International Journal of Cognitive Computing in Engineering 4 (2023) 100–108

Contents lists available at ScienceDirect

International Journal of Cognitive Computing in Engineering

journal homepage: https://www.keaipublishing.com/en/journals/international-journal-of-cognitive-computing-in-engineering/

Progress and prospects of data-driven stock price forecasting research

Chuanjun Zhao[a,c,*], Meiling Wu[a], Jingfeng Liu[a], Zening Duan[b], Jie li[a], Lihua Shen[d], Xuekui Shangguan[d], Donghang Liu[d], Yanjie Wang[d]

[a] School of Information, Shanxi University of Finance and Economics, Taiyuan, 030006, Shanxi, China
[b] School of Journalism and Mass Communication, University of Wisconsin-Madison, Madison, 53706, WI, USA
[c] Financial Innovation and Big Data Statistical Analysis Laboratory, Shanxi University of Finance and Economics, Taiyuan, 030006, Shanxi, China
[d] Shanxi Information Technology Application Innovation Engineering Research Center, Taiyuan, 030006, Shanxi, China

A R T I C L E I N F O

Keywords:
Stock price forecasting
Time series
Machine learning
Numerical and textual data
Research review

A B S T R A C T

With the rapid development of social economy and the continuous improvement of stock market, stock investment has become more and more widely concerned. Stock price prediction has become an important research direction in the field of cognitive computing in engineering. Data-driven stock price forecasting aims to predict future stock price trends based on historical values and textual data, which can effectively help people reduce risks and improve returns in the process of stock investment. The article reviews the literature on stock price forecasting methods, and classifies stock price forecasting methods from two different perspectives of model and feature. According to different model angles, the existing stock price prediction methods can be divided into statistical analysis methods, traditional machine learning methods and deep learning methods. According to different characteristic angles, the existing stock price prediction methods can be divided into those based on numerical data and those based on text mixed with numerical data. Finally, we summarize the research challenges faced by stock price prediction and provide future research directions.

1. Introduction

It has been shown that stock prices do not follow exactly the characteristics of random wandering and the assumption of efficient markets, while in the stock market their price fluctuations have a long-term memory (Chen et al., 2022; Liu et al., 2018). The stock market is characterized by high risk and high return (Zhang et al., 2022). Both securities institutions and investors are very interested in the predictive analysis of the stock market and try to find the trend of stock price fluctuations. Stock price forecasting aims to forecast future stock price trends based on historical stock price movements. It can help institutions and investors to grasp the operation of the stock market, effectively helping people to reduce risks and increase returns in the process of stock investment (Agustini et al., 2018; Jing et al., 2022; Zhang et al., 2021). It is an important research component of cognitive computing in engineering.

As part of the financial market, the trend of stock price fluctuations reflects the changing patterns of the economic market. The stock market is a complex system. First, its fluctuations are influenced not only by the basic situation of stock companies, but also by uncertainties such as national economic policies, industry development status, investors decisions and news opinion (Bogle and Potter, 2015; Sumathi et al., 2022; Zhang et al., 2020). Second, the stock data itself, as time-series data, has non-linear and non-smooth characteristic. Third, the stock market also tends to be stochastic. These leads to a certain difficulty in stock price prediction.

In recent years, researchers have made stock price predictions based on different learning methods and different data. Learning methods include statistical analysis methods, traditional machine learning based methods and deep learning based methods. Data has both numerical data and textual data. From traditional fundamental analysis methods (Grimm, 2012; Jain et al., 2022; Wang and Wang, 2019) to simple regression analysis (Sahoo and Charlapally, 2015; Xie et al., 2009), from time series statistical analysis methods represented by autoregressive moving average models (ARMA) (Fu and Li, 2012) to machine learning models using Bayesian network (BN) (Setiani et al., 2021; Yi and Kita, 2012), wavelet neural network (WNN) (Kulaglic and Ustundaug, 2018; Wang et al., 2017), and deep learning models represented by recurrent neural network (RNN) (Berradi and Lazaar, 2019; Pahlawan et al., 2021; Zhu, 2020), long-short term memory (LSTM) (Hargreaves and Chen, 2020; Hu, 2021; Rather, 2021), from numerical data to data-driven analysis of multi-source heterogeneous data (i.e., numerical and textual data) (Ding et al., 2014; Dong et al., 2017; Nguyen and Shirai, 2015; Schumaker and Chen, 2009; Vega, 2006; Yu et al., 2019; Zhang et al., 2018; Zhao et al., 2022; Zhou, 2018). They have all achieved bet-

* Corresponding author.
E-mail address: zhaochj@sxufe.edu.cn (C. Zhao).

https://doi.org/10.1016/j.ijcce.2023.03.001
Received 20 December 2022; Received in revised form 28 February 2023; Accepted 1 March 2023
Available online 6 March 2023

ter results in the field of stock price prediction. Stock price prediction, as one of the research hotspots in the field of data mining, is more reliable and accurate based on data-driven methods. Numerical data contains historical trading data such as stock opening and closing prices and volume, and textual data contains stock-related financial news and stock bar comments. The data-driven approach automatically learns the parameters of statistical or machine learning models from numerical and textual data, which effectively reduces the cost of manual forecasting and improves the accuracy and robustness of forecasting results.

This paper first introduces the background knowledge related to stock price forecasting in Section 2. This paper collects and organizes literature related to stock price forecasting, and reviews domestic and international research related to stock price forecasting from two perspectives. Stock price forecasting is divided into statistical analysis methods, traditional machine learning-based methods and deep learning-based methods according to the perspective of different models in Section 3. Stock price forecasting methods based on numerical data and text data according to the source of information are summarized and analyzed in Section 4. The current research challenges and possible future breakthroughs in stock price forecasting are also analyzed in Section 5. Finally, Section 6 concludes the whole paper.

2. Background knowledge

2.1. Problem description

Stock price forecasting is based on accurate statistical data and historical information of the stock market. Based on the history, current situation and laws of the stock market, researchers use reasonable methods to forecast the data of stocks and their trends in the future period. Its essence is financial time series analysis. In stock price forecasting, there usually exist several historical trading data such as opening price, closing price, volume and turnover. The time series of stock prices over a continuous period of time is noted as

$$C_1, C_2, C_3, \ldots, C_t, C_{t+1} \tag{1}$$

where C_t denotes the price of the stock on day t, and C_{t+1} is the forecast price of the stock on the next day. If $C_{t+1} > C_t$, the stock price rises; if $C_{t+1} = C_t$, the stock price remains unchanged; if $C_{t+1} < C_t$, the stock price falls.

