prediction helps the organisations and the stake holders to keep track of the market and choose accordingly whether to sell, buy or withheld the stock in order to maximise the profit. Our survey aims to help the stock brokers and investors for investing money in the stock market. Articles surveyed mainly focus on feature selection for the stock price prediction of either single stock index or multiple stock market indexes. The review of literature on stock price prediction reveals that the stock price prediction is highly dependent on the choice of dataset and input features which may vary from one methodology to another. A very significant part of decision making is based on data collection and consolidation via data analytic modelling, so this kind of a parsimonious model definitely will be very interesting to the decision makers.

Average number of input variables used in most of the models is between four to ten, however, there are models where more than 30 input variables have been used. The study reveals that the integration of different types of features like technical, macroeconomic and sentimental may result in better prediction model. Amongst many feature selection methods, the most widely used methods are PCA, GA and decision trees as they contribute in selecting good number of representative features. The study suggests that the fusion of various feature selection techniques may be used for better accuracy in stock price prediction. SVM and ANN are found to be widely used prediction models. On the basis of our survey, it is observed that the usage of a large set of input features may increase the performance efficiency of the prediction model. But, large set of input features also increases the computational complexity of the model. To simplify the intricacy of the model and achieve frugality, decreasing the number of input variables required for estimating is very critical which is a well-known factor, and hence to quicken the decision making operation and reduce the computation time is also critical. In turn expanding the number of variables becomes harder to access and/or collect as well resulting in high computational complexity. The study put forward the amalgamation of different feature selection techniques on a large set of integrated features to achieve higher degree of prediction on stock market with an aim of reducing the computational complexity of the model. In addition, presently it is difficult to specify the most effective approach in stock prediction territory, which may be considered as a future research issue to perform comparative survey based on all these existing approaches.

References

Afolabi, M.O. and Olude, O. (2007) 'Predicting stock prices using a hybrid Kohonen self organizing map (SOM)', in *Proceedings of the 40th Hawaii International Conference on System Sciences*.

Altinbas, H. and Biskin, O.T. (2015) 'Selecting macroeconomic influencers on stock markets by using feature selection algorithms', in *Procedia Economics and Finance*, Vol. 30, pp.22–29.

Anish, C.M. and Majhi, B. (2016) 'Hybrid nonlinear adaptive scheme for stock market prediction using feedback FLANN and factor analysis', in *Journal of the Korean Statistical Society*, Vol. 45, No. 1, pp.64–76.

Arasaratnam, I. and Haykin, S. (2008) 'Nonlinear Bayesian filters for training recurrent neural networks', *Mexican International Conference on Artificial Intelligence*, Springer, Berlin, Heidelberg, pp.12–33.

Barak, S., Dahooie, J.H. and Tichý, T. (2015) 'Wrapper ANFIS-ICA method to do stock market timing and feature selection on the basis of Japanese candlestick', in *Expert Systems with Applications*, Vol. 42, No. 23, pp.9221–9235.

Bhattacharya, D., Konar, A. and Das, P. (2016) 'Secondary factor induced stock index time-series prediction using self-adaptive interval type-2 fuzzy sets', in *Neurocomputing*, Vol. 171, pp.551–568.

Chaudhuri, A. and De, K. (2011) 'Fuzzy support vector machine for bankruptcy prediction.', *Applied Soft Computing*, Vol. 11, No. 2, pp.2472–2486.

Chen, Y. and Hao, Y. (2017) 'A feature weighted support vector machine and K-nearest neighbor algorithm for stock market indices prediction', in *Expert systems with Applications*, Vol. 80, pp.340–355.

Chong, E., Han, C. and Park, F.C. (2017) 'Deep learning networks for stock market analysis and prediction: methodology, data representations, and case studies', in *Expert systems with Applications*, Vol. 83, pp.187–205.

Cristianini, N. and Shawe-Taylor, J. (2000) *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge University Press.

Cybenko, G. (1989) 'Approximation by superpositions of a sigmoidal function', *Mathematics of Control, Signals and Systems*, Vol. 2, No. 4, pp.303–314.

Das, S.P. and Padhy, S. (2012) 'Support vector machines for prediction of futures prices in Indian stock market', in *International Journal of Computer Applications*, Vol. 41, pp.975–8887.

