Theory and Practice of Science and Technology
2025, VOL. 6, NO. 8, 117-123
DOI: 10.47297/taposatWSP2633-456922.20250608

A Review of Stock Price Prediction Research Based on Artificial Intelligence and Big Data Technology

Ziqi Li

School of Economics and Management, University of Science and Technology Beijing, Beijing, 100083, China

ABSTRACT

Stock price prediction in the financial sector has long been a hot topic of research. In recent years, with the continuous advancement of artificial intelligence and big data technologies, stock price prediction has gradually shifted from traditional statistical modeling paradigms toward an intelligent direction driven by multi-source heterogeneous data and dominated by deep learning. This study systematically reviews the recent research progress in the application of artificial intelligence and big data technologies to stock price prediction. It analyzes the field from multiple dimensions, including feature construction and fusion strategies, the evolution of modeling methods, and the application of big data processing frameworks. The study summarizes the current mainstream research paths and typical achievements, and discusses the challenges and constraints faced by various methods in practical applications. Additionally, this study summarizes the existing research findings and conducts an in-depth analysis of the limitations in the research.

KEYWORDS

Stock price prediction; Artificial intelligence; Big data technology, Multi-source heterogeneous data; Feature fusion

1 Background

As an important barometer of economic activity, stock market trends have long been a hot topic of research in the financial sector. Due to the highly nonlinear, volatile, and unpredictable nature of stock prices, traditional statistical modeling methods often encounter limitations when dealing with complex market behavior, making it difficult to effectively capture the dynamic characteristics of market changes. Additionally, as demands for higher precision and accuracy in predicting financial market volatility patterns have grown, traditional single-domain data statistical analysis has become insufficient to meet these needs.

With the continuous evolution of artificial intelligence technology, particularly machine learning and deep learning, their application in stock price prediction has become increasingly widespread. These technologies significantly outperform traditional methods in feature extraction, nonlinear modeling, and multi-source data fusion, thereby effectively enhancing the accuracy and adaptability of predictive models. In the meantime, The development of big data technology has provided robust data support for stock prediction, enabling efficient processing and input of multi-source heterogeneous data from trading markets, news texts, social media, and financial statements into models, thereby enhancing the comprehensiveness, dynamism, and real-time nature of predictions.

This study reviews the research on artificial intelligence and big data technology in the field of stock prediction, integrating and summarizing existing cutting-edge literature. On one hand, it systematically summarizes the progress and characteristics of current research in model construction, data processing, and prediction performance. On the other hand, it helps identify the challenges and shortcomings in current research, clarifies future development trends, and provides theoretical support and methodological references for subsequent research work and actual financial investment decisions.

2 Feature selection and construction in stock prediction

In stock price forecasting, stock price fluctuations are influenced by numerous factors, including time-series indicators, fundamental indicators, technical indicators, and even investor sentiment and news sentiment. Therefore, in-depth analysis of these characteristics is a critical factor in constructing stock models. Utilizing as many multi-source indicators and constructing features as possible can improve the accuracy of stock price forecasting and enhance the interpretability of stock price fluctuations.