2.2. Dataset resources

In terms of resources, stock price prediction datasets can be divided into two types based on information sources: numerical data-based and numerical and text data-based. Among them, numerical data-based datasets contain only numerical data, while numerical and text data-based datasets contain data from both numerical and textual information. Among the datasets based on numerical data such as SSE Composite Index (000001) (Fu and Li, 2012; Huang and Liu, 2019; Liu et al., 2018; Pan and Ding, 2000; Wu and Yang, 2013; Zhang et al., 2012), CSI 300 Index (000300) (Cui et al., 2019; Hu, 2021; Shi and Gao, 2015; Song et al., 2019; Zhang, 2020; Zhou, 2018), S&P 500 index (Chen and Huang, 2021; Ding et al., 2014; Hoseinzade and Haratizadeh, 2019; Schumaker and Chen, 2009; Shen et al., 2018; Wang et al., 2021), time-series data of the CSI 500 Index (000905) (Bao et al., 2020), Hengseng (HSI) index (Shen et al., 2018; Tang and Sun, 2003), and closing price datasets of eight stocks including APA and TSLA (Kan, 2019). The datasets based on numerical and textual data include closing price data and related messages for XOM, DELL, EBAY, IBM, KO (Nguyen and Shirai, 2015), Alibaba stock data and related news information data (Zhang et al., 2018), etc. Table 1 summarizes some representative dataset resources for stock price prediction.

2.3. Stock price forecasting platform

As financial stock markets continue to mature and stock price forecasting research continues to emerge, researchers have now developed a large number of stock price forecasting platforms and tools. Some of the published stock price forecasting platforms are shown in Table 2.

2.4. Technical specifications

Technical indicators are based on statistics derived from stock trading data as an analysis system. The trading data includes stock closing price, turnover rate, volatility and other factors. The main purpose of technical indicators is to analyze the market pattern of stock price fluctuations and to determine the main trends and timing of buying and selling. The commonly used technical indicators are shown in Table 3.

2.5. Evaluation indicators

The so-called evaluation indicators are the criteria for evaluating a model's ability to predict data as well as its high or low generalization ability, and there are six main indicators as follows:

(1) Accuracy is the most commonly used evaluation index, describing the ratio of the number of samples with accurate predictions to the number of all samples. Its expression is shown in Eq. (2):

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{2}$$

where TP and TN are the data judged to be correct, FP and FN are the data judged to be incorrect.

(2) Mean absolute error (MAE) indicates the average of the absolute deviation of each predicted value. Its expression is shown in Eq. (3):

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |forecast(i) - actual(i)| \tag{3}$$

where $forecast(i)$ is the predicted value of the stock on a given day, $actual(i)$ is the actual value of the stock on a given day, and n is the sample size of the test data set.

(3) The expression of mean absolute percentage error (MAPE) is shown in Eq. (4):

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} \left|\frac{forecast(i) - actual(i)}{actual(i)}\right| \tag{4}$$

(4) Mean square error (MSE) is the mean value of the sum of squares of the errors of deviations of the predicted values from the original values. The smaller the value, the more accurate the model prediction is. Its expression is shown in Eq. (5):

$$MSE = \frac{1}{n}\sum_{i=1}^{n} ((forecast(i) - actual(i))^2 \tag{5}$$

(5) Root mean squared error (RMSE) is a deformation of MSE, which is a good measure of how much the data deviate from the true value. Its expression is shown in Eq. (6):

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} ((forecast(i) - actual(i))^2} \tag{6}$$

(6) Sharpe ratio is used to provide a comprehensive description of the risk and reward of a portfolio. The higher the value, the better the performance of the stock or portfolio. Its expression is shown in Eq. (7):

$$Sharperatio = \frac{E(R_p) - R_f}{\sigma_p} \tag{7}$$

where $E(R_p)$ denotes the payoff rate, R_f denotes the payoff rate of the risk-free investment, and σ_p denotes the standard variance.

Table 1
Dataset resources of stock price forecasting.

Data set introduction	Data Selection Period	Data source	Literature
SSE Composite Index	December 29, 1990 - May 18, 2018	wind database	Liu et al. (2018)
Jakarta composite Index Daily Closing Price Data	January 2014 to December 2014	Yahoo Finance	Agustini et al. (2018)
NIKKEI225 P/E Ratio	February 22, 1985 to December 30, 2008	Undisclosed	Yi and Kita (2012)
AAPL Stock Price Data	May 2008 to May 2018	national association of securities dealers automated quotations	Kulaglic and Ustundaug (2018)
Trading data for 100 stocks on the NDAQ exchange	December 31, 2014 to December 31, 2019	Application Programming Interface (API) service of Yahoo	Zhou et al. (2020)
total Maroc historical data	February 8, 2018 to May 17, 2018	the website of Casablanca stock exchange	Berradi and Lazaar (2019)
Trading data, technical indicators and valuation indicators of CSI 300 Index	January 5, 2015 to December 31, 2019	Guotaian database and wind database	Hu (2021)
Closing price data and related messages for XOM, DELL, EBAY, IBM, KO	July 23, 2012 to July 19, 2013	Yahoo Finance, Yahoo finance message board	Nguyen and Shirai (2015)
S&P data and publicly available financial news	October 2006 to November 2013	reuters and bloomberg	Ding et al. (2014)
Alibaba Stock Data and Related News Information	September 19, 2014 to March 24, 2017	Thomson reuters databasegooseeker software	Zhang et al. (2018)
weakly data of DSE stock	January 1, 2016 to July 31, 2018	Dhaka stock exchange data archive	Shi et al. (2020)
HSIAXnd S&P 500 Index Data	August 23, 1991 to August 23, 2017	Yahoo Finance	Shen et al. (2018)
A number of indicators for Auto-desk (002227)	102 trading days in 2015	communication stock trading software	Xiao et al. (2020)
Shanghai Stock Exchange SSE Composite Index	November 30, 2009 to April 29, 2010	Undisclosed	Wu and Yang (2013)
BSE-30 and Nifty-100 weekly data	April 2011 to December 2016	Yahoo Finance	Musa and Joshua (2020)

Table 2
Stock price prediction platform, website, features and models.

Prediction Platform	Website	Features	Model
Stock price forecasting based on regression analysis	https://urw.cn/K4Knp5	Solve the regression equation to find the coefficients and predict the future price of the stock by using the coefficients.	AR Model
Forecasting stock return time series using ARIMA model	http://m6z.cn/6vpwm1	Applying the ARIMA model to predict stock price returns using the R programming language.	ARIMA Model
Stock price prediction with Stocker tool	https://t.hk.uy/aWUw	Stocker uses an intelligent software package developed by Facebook for additive modeling, creating models in code and making predictions.	Fbprophet Timing Prediction Library
Stock prediction system based on big data analysis	http://m6z.cn/69wHVk	Stock data is crawled and analyzed using a crawler, and the forecast data is presented using a web service.	LSTM Model
Using machine learning to predict stock prices	https://segmentfault.com/a/11900000171 2986	The features are extracted from the wavelet denoised data using Stacked Autoencoders and predicted using LSTM.	LSTM Model
Deep learning LSTM-RNN model to predict stocks	http://45dwz.com/ bqMEg	The advantages of LSTM and RNN models are used for weighted averaging to improve the model prediction accuracy.	LSTM and RNN Model
Web-based stock forecasting system	http://m6z.cn/5xfAsy	Django-based web application using an LSTM model proposed by jaungiers.	LSTM Model
Stock market trend prediction based on stock sentiment analysis	https://www.rainbowsnt.com.cn	Predicting stock upward and downward trends by analyzing the emotional polarity of different Internet users posting stock reviews.	Unsupervised analysis methods

Table 3
Common technical indicators of stock price forecasting.