Das, S.P. and Padhy, S. (2018) 'A novel hybrid model using teaching-learning-based optimization and a support vector machine for commodity futures index forecasting', in *Int. J. Mach. Learn. & Cyber*, DOI: 10.1007/s13042-015-0359-0.

Dash, R. and Dash, P.K. (2016) 'A hybrid stock trading framework integrating technical analysis with machine learning techniques', in *The Journal of Finance and Data Science*, Vol. 2, No. 1, pp.42–57.

Dash, R., Dash, P.K. and Bisoi, R. (2014) 'A self-adaptive differential harmony search based optimized extreme learning machine for financial time series prediction', in *Swarm and Evolutionary Computation*, Vol. 19, pp.25–42.

de Fortuny, E.J., De Smedt, T., Martens, D. and Daelemans, W. (2014) 'Evaluating and understanding text-based stock price prediction models', in *Information Processing and Management*, Vol. 50, pp.426–441.

Fischer, K.P. and Palasvirta, A.P. (1990) 'High road to a global marketplace: the international transmission of stock market fluctuations', *Financial Review*, Vol. 25, No. 3, pp.371–394.

Göçken, M., Özçalıcı, M., Boru, A. and Dosdŏgru, A.T. (2016) 'Integrating metaheuristics and artificial neural networks for improved stock price prediction', in *Expert Systems with Applications*, Vol. 44, pp.320–331.

Grogan, C., Abraham, A., Ramos, V. and Han, S.Y. (2005) 'Stock market prediction using multi expression programming', in *Proceedings of the 12th Portuguese Conference on Artificial Intelligence*.

Gunduz, H. and Cataltepe, Z. (2015) 'Borsa Istanbul (BIST) daily prediction using financial news and balanced feature selection', in *Expert Systems with Applications*, Vol. 42, No. 22, pp.9001–9011.

Hagenau, M., Liebmann, M. and Neumann, D. (2013) 'Automated news reading: stock price prediction based on financial news using context-capturing features', in *Decision Support Systems*, Vol. 55, No. 3, pp.685–697.

Huang, C-J., Yang, D-X. and Chuang, Y-T. (2008) 'Application of wrapper approach and composite classifier to the stock trend prediction', in *Expert Systems with Applications*, Vol. 34, No. 4, pp.2870–2878.

Huang, C-L. and Tsai, C-Y. (2009) 'A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting', in *Expert Systems with Applications*, Vol. 36, No. 2, pp.1529–1539.

Huang, W., Nakamori, Y. and Wang, S-Y. (2005) 'Forecasting stock market movement direction with support vector machine', in *Computers & Operations Research*, Vol. 32, No. 10, pp.2513–2522.

Huang, Z. and Shyu, M-L. (2010) 'k-NN based LS-SVM framework for long-term time series prediction', *IEEE International Conference on Information Reuse & Integration*, pp.69–74, IEEE.

Kumar, D., Meghwani, S.S. and Thakur, M. (2016) 'Proximal support vector machine based hybrid prediction models for trend forecasting in financial markets', *Journal of Computational Science*, Vol. 17, pp.1–13.

Kwon, Y-K. and Moon, B-R. (2007) 'A hybrid neurogenetic approach for stock forecasting', in *IEEE Transactions on Neural Networks*, Vol. 18, No. 3, pp.851–864.

Lee, M-C. (2009) 'Using support vector machine with a hybrid feature selection method to the stock trend prediction', in *Expert Systems with Applications*, Vol. 36, No. 8, pp.10896–10904.

Lee, S.J. and Ouyang, C.S. (2003) 'A neuro-fuzzy system modeling with self-constructing rule generationand hybrid SVD-based learning.', *IEEE Transactions on Fuzzy Systems*, Vol. 11, No. 3, pp.341–353.

Lin, X., Yang, Z. and Song, Y. (2009) 'Short-term stock price prediction based on echo state networks', in *Expert Systems with Applications*, Vol. 36, No. 3, pp.7313–7317.

Ng, W.W.Y., Liang, X-L., Li, J., Yeung, D.S. and Chan, P.P.K. (2014) 'LG-trader: stock trading decision support based on feature selection by weighted localized generalization error model', in *Neurocomputing*, Vol. 146, pp.104–112.