Luo Yufeng (2013) [1] discretized historical stock prices into different price intervals, constructed a Markov transition matrix, and used the chi-square test based on actual case analysis to prove that it met the Markov test and was consistent with the actual steady state, thereby completing the prediction of stock prices. Ariyo et al. (2014) [2] also used historical stock prices as a single indicator, employing the ARIMA time series model to predict stock prices and uncover the intrinsic relationships within the time series, ultimately confirming the effectiveness of statistical models in short-term stock price forecasting. With the explosive growth of financial market data dimensions, multi-source heterogeneous feature fusion integrates cross-domain information such as technical indicators, text sentiment analysis, and macroeconomic data, significantly enhancing the comprehensiveness and dynamic adaptability of predictive models. Zhang Kai (2017) [3] used a bidirectional LSTM (BiLSTM) to analyze the sentiment of stock forum comments and found that sentiment polarity has predictive power for key turning points in stock prices. Tsai (2019) [4] proposed a multi-factor fuzzy time series (FTS) model that integrates technical indicators, stock indices, trading volume, and the daily differences between two stock market indices as key factors, maintaining high predictive accuracy even in complex market environments. Ayala et al. (2021) [5] employed TEMA and MACD trading strategies combined with moving averages and exponential averages to derive technical analysis indicators. By integrating machine learning for prediction, this approach constitutes an efficient and robust stock trading strategy, particularly effective in markets with clear long-term trends. Zhang et al. (2023) [6] used a fixed-effects model to demonstrate that policy uncertainty effects (EPU) indirectly stimulate stock liquidity by igniting retail investors' enthusiasm for investment, attracting speculative trading by stock investors, and thereby influencing stock price volatility. Multiple studies have shown that multi-source heterogeneous feature fusion enhances the predictive capabilities of stock models by integrating cross-domain data such as technical indicators, text sentiment analysis, and macroeconomic data, driving the evolution of predictive models from static fitting to dynamic adaptation. On one hand, multi-source data is gradually achieving a comprehensive portrayal of market-driving factors, thereby improving the model's ability to explain stock prices. On the other hand, the fusion of multi-source heterogeneous features, combined with deep prediction models, hybrid models, and reinforcement learning models, can better achieve dynamic weight allocation, enabling models to dynamically capture market trends. This not only enhances predictive capability but also improves real-time performance and generalization ability. However, as data dimensions increase and data volumes grow, challenges arise, including multi-modal semantic alignment, dynamic weight adaptation, and the collection and filtering of massive amounts of data and noise.

3 Stock prediction model construction

3.1 Machine learning

Machine learning algorithms, with their nonlinear modeling capabilities, have broken through the limitations of traditional statistical models and become an important tool in the field of stock prediction. From the early days of artificial neural networks (ANN) fitting nonlinear relationships in stock prices, to algorithms like KNN and random forests (RF) mining trading patterns, researchers have validated the effectiveness of machine learning in scenarios such as trend prediction and risk assessment through feature engineering and algorithm optimization, laying the methodological foundation for the introduction of deep learning technologies.

Haider (2011) [7] used artificial neural networks (ANN) for backpropagation training with different input indicators, resulting in an average error rate of 3.71% on GI and EPS, with a 2.18% reduction in error on GI, P/E, NAV, EPS, and Volume. Alkhatib (2013) [8] used the KNN algorithm to calculate the similarity between test samples and training samples using Euclidean distance. Through cross-validation, the nearest neighbor k was determined to be 5, and the weighted average of the closing prices of the neighbors was taken as the predicted value. Lin Nana (2018) [9] constructed a Random Forest model and a trading indicator system to predict the rise and fall of the A-share market, verifying the effectiveness of machine learning in financial time series data modeling. The overall accuracy rate achieved was 83%. Lü Linfeng (2019) [10] used an SVM model and tested its annualized return under different market conditions. The annualized return of the SVM prediction account was 16.89% higher than that of the original portfolio, demonstrating high annualized returns during periods of sharp rises and falls and resilience during slow bear markets, thereby confirming the model's strong generalization ability. Machine learning algorithms, with their powerful nonlinear modeling capabilities, have broken free from the reliance on linear relationships inherent in traditional statistical methods, significantly improving the accuracy of predictions. Nevertheless, single machine learning models remain limited, and many studies in this field have begun to explore ensemble learning methods that integrate multiple model algorithms, combining the unique advantages of each model. However, most current methods still face challenges such as high market noise and strong heterogeneity. Nonetheless, they provide a solid technical foundation and validation basis for the evolution of deep learning methods. In the future, it is anticipated that ensemble learning combined with multi-source data fusion will further enhance model robustness and practical applicability.

3.2 Deep learning

Deep learning achieves automatic extraction and mining of complex time-series features in stock markets through hierarchical representations of multi-layer neural networks, making it a current research hotspot in predictive modeling. Convolutional neural networks (CNNs) capture spatial features, recurrent neural networks (RNNs/LSTMs) model time-series dependencies, and attention mechanisms focus on key information, giving the model significant advantages in multi-source data fusion and long-term forecasting.