Indicator Name	Type	Usage
MACD	Trend type	When MACD turns from negative to positive, buy; when MACD turns from positive to negative, sell.
EMV	Trend type	Buy when EMV crosses the 0 axis from bottom to top; sell when EMV crosses the 0 axis from top to bottom.
DMI	Trend type	Buy when +DI crosses up -DI; sell when + DI crosses down -DI.
KDJ	Overbought and oversold type	Buy when the KDJ forms a golden cross; sell when the KDJ forms a dead cross.
RSI	Overbought and oversold type	If the RSI breaks above the previous high platform, buy to rise; if the RSI falls below the previous low line, sell.
ROC	Overbought and oversold type	Buy when ROC breaks the centerline upwards; sell when ROC breaks the centerline downwards.
VR	Energy type	Buy when VR takes the value in the low price area; sell when VR takes the value in the alert area.
OBV	Volume type	Buy when the OBV line rises slowly; sell when the OBV line rises sharply.
BOLL	Path type	Sell when the price crosses the upper pressure line; buy when the price crosses the lower support line.

2.6. Critical technical problem

In the stock price forecasting task, the following three key technical issues are faced.

(1) Stock prices are influenced by various factors. Stock markets are complex and volatile systems, usually influenced by various factors such as national policies, economic conditions, and industry developments. Incomplete and asymmetric information leads to relative difficulties in predicting stock prices accurately.

(2) Stock prices are non-stationary. Stock prices, as time-series data, are non-linear and non-stationary characteristics due to the influence of various aspects and their own reasons. Therefore, the forecasting model should be highly capable of handling non-linear problems.

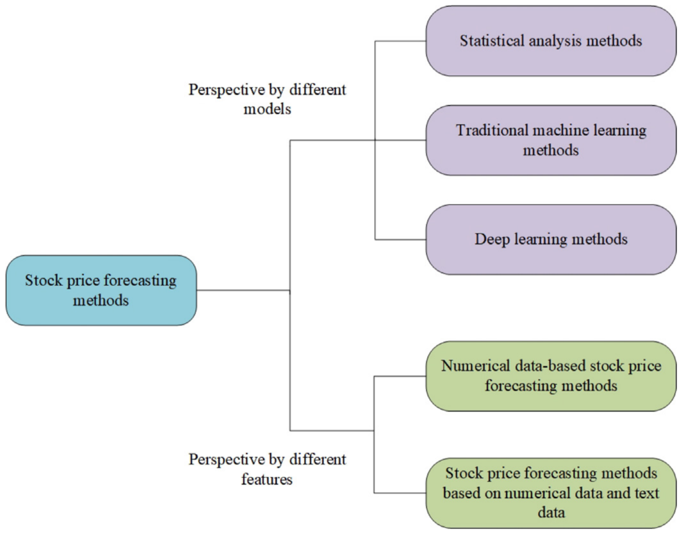

Fig. 1. Stock price forecasting methods under the perspective of different model and feature.

(3) There is randomness in the stock market. Investors are influenced by various aspects of forum speech will produce emotional fluctuations, which have a certain impact on investors' decisions and often lead to stock price fluctuations. Data noise has an impact on the accuracy of stock price prediction.

2.7. Research ideas

According to different research ideas, the stock price forecasting methods under the perspective of different model and feature are shown in Fig. 1.

(1) From the perspective of different models, the stock price forecasting methods can be divided into statistical analysis methods, traditional machine learning based methods and deep learning based methods. We analyze the application of the three types of methods in stock price forecasting and the construction of combined models for forecasting using the advantages of each model.

(2) Studies based on numerical data and studies based on numerical and textual data are classified from the perspective of different features. The stock market is a complex system consisting of multiple elements. The multi-source heterogeneous forecasting technology builds stock price forecasting models that incorporate multiple features. Commentary information contains financial websites, investor decisions and various indicators for forecasting.

3. Stock price forecasting methods from different methodological perspectives

Stock market is uncertain and vulnerable to multiple factors. Researchers first use statistical analysis methods such as ARMA and ARIMA to predict stock price movements. Subsequently, statistical machine learning methods are widely used in stock forecasting research. The rise of deep learning methods such as CNN and RNN are also gradually applied to this task. Representative statistical machine learning stock price forecasting methods for different years are shown in Fig. 2.

3.1. Statistical analysis methods

A large amount of stock trading data is accumulated in the stock market. Time series analysis is a mathematical technique that uses curve fitting, parameter estimation and residual analysis through systematic observations. The purpose is to construct a model and predict future trends (Zhang et al., 2009).

The ARMA model is a unique high precision linear time series forecasting model to estimate smooth irregular fluctuations. Fu and Li (2012) combined with Eviews software to forecast and analyze the closing price of the stock of China Sports Industry. The ARMA model could predict the stock price more accurately and had a greater reference value. The formula of the ARMA model was shown in Eq. (8):

$$x_t = \emptyset_0 + \sum_{i=1}^{p} \emptyset_i x_{t-i} + a_t - \sum_{i=1}^{q} \theta_i a_{t-i} \tag{8}$$

where x_t was the predicted value, and p and q were the orders of autoregression and moving average, respectively.

To address the characteristic of stock prices being nonlinear, autoregressive moving average model (ARIMA) based on ARMA is proposed. ARIMA not only has the advantages of the ARMA model, but also can transform the non-stationary time series into a stationary time series for analysis using the difference method. Majumder et al. (2019) used feedforward neural network, linear model, Holt-Winter method and ARIMA to forecast several Bangladesh stocks. The results showed that ARIMA model had the highest prediction accuracy.

The general ARIMA model ignores the residual equation of non-white noise, and the analysis method based on the variance chi-square of the whole series cannot meet the needs. There is ARCH effect on the residuals, thus introducing the generalized autoregressive conditional heteroskedasticity (GARCH) model, which can predict the non-stationary time series and can better represent the trend of stock data. Zhang et al. (2012) applied a GARCH-like model to the daily closing price of the CSI 300 index and transformed it into a smooth series for predictive analysis of stock prices and returns, where the daily returns were calculated in Eq. (9):

$$r_t = \ln\left(\frac{C_t}{C_{t-1}}\right) \tag{9}$$

where r_t was the return at moment t and C_t was the closing price at moment t, yielding that the method was a better predictor of stock prices in the short run.

3.2. Traditional machine learning based approach

Due to the complexity of the stock market, models based on statistical analysis need to be based on strict assumptions. In the era of rapid explosion of big data, the applicability of statistical analysis models in stock price forecasting is severely reduced (Li, 2012). Machine learning is also used to predict stock prices because of its advantages in handling large amounts of data. It can model non-stationary and non-linear data, and adapt features that make them universally applicable in time series forecasting (Song, 2018).

Studies have shown that the distribution of stock price fluctuations does not follow a normal distribution (Takayasu, 2006). The BN network is fast, easy to train. It can handle digitized values alone, and use makes stock price forecasting possible without a white noise model. Yi and Kita (2012) first converted continuous P/E ratios to digitized values using a clustering algorithm, and then aggregated the samples by the Ward method to predict the stock price return of NIKKEI stock average Toyota Motor Corporation using BN. Among them, the distance metric in Ward's method was formulated as follows:

$$E(C_i) = \sum_{z \in C_i} d(z, c_i)^2 \tag{10}$$

where z was the sample group, c_i was the center of the sample group, and d denoted the distance between them.