Nguyen, T.H., Shirai, K. and Velcin, J. (2015) 'Sentiment analysis on social media for stock movement prediction', in *Expert Systems with Applications*, Vol. 42, No. 2, pp.9603–9611.

Ni, L-P., Ni, Z-W. and Gao, Y-Z. (2011) 'Stock trend prediction based on fractal feature selection and support vector machine', in *Expert Systems with Applications*, Vol. 38, No. 5, pp.5569–5576.

Olson, D. and Mossman, C. (2003) 'Neural network forecasts of Canadian stock returns using accounting ratios', *International Journal of Forecasting*, Vol. 19, No. 3, pp.453–465.

Oztekin, A., Kizilaslan, R., Freund, S. and Iseri, A. (2016) 'A data analytic approach to forecasting daily stock returns in an emerging market', in *European Journal of Operational Research*, Vol. 253, No. 3, pp.697–710.

Patel, J., Shah, S., Thakkar, P. and Kotecha, K. (2015) 'Predicting stock market index using fusion of machine learning techniques', *Expert Systems with Applications*, Vol. 42, No. 4, pp.2162–2172.

Powell, N., Foo, S.Y. and Weatherspoon, M. (2008) 'Supervised and unsupervised methods for stock trend forecasting', in *40th Southeastern Symposium on System Theory*, University of New Orleans, New Orleans, LA, USA, 16–18 March.

Qian, B. and Rasheed, K. (2007) 'Stock market prediction with multiple classifiers', in *Applied Intelligence*, Vol. 26, No. 1, pp.25–33.

Schumaker, R. and Chen, H. (2006) 'Textual analysis of stock market prediction using financial news articles', in *Proceedings of the Twelfth Americas Conference on Information Systems*, Acapulco, Mexico, 4th–6th August.

Sehgal, V. and Song, C. (2007) 'SOPS: stock prediction using web sentiment', in *Proceedings of the Seventh IEEE International Conference on Data Mining – Workshops*.

Specht, D.F. (1990) 'Probabilistic neural networks', *Neural Networks*, Vol. 3, No. 1, pp.109–118.

Su, C-H. and Cheng, C-H. (2016) 'A hybrid fuzzy time series model based on ANFIS and integrated nonlinear feature selection method for forecasting stock', in *Neurocomputing*, Vol. 205, pp.264–273.

Sukhija, S. (2014) 'An explicit model on fundamental factors affecting stock prices of BSE listed companies in India: an inter industry approach', in *European Journal of Business and Management*, Vol. 6, No. 37.

Tan, T.Z., Quek, C. and Ng, G.S. (2005) 'Brain-inspired genetic complementary learning for stock market prediction', in *Proceedings of the IEEE Congress on Evolutionary Computation*.

Tsai, C-F. and Hsiao. Y-C. (2010) 'Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches', in *Decision Support Systems*, Vol. 50, No. 1, pp.258–269.

Wang, L., Wang, Z., Zhao, S. and Tan, S. (2015) 'Stock market trend prediction using dynamical Bayesian factor graph', in *Expert Systems with Applications*, Vol. 42, Nos. 15–16, pp.6267–6275.

Zbikowski, K. (2015) 'Using volume weighted support vector machines with walk forward testing and feature selection for the purpose of creating stock trading strategy', in *Expert Systems with Applications*, Vol. 42, No. 4, pp.1797–1805.

Zhang, X., Hu, Y., Xie, K., Wang, S., Ngai, E.W.T. and Liu, M. (2014) 'A causal feature selection algorithm for stock prediction modelling', in *Neurocomputing*, Vol. 142, pp.48–59.

Zhong, X. and Enke, D. (2017) 'Forecasting daily stock market return using dimensionality reduction', *Expert Systems with Applications*, Vol. 67, pp.126–139.

Zhu, B. and Niu, F. (2016) 'Investor sentiment, accounting information and stock price: evidence from China', in *Pacific-Basin Finance Journal*, Vol. 38, pp.125–134.

Zorin, A. and Borisov, A. (2002) 'Traditional and index tracking methods for portfolio construction by means of neural networks', *Network*, Vol. 1, p.1.