Zeng An (2019) [11] used an LSTM model to predict stock price movements and found that it outperformed traditional methods in short-term predictions, but its performance declined significantly as the prediction period lengthened. Zhu (2020) [12] utilized the contextual relationships in time series data, employing a two-layer RNN structure with 50 and 100 neurons respectively, and used data from the previous 5 to 10 days to predict the stock price for the current day. When predicting Apple Inc.'s stock price, the Mean Absolute Error (MAE) was as low as 5.9. Moodi (2024) [13] proposed a CNN-LSTM hybrid model combined with text sentiment analysis, using stock datasets from Tesla (TSLA) and Apple (AAPL), achieving an accuracy rate of 96.83%, outperforming LSTM and GRU models. Zhen (2025) [14] used five algorithms (Pearson, MIC, Lasso, RFE, Null Importance) to screen out 26 key features. By adding an Attention module to the original CNN-LSTM model, feature enhancement was achieved, with a focus on the impact of sentiment indicators. The model's prediction error was reduced, with an MAE as low as 0.0097 in the food and beverage industry, approaching the "zero-error" level. Deep learning, with its powerful automatic feature extraction capabilities, has become an important research direction in the field of stock prediction. By constructing a multi-layer neural network structure, the model can effectively capture the nonlinear, temporal, and multi-source heterogeneous features present in financial markets. However, data defects such as low signal-to-noise ratio and non-stationarity, along with characteristics like high complexity and black-box nature, still pose challenges such as strong sensitivity to prediction cycles and high overfitting risks. Additionally, the model exhibits weak robustness when addressing dynamic market changes. To address these issues, we enhance interpretability through model fusion, feature enhancement, and causal models, and introduce a reinforcement learning framework to achieve dynamic market adaptability.

Table 1 Comparative analysis of machine learning and deep learning models in stock market prediction

Category	Author	Year	Algorithm	Dataset performance	Advantages	limitations
Machine Learning	Zabir Haider Khan	2011	ANN	Average error of 1.53% for GI, P/E, NAV, EPS, and Volume	Can capture nonlinear relationships; applicable to multivariate analysis	The network is relatively simple and lacks explanatory power; there is a risk of overfitting.
	Khalid Alkhatib	2013	KNN	Under the data of three different companies, AIEI (RMSE = 0.0378), JOST (RMSE = 0.030), and APOT (RMSE = 0.337)	Simple and easy to implement; no parameter assumptions; real-time performance.	Poor generalization ability; poor dynamic market adaptability; limited model interpretability; low computational efficiency.
	Lin Nana	2018	Random Forest	The accuracy rate for stock price increases is 73.33%, and the accuracy rate for stock price decreases is 87.14%.	High accuracy and robustness; Strong parameter interpretability; Automatic identification of key indicators	High computational resource consumption; increased depth may result in black box characteristics; unsuitable for structural changes in the market.
	Lin Feng	2019	SVM	SVM predicts an account yield of 180.88% (annualized 35.34%).	Dynamic adaptability; strong generalization ability; strong risk control	High parameter sensitivity; risk of overfitting; other features ignored
Deep Learning	Zeng An	2019	Bidirectional LSTM	Contains 30 stocks, MAE is 0.075, RMSE is 0.114, $R^2$is 0.861	Bidirectional time modeling, utilizing both historical and future information; multi-hidden layer networks can adapt to different time window lengths;	Bidirectional LSTM networks are deeper and more computationally complex; features are single.
	Yongqiong Zhu	2020	RNN	Training and testing on AAPL stock MSE is 68.49, RMSE is 8.28 MAE is 5.90	Suitable for time series modeling, deeply capturing characteristics between time periods; strong short-term forecasting capabilities.	Gradient vanishing and gradient explosion; feature monotony; limited generalization
	FATEMEH MOODI	2024	GRU LSTM CNN-LSTM	For TSLA and AAPL stocks, the accuracy rate of CNN-LSTM was 96.83%, the accuracy rate of LSTM was 95.02%, and the accuracy rate of GRU was 95.63%.	Simultaneously utilizing technical indicators and sentiment analysis; combining the advantages of hybrid models: CNN feature extraction and LSTM temporal dependency	Relies on large amounts of data from multiple sources; API retrieval real-time performance cannot be guaranteed; generalization capabilities have not been verified, only performing well on specific stocks.
	Kehan Zhen	2025	CNN-LSTM-Attention	The MAE for ten industries is below 0.05, with the food and beverage industry having an MAE of 0.0097.	High robustness; combines feature extraction, temporal dependency, and attention enhancement to fuse model advantages; high generalization ability.	High dependence on data timeliness and high model complexity