To better and more accurately extract the stable trend of non-stationary time series, WT is applied to the analysis of stock history data by virtue of its good processing of non-stationary signals. WT can give a good partial representation of a signal transformed to each other in the time and frequency domains (Huang and Liu, 2019). The use of wavelet transform enables accurate processing of non-stationary time series data. Wang et al. (2017) used a particle swarm optimization algorithm to perform parametric optimization of the wavelet transform.

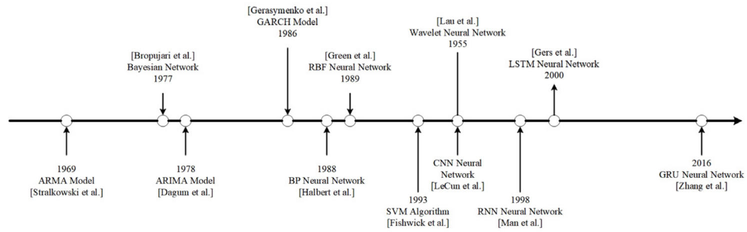

Fig. 2. Timeline of stock price forecasting methods based on statistical machine learning.

They selected appropriate initial network values for different companies and predicting the Taiwan 50 index with an accuracy of 73%.

Support vector machine (SVM) is widely used in data mining such as pattern recognition and regression estimation. SVM was adopted by Jaiwang and Jeatrakul (2016) to build a prediction model of stock buying and selling points for data of five stocks in the Stock Exchange of Thailand. Zhang (2020) used a stock investment model based on a regression fitting method to reduce the dimensionality using principal component analysis (PCA) to calculate the cumulative contribution of eigenvalues. The cumulative contribution was given by the following formula:

$$C = \sum_{k=1}^{i} \lambda_k / \sum_{k=1}^{n} \lambda_k \tag{11}$$

where n was the number of feature vectors and λ_k was the k-th feature root. The performance of SVM was improved by using particle swarm algorithm and genetic algorithm for tuning and optimizing the SVM model. However, it was found that SVM was often difficult to implement when training on large sample sets. Their training time was too long and their stability was poor (Quan et al., 2009).

In order to further reduce the prediction error and improve the accuracy of stock forecasting, a number of researchers have also optimized and improved each forecasting algorithm by making full use of the advantages of multiple models and using combinatorial models. Combinatorial models are used to derive stock forecasts based on the prediction results of a single model weighted according to different weights (Kim et al., 2013). Wang et al. (2021) combined multiple fractal detrended sliding average correlation analysis with mean-semivariance model and the results were better than traditional portfolio models and with lower risk. Ye (2017) used wavelet decomposition and wavelet reconstruction decomposed the daily closing price data of Shanghai Pudong Development Bank into reconstruction part and error part. In addition, Xiao et al. (2020) combined the least squares support vector machine integrated model ARI-MA-LA-LS-SVM for basic stock market prediction, which could improve the prediction accuracy of existing models.

3.3. Deep learning based approach

The development of computing power as well as big data technology has led to continuous breakthroughs in deep learning Monte-Serrat and Cattani (2021). Deep neural network models are widely used as a promising alternative method for time series forecasting. Wu and Yang (2013) optimized BP (backpropagation) neural network by setting the learning steps, number of hidden nodes and activation function for the Shanghai Stock Exchange SSE composite index trend prediction. Zhun and Zhu (2021) optimized the BP neural network model for the nonlinear characteristics of the stock market using the LM (levenberg-marquard) algorithm, where the LM algorithm was formulated in Eq. (12):

$$x_{t+1} = x_t - (J_t^T J_t + \mu I)^{-1} g_t \tag{12}$$

where $J_t^T J_t$ was the Hessian matrix and I was the identity matrix. Then the closing prices of SSE Composite Index and SZSI were predicted to prove the effectiveness of BP neural network in stock prediction.

Radial basis function (RBF) neural networks are commonly used in time series data analysis, identification and prediction of models due to their high prediction accuracy and fast learning rate for nonlinear data. The K-Means clustering algorithm was used by Kan (2019) to optimize the RBF neural network so as to predict the results of eight stock closing prices of APA, TSLA, CCU, TRV, ETR, MCD, PG, and GS to calculate the maximum return. The objective function to maximize the closing 6 return was formulated as follows:

$$\arg\min_{w} - \sum_{i=1}^{n} m_i subjectto \sum w_i = 1, w_i \geq 0 \tag{13}$$

where m was the expected return and w was the weight vector, yielding a higher short-term prediction accuracy of the RBF neural network.

The CNN (convolutional neural network) model is applied to the stock price prediction task. It has better feature extraction and recognition capabilities. The model can be used to study and analyze the stock quotes using k-line charts to classify the reversal points of the stocks. A CNN-based feature extraction framework was proposed by Hoseinzade and Haratizadeh (2019) for predictive analysis of different datasets of US stock market indices, discretizing the data into 0/1 values. They obtained a significant improvement in predictive performance over the latest baseline algorithms. However, in the extraction of multi-dimensional features of stocks using CNN, there is a decrease in accuracy if there are many parameters. High computational complexity and difficulty are needed in optimizing the gradient disappearance (Zhou et al., 2017).

RNN can solve the problem of data with time-series characteristics by achieving self-connection of hidden layers through the horizontal output of neurons. So it can extract time-series information from time-series data and outperforms other types of neural networks in processing time-series data. The integration of PCA and RNN was Berradi and Lazaar (2019) used by to predict the stock price of Maroc in Casablanca Stock Exchange. Their method utilized a gradient descent optimization algorithm and used principal component analysis for dimensionality reduction to improve the accuracy of the RNN model. They had excellent results for predicting the total stock price.

Although RNN works well for short-term memory, its structure is simpler and less effective for long-term memory task operations. The LSTM model better solves the problem of RNN gradient explosion and gradient disappearance by introducing the concept of gate in the model. (Rai et al., 2022) The algorithm performs better than the existing sliding window algorithm, which gives increasingly effective expectations while being more suitable for nonlinear information (Rather, 2021). Hargreaves and Chen (2020) analyzed 400 stocks in the Australian stock market and predicted five of them using LSTM, regression tree and ARIMA models. They found that the LSTM model had better predictive performance. However, the network structure was more complex due to the introduction of more weight parameters in LSTM, which increased the computational complexity.

The gated recurrent unit (GRU) neural network is proposed in 2014. It solves the problem of RNN's difficulty in handling long-distance information capture by introducing a gating structure. Shen et al. (2018) used the GRU neural network and improved it by replacing the Softmax layer of the GRU network with SVM. They pre-

Table 4
Different stock price forecasting methods and forecast effects.

Methods	Model	Literature	Dataset	Results
Statistical analysis methods	ARMA model	Fu and Li (2012)	Closing price data of CTSI shares	Relative Error=1.06%
	ARIMA model	Shi et al. (2020)	Weekly trading data of DSE stock	Accuracy=82.1%
	GARCH model	Zhang et al. (2012)	CSI 300, SSE and SSCI closing prices	Accuracy=94.97%
Traditional machine learning methods	Bayesian networks	Setiani et al. (2021); Yi and Kita (2012)	NIKKEI225 P/E ratio (Yi and Kita, 2012),Bank Rakyat Indonesia Stock Data (Setiani et al., 2021)	RMSE=5.0698 (Yi and Kita, 2012), Euclidean distance=0.0426 (Setiani et al., 2021)
	Wavelet transform	Kulaglic and Ustundaug (2018); Wang et al. (2017)	Apple Stock Price Data (Kulaglic and Ustundaug, 2018),TWSE50 Index Data (Wang et al., 2017)	RMSE=3.51 (Kulaglic and Ustundaug, 2018),Accuracy =73.00% (Wang et al., 2017)
	SVM algorithm	Jaiwang and Jeatrakul (2016); Zhang (2020)	Stock data of the Stock Exchange of Thailand (Jaiwang and Jeatrakul, 2016),Shanghai and Shenzhen index data and 24 individual stocks trading data and financial news (Zhang, 2020)	Accuracy =72.45% (Jaiwang and Jeatrakul, 2016),Accuracy =84.21% (Zhang, 2020)
	ARIMA-SVR model	Ye (2017)	Shanghai Pudong Development Bank Daily Closing Price Data	MSE=0.57
	ARI-MA -LS-SVM model	Xiao et al. (2020)	Auto-desk (002227)'s various metrics data	RMSE=0.108
Deep learning methods	BP Neural network	Wu and Yang (2013); Zhun and Zhu (2021)	SSE Composite Index Data (Wu and Yang, 2013), CSI 300 Index Data (Zhun and Zhu, 2021)	Accuracy =94.00% (Wu and Yang, 2013), Accuracy =98.69% (Zhun and Zhu, 2021)
	RBF Neural network	Shi and Gao (2015)	CSI 300 Index Data	MAE=18.6743
	CNN model	Chen and Huang (2021); Hoseinzade and Haratizadeh (2019); Zhou et al. (2020)	S&P 500 Index Data (Hoseinzade and Haratizadeh, 2019), NDAQ Stock Trading Data (Zhou et al., 2020), S&P 500 Index (Chen and Huang, 2021)	F-measure=0.55 (Hoseinzade and Haratizadeh, 2019), Accuracy=60% (Zhou et al., 2020), Accuracy=67% (Chen and Huang, 2021)
	RNN model	Berradi and Lazaar (2019); Pahlawan et al. (2021); Zhu (2020)	XYZ Stock Price Data (Pahlawan et al., 2021), APPLE Stock Data (Zhu, 2020), Total Maroc Historical Data (Berradi and Lazaar, 2019)	MAPE=1.558% (Pahlawan et al., 2021), Accuracy=95% (Zhu, 2020), MSE=0.00596 (Berradi and Lazaar, 2019)
	LSTM neural network	Hargreaves and Chen (2020); Hu (2021)	Australian Stock Market Data (Hargreaves and Chen, 2020), CSI 300 Index Data (Hu, 2021)	Sharpe Ratio=2.13 (Hargreaves and Chen, 2020), RMSE=0.1420 (Hu, 2021)

dicted the trading signals of HSI index, German DAX index and S&P500 index with high prediction accuracy.

3.4. Performance comparison of different stock price forecasting methods

The time complexity of representative methods of stock price prediction under different methodological perspectives and the prediction results in stock historical data are compared in Table 4. As can be seen from Table 4, the accuracy of stock price prediction models has been increasing with the continuous improvement of machine learning methods and deep learning models. First, the deep learning-based models can automatically extract the features and rules implied by the samples and then find certain behavioral change patterns. So the stock price prediction methods are powerful in sequence modeling compared with statistical analysis and traditional machine learning. This makes the stock price prediction more intelligent and with better prediction ability. Second, the combined model fully combines the advantages of each algorithm to overcome the phenomena such as overfitting that occurs in single models. Its prediction accuracy is relatively higher than that of a single model. For example, in the literature (Xiao et al., 2020), a combined ARI-MA-LS-SVM model was used to predict several index data of Autodesk (002227). The RMSE prediction index value was obtained only 0.108. Therefore, deep learning models as well as combinatorial models are also important research methods to solve the stock price prediction problem at present. In terms of model time complexity, the Bayesian network ($O(n2^n)$) has a significantly higher time complexity than the wavelet transform ($O(n \log n)$) and the SVM algorithm ($O(n^2)$). The number of hidden layer neurons of the RBF network is much higher than that of the BP network when the training samples are increased. This makes the complexity of the RBF network increase greatly. Both CNN ($O(knd^2)$) and RNN ($O(nd^2)$) are higher than that of the BP neural network ($O(n^2)$). The LSTM model is an improvement of the RNN. It can avoid the problem of gradient disappearance and have longer memory. Its time complexity is higher than that of CNN and RNN.

4. Stock price forecasting methods from different features perspectives

4.1. Numerical data-based stock price forecasting methods

The stock price forecasting method based on numerical data is a more applied forecasting method in existing studies. It is mainly a single application of historical data of a stock for a certain period of time, such as stock market trading data of a stock for a certain period of time, to predict the price of that stock for a future period of time. There has been a large amount of research work on stock price prediction methods based on numerical data. Techniques based on machine learning, deep learning, and data mining are used in stock price prediction based on numerical data.

An eight-factor stock selection model index system based on Jian Jiao's six-factor model was constructed by Wang et al. (2016). They used random forest model (RF) to predict the rise and fall of 200 stocks, with an average prediction accuracy of 75.50%. Cui et al. (2019) used Baum-Welch algorithm and Python's Hmmlearn package to complete the training of the hidden markov model (HMM), and adopted multiple frequency features to predict the Shanghai-Shenzhen index data. RNN considered the data before and after the time period when processing the time series. Zhu (2020) normalized the price data of Apple stock in the past ten years and scaled the stock data to [0, 1] range. Then RNN neural networks were used to predict them, and the prediction accuracy reached more than 95%.

GRG nonlinear method and evolutionary method based on genetic algorithm was by adopted by Mishra et al. (2021). ARIMA and human artificial neural network model were used by Musa and Joshua (2020) to develop a hybrid model for forecasting the Nigerian Asian stock market all-share index. The results showed that the hybrid model using artificial neural network had better forecasting effect than the single differential autoregressive model or neural network model.

Table 5
Different stock price forecasting methods and forecast effects.