3.3 Reinforcement learning

Reinforcement learning, with dynamic decision-making at its core, expands stock prediction from “static trend analysis” to a comprehensive system encompassing “strategy optimization and risk management,” providing a new framework for quantitative investing. By establishing a “state-action-reward” mechanism, the algorithm can autonomously learn optimal trading strategies in uncertain market environments. From continuous action models such as DDPG and TD3 to the integration of DQN with traditional models, reinforcement learning demonstrates significant risk-return balancing capabilities in multi-asset portfolio management and dynamic position adjustments. Consequently, numerous studies have focused on applying reinforcement learning frameworks to achieve efficient dynamic asset allocation.

Qi Yue (2018) [15] used the DDPG algorithm to construct an Actor-Critic framework, processing historical stock prices, returns, and other time series data through LSTM to extract features as state information, and using asset weight allocations as action signals to achieve multi-asset portfolio management. Applying this framework achieved a total portfolio increase of 65.09%. Kabbani (2022) [16] applied the TD3 framework, leveraging its continuous action space design to enable dynamic position adjustments. Additionally, a POMDP (Partial Markov Decision Process) was employed to specifically address market uncertainty. Experimental results showed that TD3 converges faster in continuous action spaces, with cumulative returns improving by 2-3 times compared to discrete models. Song Fei (2025) [17] combined Deep Q-Network (DQN) with Support Vector Machine (SVM) and XGBoost to build a smart investment decision-making system based on predictive signals, achieving online learning and adaptive trading strategy optimization. This study demonstrates the potential of reinforcement learning in financial decision-making, but its performance highly depends on the accuracy of the underlying predictive models, making stock trend prediction a core issue in quantitative investing. Chen (2025) [18] first performs a prediction using the deepseek-r1-distill-llama-70b large language model, then integrates a risk-aware PPO adjustment module. The state space includes predictions of LLM-predicted prices, historical prices, etc., while the action space fine-tunes the LLM-predicted prices. The model's cumulative return improves by approximately 25% compared to XGBoost.

Reinforcement learning demonstrates strong market adaptability in the stock market, enabling adaptive trading strategies and balanced optimization of risk and return through continuous trial-and-error learning in uncertain environments. It particularly excels in tasks such as multi-asset allocation, position adjustment, and market trend identification. Current research extensively explores the integration of DDPG, TD3, DQN, and their variants with other models, driving the practical application of reinforcement learning in quantitative investment scenarios. However, the strong non-stationarity of the market and the profit-oriented reward function settings may overlook real-world risk management needs. Future research could integrate external predictive models, market mechanism priors, and risk-aware optimization strategies, combined with multi-objective reward function settings, to enhance its practicality and robustness in complex financial environments.

4 Combining big data technology

4.1 Hadoop

The Hadoop distributed computing framework, with its high fault tolerance and massive data processing capabilities, has become the foundational platform for financial big data analysis. In stock prediction, it effectively addresses the challenges of real-time processing of unstructured data such as social media stream data and high-frequency trading data through MapReduce parallel computing and HDFS distributed storage. This enables the integration of cross-modal data and large-scale model training to transition from theory to practice, driving the transformation of predictive research toward a big data-driven paradigm.

Attigeri (2015) [19] broke through the traditional single paradigm by utilizing the Hadoop platform to process massive amounts of real-time social media stream data during data processing, combining sentiment indicators and fundamental indicators for stock price prediction, and using distributed Hadoop to achieve efficient model training. Arun (2015) [20] combined neural network models with distributed training, using MapReduce to achieve distributed batch training. The training set was divided, with each node performing parallel gradient calculations, and the Reducer aggregating gradients to update weights. Experiments demonstrated that a 50-node Hadoop cluster processing 350 Map tasks achieved a speed improvement of up to 6 times compared to a single-node dual-core processor. Jose (2019) [21] not only utilized distributed computing frameworks to accelerate data extraction and computation but also pioneered the integration of GA with technical analysis, running trading strategies in parallel on a Hadoop cluster. This ultimately reduced the training time for 42 days of order book data from several days on a single machine to hours.