Methods	Model	Literature	Dataset	Results
Prediction methods based on numerical data	RF model	Ai (2020)	Computer industry 200 stocks data	Accuracy=60.43%
	HMM model	Cui et al. (2019)	CSI 300 Index Data	MAPE=0.7471
	RNN model	Zhu (2020)	Apple's stock price	Accuracy =95%
	LASSO-TLBO-SVR model	Mishra et al. (2021)	Weekly data of BSE-30 and Nifty-100	MAPE<5
Methods based on numerical data and text data	TSLDA model	Nguyen and Shirai (2015)	Closing price data and related messages for XOM, DELL, EBAY, IBM, KO	Accuracy =56.43%
	Multi-layer perceptron algorithm	Yu et al. (2019)	Stock trading data indicators and text data	Accuracy =65.91%
	LSTM neural Networks	Zhou (2018)	CSI 300 Index data and related text data	Accuracy =78.1%
	SVR algorithm	Schumaker and Chen (2009)	S&P 500 data and related financial news	MSE=0.1954

4.2. Stock price prediction methods based on numerical and textual data

Text data from socio-political events and financial news in the fintech sector to investors' comments on the investment behavior of a stock. To some extent there is an impact on the volatility of the stock market. Therefore text data mining techniques are used in stock price forecasting to assist in decision making. At present, financial text sentiment research mainly includes two categories of investor sentiment and news text sentiment information. By classifying the sentiment of stock-related news financial and other textual information, it can help users to make investment judgments about the positive and negative impacts of stocks (Gao et al., 2010). There are many sentiment classification methods, for example, Demircan et al. (2021) built a sentiment analysis model using support vector machine (SVM), random forest (RF), decision tree (DT), logistic regression (LR) and k-nearest neighbor (KNN) for Turkish language sentiment classification, and the results show that support vector machine and RF are a feasible method. Haque et al. (2023) proposed a supervised deep learning classifier based on cnn and lstm for sentiment classification on Bengali social media comments. The method proposed by Jindal et al. (2020); Kaur et al. (2021); Zhao et al. (2014, 2019, 2020, 2021) also helps us to accomplish the financial text sentiment classification task.

Vega (2006) showed experimentally that the more information investors obtained about the true value of an asset, the more they subscribed to that information. A model that used social media to capture themes was developed by Nguyen and Shirai (2015). They used their sentiment model to predict stock closing price movements with higher accuracy than using historical data alone. A new text fusion model using K-means clustering was proposed by Yu et al. (2019). The text categories were generated using a K-mean clustering method. A multi-layer perceptron algorithm was used to combine 15 original price indicators such as opening and closing prices to predict the price of the market. The accuracy of the prediction was 65.91%. Zhou (2018) mined and analyzed social media texts with sentiment and used PMI index. They calculated sentiment tendency and sentiment tendency using PMI index with Eq. (14):

$$S = \frac{\sum_{i=1}^{n} PMI_{posi} - \sum_{i=1}^{n} PMI_{negi}}{n} \tag{14}$$

where $posi$ was a positive term and $negi$ was a negative term. Ding et al. (2014) used open information extraction techniques to extract structured events from large-scale news. In order to solve the problem of news events of not capturing structured entity relationship information for news events, they obtained high prediction accuracy for S&P index data. Zhang et al. (2018) transformed Alibaba media news into structured news sentiment by dictionary matching based on L&M dictionary. Dong et al. (2017) used both sentiment dictionaries and support vector machines for sentiment analysis from the perspective of rule-based and statistical methods to obtain sentiment information.

4.3. Performance comparison of different stock price forecasting methods

The prediction results of representative methods of stock price forecasting under the perspective of different features on historical stock data were compared in Table 5. From Table 5, it can be seen that both numerical and textual data-based stock price prediction methods achieved good accuracies. The stock price prediction methods based on numerical data and text data fully consider the influence of financial news and investors' speech decisions on the stock market. These methods quantify them by conducting sentiment analysis. These methods explore the influence of text data on the stock market, and conduct feature fusion analysis with historical stock trading data to get a more comprehensive grasp of the stock market fluctuation trends. (Zhou, 2018) incorporates the CSI 300 index trading data and related stock bar post text data. The LSTM model was used for prediction with an accuracy of 78.1%. The method of stock price prediction based on numerical data and text data is a more cutting-edge method and one of the important research directions in the future.

5. Research challenges and future work

In summary, there is a wide variety of stock prediction models that have achieved good results. However, there are still many problems that have not yet been fully solved. The possible future research challenges and perspectives are mainly in the following four areas.

(1) Most studies have used a single algorithm to construct models for forecasting, while studies that combine statistical machine learning models with traditional stock analysis methods (fundamental analysis and technical indicator analysis) are lacking. The forecasting process often ignores the intrinsic value of stocks and the extrinsic influence of market factors. In the future, we can make use of the advantages of various algorithms, adjust the model characteristics, optimize the weighted average of the model so as to build a combination model for forecasting. We should make full use of factor analysis and other methods for the selection of stock-related indicator characteristics. We can also use stock market images, which can fully capture the laws of different stock markets and explore the stock market more deeply.

(2) Most current stock price forecasting methods simply use historical stock trading data for analysis and forecasting, without taking into account textual information such as financial news, company earnings reports and stock bar comments that may have an impact on investors. Therefore, the impact of textual information on the stock market should be fully considered. Deep learning should be applied to stock price prediction. We quantify the textual data such as investors' speech decisions and financial news, combine investors' attention. Also we build models to fuse them with historical stock trading data, and analyze and forecast using multiple information sources.

(3) Most models predict stock prices and their trends on a monthly or annual cycle with long prediction periods. They are less likely to be able to predict daily stock prices and their rise and fall more accurately, making it difficult to give guiding advice in actual operations. In future

research, it is possible to set up time steps, select different time steps, apply intelligent methods or pattern recognition to capture the stock's anomalies, conduct quantitative analysis through real-time detection. It is able to evaluate the expected returns, discover the quantitative characteristics of stock trends with large fluctuations, and effectively give buy and sell signals among non-smooth stocks according to the stock price returns and the magnitude of increases and decreases.

(4) The current forecasting models can basically satisfy the prediction of future stock prices. But most of the models make a general prediction of the future trend of stocks with relatively low accuracy. Improving the prediction accuracy is still a concern. The impact of various factors on the model and the stock market, such as the trading situation of stocks, should be considered comprehensively. The model is used to study multiple types of stock markets, while optimizing the model construction in terms of parameter selection settings. The establishment of a universally applicable model enables stock price forecasting models with high accuracy while providing an indicator of the credibility of the forecast results.

6. Conclusion

This paper reviews the background knowledge related to stock price forecasting, as well as existing stock price forecasting methods and applications. We classify and summarize the existing work on stock price forecasting from two perspectives, according to different models and different features. In particular, the datasets, literature sources and performance used in different stock price forecasting methods and forecasting methods from the perspective of different features are compared. Deep learning-based models achieve better results. Combinatorial models are an important current research approach. Stock price forecasting methods based on numerical and textual data are cutting-edge research directions. Finally, the remaining challenges of stock price prediction are summarized and its future research directions are outlined.

Declaration of Competing Interest

There is no conflict of interest.

Acknowledgments

This study is supported by national natural science foundation of China (Grant No.61906110), humanities and social sciences project of the ministry of education (Grant No.22YJAZH092), outstanding innovation project for graduate students in Shanxi province (Grant No.2022Y535).

References

Agustini, W. F., Affianti, I. R., & Putri, E. R. (2018). Stock price prediction using geometric Brownian motion. In *Journal of Physics: Conference Series: vol. 974* (p. 012047). IOP Publishing.

Ai, Y. (2020). Research on trend prediction of CSI 300 index based on random forest optimization. Master's Thesis.Shandong University.

Bao, Z., Guo, J., Xie, Y., & Zhang, W. (2020). Model for stock price trend prediction based on LSTM and GA. *Comput Sci, 47*, 467–473.

Berradi, Z., & Lazaar, M. (2019). Integration of principal component analysis and recurrent neural network to forecast the stock price of casablanca stock exchange. *Procedia Computer Science, 148*, 55–61.

Bogle, S., & Potter, W. (2015). Sentamal-a sentiment analysis machine learning stock predictive model. In *Proceedings on the international conference on artificial intelligence (ICAI)* (pp. 610–615). The Steering Committee of The World Congress in Computer Science.

Chen, M.-Y., Sangaiah, A. K., Chen, T.-H., Lughofer, E. D., & Egrioglu, E. (2022). Deep learning for financial engineering. *Computational Economics, 59*(4), 1277–1281.

Chen, Y.-C., & Huang, W.-C. (2021). Constructing a stock-price forecast cnn model with gold and crude oil indicators. *Applied Soft Computing, 112*, 107760.

Cui, X., Shang, W., Jiang, F., & Wang, S. (2019). Stock index forecasting by hidden markov models with trends recognition. In *2019 IEEE international conference on big data* (pp. 5292–5297). IEEE.

Demircan, M., Seller, A., Abut, F., & Akay, M. F. (2021). Developing turkish sentiment analysis models using machine learning and e-commerce data. *International Journal of Cognitive Computing in Engineering, 2*, 202–207.

Ding, X., Zhang, Y., Liu, T., & Duan, J. (2014). Using structured events to predict stock price movement: An empirical investigation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1415–1425).

Dong, L., Wang, Z., & Xiong, D. (2017). Stock index prediction based on text information. *Acta entiarum Naturalium Universitatis Pekinensis, 53*(2), 273–278.

Fu, Y., & Li, F. (2012). Application of ARMA model in sports stock price prediction in China. *Statistics and Decision, 21*(3), 101–103.

Gao, Y., Zhou, L., Zhang, Y., Xing, C., Sun, Y., & Zhu, X. (2010). Sentiment classification for stock news. In *5th international conference on pervasive computing and applications* (pp. 99–104). IEEE.

Grimm, R. C. (2012). Fundamental analysis as a traditional austrian approach to common stock selection. *Quarterly Journal of Austrian Economics, 15*(2), 221–237.

Haque, R., Islam, N., Tasneem, M., & Das, A. K. (2023). Multi-class sentiment classification on bengali social media comments using machine learning. *International Journal of Cognitive Computing in Engineering.*

Hargreaves, C., & Chen, L. (2020). Stock prediction using deep learning with long-short--term-memory networks. *International Journal of Electronic Engineering and Computer Science, 5*(3), 22–32.

Hoseinzade, E., & Haratizadeh, S. (2019). CNNpred: CNN-based stock market prediction using a diverse set of variables. *Expert Systems with Applications, 129*, 273–285.

Hu, Y. (2021). Stock forecast based on optimized LSTM model. *Computer Science, 48*(6A), 151–157.

Huang, S., & Liu, S. (2019). Machine learning on stock price movement forecast: the sample of the taiwan stock exchange. *International Journal of Economics and Financial Issues, 9*(2), 189–201.

Jain, G., Mahara, T., Sharma, S., & Sangaiah, A. K. (2022). A cognitive similarity-based measure to enhance the performance of collaborative filtering-based recommendation system. *IEEE Transactions on Computational Social Systems, 9*(6), 1785–1793.

Jaiwang, G., & Jeatrakul, P. (2016). A forecast model for stock trading using support vector machine. In *2016 international computer science and engineering conference (ICSEC)* (pp. 1–6). IEEE.

Jindal, R., Seeja, K., & Jain, S. (2020). Construction of domain ontology utilizing formal concept analysis and social media analytics. *International Journal of Cognitive Computing in Engineering, 1*, 62–69.

Jing, W., Sangaiah, A. K., Wei, L., Shaopeng, L., Lei, L., & Ruishi, L. (2022). Multi-view fusion for recommendation with attentive deep neural network. *Evolutionary Intelligence, 15*(4), 2619–2629.

Kan, Z. (2019). Research on stock forecasting based on improved machine learning method. Master's thesis.Changchun: Changchun University of Science and Technology.

Kaur, P., Kumar, H., & Kaushal, S. (2021). Affective state and learning environment based analysis of students performance in online assessment. *International Journal of Cognitive Computing in Engineering, 2*, 12–20.

Kim, H., Jung, J., & Kim, S. (2013). Investment performance of markowitz's portfolio selection model over the accuracy of the input parameters in the korean stock market. *Journal of the Korean Operations Research and Management Science Society, 38*(4), 35–52.

Kulaglic, A., & Ustundaug, B. B. (2018). Stock price forecast using wavelet transformations in multiple time windows and neural networks. In *2018 3rd international conference on computer science and engineering (UBMK)* (pp. 518–521). IEEE.

Li, L. (2012). Effectiveness analysis of chinese stock market based on statistical model. *Shanxi Finance and Economics University, 34*(S1), 70–71.

Liu, F., Hu, X., & Li, F. (2018). Research on the validity of china's stock market based on random walk test. In *Third international conference on economic and business management (FEBM 2018)* (pp. 171–174). Atlantis Press.

Majumder, M. M. R., Hossain, M. I., & Hasan, M. K. (2019). Indices prediction of bangladeshi stock by using time series forecasting and performance analysis. In *2019 international conference on electrical, computer and communication engineering (ECCE)* (pp. 1–5). IEEE.

Mishra, S., Padhy, S., Mishra, S. N., & Misra, S. N. (2021). A novel lasso-tlbo-svr hybrid model for an efficient portfolio construction. *The North American Journal of Economics and Finance, 55*, 101350.

Monte-Serrat, D. M., & Cattani, C. (2021). Interpretability in neural networks towards universal consistency. *International Journal of Cognitive Computing in Engineering, 2*, 30–39.

Musa, Y., & Joshua, S. (2020). Analysis of arima-artificial neural network hybrid model in forecasting of stock market returns. *Asian Journal of Probability and Statistics, 6*(2), 42–53.

Nguyen, T. H., & Shirai, K. (2015). Topic modeling based sentiment analysis on social media for stock market prediction. In *Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (volume 1: Long papers)* (pp. 1354–1364).

Pahlawan, M. R., Riksakomara, E., Tyasnurita, R., Muklason, A., Mahananto, F., & Vinarti, R. A. (2021). Stock price forecast of macro-economic factor using recurrent neural network. *IAES International Journal of Artificial Intelligence, 10*(1), 74–83.

Pan, J., & Ding, M. (2000). Gp distribution model and stock return analysis. *Journal of Peking University Science & Technology Edition, 36*(3), 295–306.

Quan, L., Jiang, X., Zhao, J., & Wang, D. (2009). Research of selecting stocks based on classification method of SVM. *Journal of Shanghai Jiaotong University, 43*(9), 1412–1416.

Rai, N., Kumar, D., Kaushik, N., Raj, C., & Ali, A. (2022). Fake news classification using transformer based enhanced LSTM and bert. *International Journal of Cognitive Computing in Engineering, 3*, 98–105.

Rather, A. M. (2021). LSTM-based deep learning model for stock prediction and predictive optimization model. *EURO Journal on Decision Processes, 9*, 100001.