Hadoop, with its high fault tolerance and robust offline batch processing capabilities, is well-suited for processing complex, large-scale historical market data, sentiment information, and transaction records, providing a stable distributed data processing platform for stock prediction. However, Hadoop has certain limitations in terms of real-time performance and struggles to meet the demands of rapidly changing financial markets.

Table 2 Comparative analysis of reinforcement learning models in stock market prediction

Category	Author	Year	Algorithm	Dataset performance	Advantages	limitations
Reinforcement learning	Qiyue	2018	DDPG	16 CSI 100 Index constituent stocks. The DDPG portfolio achieved a cumulative return of 65.09%, which is 2.5 times that of the control group.	Processing with continuous action space, strong strategy adaptability; using LSTM to capture market time series data	The model is highly dependent on data; its handling of extreme events has not been verified.
	TAYLAN KABBANI	2022	TD3	2 stocks (QCOM, MSFT) with initial capital of $10,000, cumulative return of $17,721, and Sharpe ratio of 3.14; 10 stocks (AAPL, MSFT, etc.) with initial capital of $100,000, cumulative return of $110,308, and Sharpe ratio of 2.68.	Continuous action space adapts to dynamic position adjustments, reducing value function overestimation through dual critic networks and delayed updates.	High computational complexity; strong data dependency; zero slippage assumption deviates significantly from reality in highly volatile markets; weak generalization across different markets.
	Song Fei	2025	neural network+DQN	China Merchants Bank, Taihe Technology, and four additional stocks The yields on China Merchants Bank and Taihe Technology reached 228% and 356%, respectively.	Neural network fitting of Q values to solve high-dimensional space prediction problems; dynamic market adaptability	Discrete action restrictions, unable to refine positions; highly data-dependent
	Qizhao Chen	2025	LLM+PPO	Apple, HSBC, Pepsi, Tencent, and Toyota: MAPE = 0.081, RMSE = 0.028, Sharpe ratio = 1.04 Maximum drawdown = -0.22, cumulative return = 0.23	Combines the advantages of LLM fusion; PPO incorporates VaR/CVaR into the reward function to reduce risk losses; dynamic adjustment for enhanced real-time performance; cross-domain integration.	High computational complexity of fusion models; insufficient coverage of extreme risks; high data dependency.

4.2 Spark

Spark leverages its advantages in in-memory computing and stream processing to address Hadoop's limitations in real-time scenarios, making it the core technological backbone for big data applications in stock prediction. Its distributed parallel computing framework not only accelerates the training efficiency of machine learning models (e.g., reducing random forest training time by 60%) but also enables real-time analysis of dynamic data such as market sentiment and news sentiment through Spark Streaming, providing an efficient computational engine for high-frequency trading strategies and adaptive prediction models.

Das et al. (2018) [22] processed Twitter stream data using the Spark platform, integrating public sentiment into the prediction framework through sentiment analysis, thereby achieving the fusion of big data technology with traditional prediction methods. Jiang Xianya (2019) [23] further utilized the Spark framework to distribute and parallelize the execution of four machine learning models (Naive Bayes, random forest, decision tree, and logistic regression algorithms), all of which significantly improved training efficiency, with the random forest algorithm reducing execution time by 60% compared to single-machine execution. Kabbani (2022) [24] leveraged Spark's distributed computing capabilities to process 9.2GB of news data, enhancing sentiment analysis efficiency by over 400% through parallel text processing. This enables real-time tracking of market sentiment with daily data updates. Spark provides robust computational power for multi-source data-driven investment decision-making. Koduppanapolackal (2025) [25] distributed the training tasks of the LSTM model across cluster nodes using a Spark-based model, enabling each node to independently compute gradients, with the Reducer aggregating global parameters. This efficient computation simultaneously achieved grid search optimization of hyperparameters, improving computational resource utilization while enhancing the Sharpe ratio of its personalized investment strategy by 0.2–0.5 compared to generic models.

Compared to Hadoop's offline batch processing advantages, Spark focuses on in-memory computing and stream processing capabilities, making it more suitable for addressing the real-time prediction requirements of “low latency and high responsiveness” in financial markets. It demonstrates greater flexibility and computational efficiency in areas such as multi-source data fusion, online model training, and real-time sentiment analysis.