Sahoo, P., & Charlapally, K. (2015). Stock price prediction using regression analysis. *International Journal of Scientific & Engineering Research, 6*(3), 1655–1659.

Schumaker, R. P., & Chen, H. (2009). A quantitative stock prediction system based on financial news. *Information Processing & Management, 45*(5), 571–583.

Setiani, I., Tentua, M. N., & Oyama, S. (2021). Prediction of banking stock prices using naive bayes method. In *Journal of physics: Conference series: vol. 1823* (p. 012059). IOP Publishing.

Shen, G., Tan, Q., Zhang, H., Zeng, P., & Xu, J. (2018). Deep learning with gated recurrent unit networks for financial sequence predictions. *Procedia computer science, 131*(C), 895–903.

Shi, J., Liu, W., Feng, Z., Zhou, Y., et al., (2020). Analysis and forecast of stock price law based on Arima model. *Statistics and Application, 9*(1), 101–114.

Shi, W., & Gao, Y. (2015). Prediction of the chinese stock index futures market based on new EMD-RBF model. *Journal of Quantitative Economics, 32*(1), 47–51.

Song, G., Zhang, Y., Bao, F., & Qin, C. (2019). Stock prediction model based on particle swarm optimization LSTM. *Journal of Beijing University of Aeronautics and Astronautics, 45*(12), 2533–2542.

Song, Y. (2018). *Stock trend prediction: Based on machine learning methods*. UCLA Ph.D. thesis..

Sumathi, A., Javadpour, A., Pinto, P., Sangaiah, A. K., Zhang, W., & Mahmoodi Khaniabadi, S. (2022). Newtr: A multipath routing for next hop destination in internet of things with artificial recurrent neural network (rnn). *International Journal of Machine Learning and Cybernetics, 13*(10), 2869–2889.

Takayasu, H. (2006). *Practical fruits of econophysics*. Springer.

Tang, H., & Sun, Z. (2003). A cointegration analysis among four exchange rates and four stock indices. *Acta Scicentiarum Naturalum Universitis Pekinesis, 39*(5), 601–606.

Vega, C. (2006). Stock price reaction to public and private information. *Journal of Financial Economics, 82*(1), 103–133.

Wang, F., Ye, X., Chen, H., & Wu, C. (2021). A portfolio strategy of stock market based on mean-mf-x-dma model. *Chaos, Solitons & Fractals, 143*, 110645.

Wang, K., Yang, C., & Chang, K. (2017). Stock prices forecasting based on wavelet neural networks with pso. In *Matec web of conferences: vol. 119* (p. 01029). EDP Sciences.

Wang, S., Cao, Z.-f., & Chen, M. (2016). Research on application of random forests in the quantitative stock selection model. *Operations Research and Management Science, 25*(3), 163.

Wang, Z., & Wang, S. (2019). Fundamental price effect analysis: Evidence from chinese a-shares. *Research on Financial and Economic Issues, 4*, 53–61.

Wu, M., & Yang, Y. (2013). The research on stock price forecast model based on data mining of bp neural networks. In *2013 third international conference on intelligent system design and engineering applications* (pp. 1526–1529). IEEE.

Xiao, C., Xia, W., & Jiang, J. (2020). Stock price forecast based on combined model of ARI-MA-LS-SVM. *Neural Computing and Applications, 32*(10), 5379–5388.

Xie, X., Xie, J., Ren, J., & Yu, S. (2009). Dynamic decision making under approaching real risk scenarios. *Journal of Peking University (Natural Science Edition), 45*(5), 884–890.

Ye, T. (2017). Stock forecasting method based on wavelet analysis and arima-svr model. In *2017 3rd international conference on information management (ICIM)* (pp. 102–106). IEEE.

Yi, Z., & Kita, E. (2012). Stock price forecast using bayesian network. *Expert Systems with Applications, 39*(8), 6729–6737.

Yu, C., Gong, Y., Wang, F., & An, L. (2019). Predicting stock prices with text and price combined model. *Data Analysis and Knowledge Discovery, 2*(12), 33–42.

Zhang, C., Ding, J., Zhan, J., & Li, D. (2022). Incomplete three-way multi-attribute group decision making based on adjustable multigranulation pythagorean fuzzy probabilistic rough sets. *International Journal of Approximate Reasoning, 147*, 40–59.

Zhang, C., Li, D., & Liang, J. (2020). Multi-granularity three-way decisions with adjustable hesitant fuzzy linguistic multigranulation decision-theoretic rough sets over two universes. *Information Sciences, 507*, 665–683.

Zhang, C., Li, D., Liang, J., & Wang, B. (2021). Magdm-oriented dual hesitant fuzzy multigranulation probabilistic models based on multimoora. *International Journal of Machine Learning and Cybernetics, 12*, 1219–1241.

Zhang, J., Shan, R., & Su, W. (2009). Applying time series analysis builds stock price forecast model. *Modern Applied Science, 3*(5), 152–157.

Zhang, L. (2020). Research on stock market trend analysis and prediction based on svm. Master's thesis.Nanjing: Nanjing University of Posts and Telecommunications.

Zhang, L., Fu, S., & Li, B. (2018). Research on stock price forecast based on news sentiment analysis-a case study of alibaba. In *International Conference on Computational Science* (pp. 429–442). Springer.

Zhang, L., Luo, Y., & Tang, Y. (2012). Analysis of the chinese stock market returns based on the garch models. *Journal of Sichuan University (Natural Science Edition), 49*(1), 15–22.

Zhao, C., Liu, J., Wu, M., & Chang, X. (2022). Research on stock price prediction based on lstm with numerical and text features. *Journal of Shanxi University Natural Science Edition*.

Zhao, C., Wang, S., & Li, D. (2014). Fuzzy sentiment membership determining for sentiment classification. In *2014 ieee international conference on data mining workshop* (pp. 1191–1198). IEEE.

Zhao, C., Wang, S., & Li, D. (2019). Exploiting social and local contexts propagation for inducing Chinese microblog-specific sentiment lexicons. *Computer Speech & Language, 55*, 57–81.

Zhao, C., Wang, S., & Li, D. (2020). Multi-source domain adaptation with joint learning for cross-domain sentiment classification. *Knowledge-based systems, 191*, 105254.

Zhao, C., Wang, S., Li, D., Liu, X., Yang, X., & Liu, J. (2021). Cross-domain sentiment classification via parameter transferring and attention sharing mechanism. *Information sciences, 578*, 281–296.

Zhou, F., Jin, L., & Dong, J. (2017). Review of convolutional neural network. *Chinese Journal of Computers, 40*(6), 1229–1251.

Zhou, L. (2018). Research on stock market forecast based on LSTM and investor sentiment. Master's thesis.Wuhan: Central China Normal University.

Zhou, Q., Liu, N., & Shen, L. (2020). A stock price prediction model based on stock charts and deep cnn. *Finance, 10*(4), 334–342.

Zhu, Y. (2020). Stock price prediction using the RNN model. *Journal of Physics: Conference Series, 1650*(3). 032103 (7pp)

Zhun, J., & Zhu, P. (2021). Prediction of stock index based on bp neural network. In *The 2nd international conference on education technology, economic management and social sciences (ETEMSS 2021)* (pp. 192–197). ETEMSS.