5 Conclusion

5.1 Research summary

This study systematically reviews the research progress of artificial intelligence and big data technology in the field of stock price prediction, organizing the findings into three categories: data characteristics, algorithm models, and big data computing frameworks. It summarizes a large number of relevant literature in recent years and explores the application of different methods in the field of stock prediction.

At the data feature level, researchers in this field have gradually transitioned from traditional single features such as historical stock prices to technical indicators obtained through feature engineering, and further to comprehensive feature construction schemes that integrate multi-source heterogeneous data such as text sentiment, macroeconomics, and policy uncertainty. Extracting data features related to stock price fluctuations from hundreds of features not only enhances the interpretability of models in complex and dynamic markets and explores changes in multi-dimensional markets but also improves modeling efficiency and model stability.

At the algorithmic model level, machine learning and deep learning serve as the mainstream modeling frameworks in stock prediction research. The former leverages its nonlinear fitting capabilities to effectively uncover price fluctuation patterns, while the latter demonstrates stronger predictive capabilities in complex markets through automatic feature extraction and sequence modeling. Reinforcement learning, as a dynamic strategy framework adept at capturing market changes, effectively enhances the robustness of models in predicting dynamic market changes. When combined with decision optimization and risk control, it can drive the intelligent development of quantitative investment strategies.

At the big data computing framework level, this study lists two big data platforms commonly used in the field of stock prediction. Hadoop and Spark provide underlying support for the processing of large-scale financial data. Hadoop is suitable for batch processing and model training, while Spark compensates for its lack of real-time capabilities, enabling online learning and rapid response to public opinion. Together, they form the data infrastructure of financial intelligent prediction systems.

5.2 Limitations analysis

Currently, research in the field of stock research has made varying degrees of progress. However, the highly uncertain behavior inherent in financial markets, as well as issues such as partial data gaps, noise, and non-stationarity in financial data, remain as challenges that scholars in this field are committed to addressing. First, the high noise and non-stationarity of financial data conflict with the high data quality requirements of machine learning and deep learning models, which can lead to a decline in model accuracy. Second, although the task is to predict stock prices, the varying market environments, policy factors, and financing conditions across industries and markets pose challenges to the model's generalization capabilities. Third, while deep learning and reinforcement learning can achieve higher prediction accuracy compared to statistical models or machine learning, their lack of interpretability creates a black-box problem, making it difficult to meet the dual demands of financial regulation and investor trust. Finally, stock markets are often influenced by extreme, sudden events that cause significant volatility, and such volatility is often difficult to model using historical data. Existing models face numerous challenges in capturing these environmental factors and potential influencing factors.

Despite market randomness and environmental uncertainty, with the continuous development of technology, a large number of new algorithms have emerged, and the development of LLM has further promoted the development of stock prediction research. In the future, there will be opportunities to fully capture multi-source, heterogeneous, and multi-modal data types and environmental features, effectively capture and analyze relevant lag signals, and combine dynamic prediction strategies to effectively avoid risks when sudden events occur. Therefore, stock prediction research based on models and algorithms not only holds significant academic value at the artificial intelligence algorithm level but also demonstrates broad application prospects.

References

[1] Luo Yufeng, Zhang Yuangu. Application of Markov Chains in Stock Market Investment [J]. Science and Education Guide (Mid-Month Edition), 2013, (22): 165-166.

[2] Ariyo, A. A., Adewumi, A. O., & Ayo, C. K. (2014). Stock Price Prediction Using the ARIMA Model. 2014 UKSim-AMSS 16th International Conference on Computer Modelling and Simulation.

[3] Zhang Kai, Ren Weiping, Zhang Yangsen, et al. Research on Stock Prediction Methods Based on Stock Investor Comments [J]. Journal of Beijing University of Information Science and Technology (Natural Science Edition), 2017, 32(05): 67-71.

[4] TSAI M-C, CHENG C-H, TSAI M-I. A multifactor fuzzy time-series fitting model for forecasting the stock index[J]. Symmetry, 2019, 11(12):1474.

[5] AYALA J, GARCíA-TORRES M, NOGUERA J L V, et al. Technical analysis strategy optimization using a machine learning approach in stock market indices[J]. KnowledgeBased Systems, 2021, 225:107119.

[6] ZHANG L, CHEN W, HU N. Economic policy uncertainty and stock liquidity: evidence from China[J].International Journal of Emerging Markets, 2023, 18(1):22-44.

[7] Haider Khan, Z., Sharmin Alin, T., & Hussain, A. (2011). Price Prediction of Share Market Using Artificial Neural Network "ANN." International Journal of Computer Applications, 22(2), 42–47.

[8] Alkhatib, K. ,Najadat, H. ,Hmeidi, I. , &Shatnawi, M. K. A. . (2013). Stock price prediction using k-nearest neighbor (knn) algorithm. Ijbhtnet Com.

[9] Lin Nana, Qin Jiangtao. Research on the prediction of A-share stock price fluctuations based on random forests [J]. Journal of Shanghai University of Technology, 2018, 40(03): 267-273+301.

[10] Lv, Linfeng. Research on Intraday Arbitrage of Stocks Based on SVM [J]. Business News, 2019, (11): 39+41.

[11] Zeng An, Nie Wenjun. Stock Recommendation System Based on Deep Bidirectional LSTM [J]. Computer Science, 2019, 46(10): 84-89.

[12] Zhu, Y. (2020). Stock price prediction using the RNN model. Journal of Physics: Conference Series, 1650, 032103.

[13] Fatemeh Moodi, Amir Jahangard Rafsanjani, Sajad Zarifzadeh, & Ali, M. (2024). Fusion of technical indicators and sentiment analysis in a hybrid framework of deep learning models for stock price movement prediction.IEEE Access, 1–1.

[14] Zhen, K., Xie, D., & Hu, X. (2025). A multi-feature selection fused with investor sentiment for stock price prediction. Expert Systems with Applications, 127381–127381.

[15] Qi Yue, Huang Shuohua. Portfolio Management Based on the Deep Reinforcement Learning DDPG Algorithm [J]. Computers and Modernization, 2018, (05): 93-99.

[16] Kabbani, Taylan, and Ekrem Duman. "Deep Reinforcement Learning Approach for Trading Automation in the Stock Market." IEEE Access, 2022, pp. 1–1,

[17] Song Fei, Tian Chenlei, and Feng Chuanwei. Research on Stock Quantitative Investment Based on Deep Reinforcement Learning [J]. Journal of Nanjing Normal University (Natural Science Edition), 2025, 48(01): 85-92.

[18] Chen, Q. (2025) A Two-Stage Framework for Stock Price Prediction LLM-Based Forecasting with Risk-Aware PPO Adjustment. Journal of Computer and Communications, 13, 120-139. - References - Scientific Research Publishing. (2025). Scirp.org. https://www.scirp.org/reference/referencespapers?referenceid=4012850.

[19] Attigeri, G. V., Manohara Pai M M, Pai, R. M., & Nayak, A. (2015). Stock market prediction: A big data approach. TENCON 2015 - 2015 IEEE Region 10 Conference.

[20] Jose, J., Mana, S. C., & Samhitha, B. K. (2019). An efficient system to predict and analyze stock data using Hadoop techniques. International Journal of Recent Technology and Engineering (IJRTE),(2), 2277-3878.

[21] Arun, D., Vanita, K., & Mittal, A. P. (2015). Stock market prediction using Hadoop Map-Reduce ecosystem. In Computing for Sustainable Global Development (INDIACom), 2015 2nd International Conference on. IEEE.

[22] Das, S., Behera, R. K., kumar, M., & Rath, S. K. (2018). Real-Time Sentiment Analysis of Twitter Streaming data for Stock Prediction. Procedia Computer Science, 132, 956–964.

[23] Xianya, J., Mo, H., & Haifeng, L. (2019). Stock Classification Prediction Based on Spark. Procedia Computer Science, 162, 243–250.

[24] Kabbani, T., & Usta, F. E. (2022). Predicting The Stock Trend Using News Sentiment Analysis and Technical Indicators in Spark. ArXiv: 2201.12283 [Cs, Q-Fin]. https://arxiv.org/abs/2201.12283.

[25] Koduppanapolackal, J. J., Ahamed, N., Suresh, A., U.A. Adwaith, & P. Gouthaman. (2025). Real-Time Stock Price Prediction and Visualization Using LSTM and Matplotlib. 605–609.