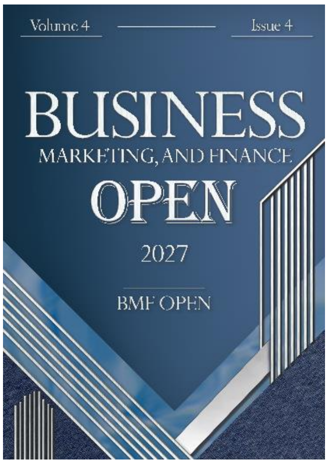
Designing a Model for Stock Price Prediction Based on the Integration of Machine Learning Models

Farhang Sahabi[1], Mehdi Zeynali [2,*] and Yagoub Alavimatin [3]

[1] Department of Management, Ta.C., Islamic Azad University, Tabriz, Iran;
[2] Department of Accounting, Ta.C., Islamic Azad University, Tabriz, Iran;
[3] Department of Management, Ta.C., Islamic Azad University, Tabriz, Iran;

* Correspondence: Dr.zeynali@iau.ac.ir



Citation: Sahabi, F., Zeynali, M., & Alavimatin, Y. (2027). Designing a Model for Stock Price Prediction Based on the Integration of Machine Learning Models. *Business, Marketing, and Finance Open*, 4(4), 1-21.

Received: 19 March 2026
Revised: 23 June 2026
Accepted: 30 June 2026
Initial Publication: 01 July 2026
Final Publication: 01 July 2027

Abstract: This study aimed to design and evaluate an integrated machine learning model for predicting the stock prices of companies listed on the Tehran Stock Exchange using market, technical, trading, and financial predictors. This applied, quantitative, longitudinal, and predictive-modeling study was conducted on companies listed on the Tehran Stock Exchange during 2014–2024. After applying inclusion and exclusion criteria, 148 companies were selected as the final sample, and 301,284 firm-day observations were analyzed. The predictor variables included historical price data, trading volume, trading value, number of transactions, market capitalization, technical indicators, volatility indicators, and selected financial ratios. Data were cleaned, adjusted, normalized, and transformed into a supervised learning structure. The dataset was divided chronologically into training, validation, and test sets. Several machine learning algorithms, including multiple linear regression, ridge regression, decision tree, support vector regression, random forest, gradient boosting, extreme gradient boosting, and long short-term memory neural network, were trained and compared. Finally, an integrated stacking ensemble model was developed. The results showed that nonlinear machine learning models outperformed linear models in predicting adjusted closing stock prices. Among the individual models, extreme gradient boosting achieved the strongest performance, with MAE = 398.26, RMSE = 642.37, MAPE = 6.18%, R^2 = 0.948, and directional accuracy = 72.61%. The final stacking ensemble model demonstrated the best overall predictive performance, with MAE = 331.44, RMSE = 548.63, MAPE = 5.08%, R^2 = 0.964, and directional accuracy = 76.39%. Feature-importance analysis indicated that lagged adjusted price, moving averages, market capitalization, trading value, earnings per share, MACD, RSI, trading volume, volatility, and return on equity were the most influential predictors. The findings indicated that integrating multiple machine learning models improves stock price prediction accuracy compared with single-model approaches. The proposed stacking ensemble model can serve as an effective decision-support tool for investors, analysts, and portfolio managers in emerging stock markets.

Keywords: Stock price prediction; machine learning; stacking ensemble; Tehran Stock Exchange; financial forecasting; artificial intelligence.

1. Introduction

Stock price prediction has long been one of the most complex and strategically important subjects in financial economics, investment management, and computational finance. The stock market reflects a dynamic interaction among firm-level fundamentals, macroeconomic expectations, investor psychology, liquidity conditions, institutional behavior, political developments, and speculative trading pressures. As a result, predicting stock prices is not merely a statistical task but a multidimensional analytical problem in which nonlinear patterns, temporal dependencies, sudden structural breaks, and noisy market signals must be identified and modeled. Traditional theories of financial markets have often emphasized the difficulty of forecasting prices because publicly available information is rapidly absorbed into market valuations. However, the increasing availability of large-scale financial data and the rapid development of artificial intelligence have created new opportunities for extracting hidden patterns from complex market environments. In this context, machine learning has become one of the most widely discussed approaches for improving the accuracy, adaptability, and practical usefulness of stock price prediction models [1-3].

The growing use of artificial intelligence in finance has transformed the methodological foundations of market analysis. Conventional econometric models are usually based on assumptions such as linearity, stationarity, normality, and predefined functional relationships among variables. Although such models are useful for theoretical interpretation and basic forecasting, they often fail to capture the nonlinear, high-dimensional, and unstable nature of financial markets. Machine learning methods, by contrast, are designed to learn from data and improve prediction through pattern recognition, iterative optimization, and flexible modeling of complex relationships. In stock price prediction, these methods can process large volumes of historical prices, trading volumes, technical indicators, financial ratios, textual information, and market sentiment data. Because stock markets generate continuous and heterogeneous streams of information, machine learning provides a suitable framework for transforming raw market data into predictive signals [4-6].

Stock price forecasting is particularly difficult because price movements are affected by both observable and latent factors. Observable factors include opening price, closing price, high and low prices, trading volume, trading value, earnings per share, price-to-earnings ratio, leverage, profitability, and liquidity indicators. Latent factors include investor expectations, market fear, herd behavior, news interpretation, and reaction to uncertainty. In addition, market behavior is not constant over time; relationships that are useful in one period may weaken or disappear in another. For this reason, prediction models must be capable of learning nonlinear interactions and adapting to changing market regimes. Studies on stock market prediction have increasingly shown that machine learning algorithms can provide stronger predictive performance than purely linear or rule-based methods because they can identify complex and hidden structures in financial time series [7-9].

Among the machine learning methods used in stock price prediction, regression-based algorithms, decision trees, random forests, support vector machines, gradient boosting models, artificial neural networks, and deep learning architectures have received extensive attention. Each of these approaches has specific advantages and limitations. Linear regression and regularized regression models are simple, interpretable, and computationally efficient, but their predictive power is limited when the relationship between predictors and stock prices is nonlinear. Decision tree models can detect nonlinear thresholds and interactions, but they are often unstable when used alone. Random forest models improve this weakness by combining multiple trees and reducing variance. Support vector regression can handle nonlinear relationships through kernel functions, but it may be computationally expensive for large datasets. Gradient boosting and extreme gradient boosting models are powerful ensemble methods that sequentially improve weak learners and often produce high accuracy in structured financial datasets. Neural networks and deep learning models are particularly useful when the objective is to capture temporal dependencies and sequential patterns in historical market data [5, 10-12].

Despite the progress achieved by individual machine learning algorithms, the use of a single model may be insufficient for stock price prediction. Financial markets are complex systems in which different types of models may capture different aspects of price behavior. For example, tree-based models may be effective in modeling nonlinear interactions among technical and financial variables, while recurrent neural networks may be better suited for learning temporal structures in sequential data. Similarly, support vector regression may be useful for identifying nonlinear boundaries, while gradient boosting models may improve performance by reducing prediction error iteratively. Therefore, integrated or hybrid models are increasingly recommended in the literature because they can combine the strengths of multiple algorithms and reduce the limitations of relying on a single predictive structure. Hybrid and ensemble strategies are especially useful when financial data are noisy, unstable, and influenced by multiple sources of uncertainty [3, 13, 14].

The integration of machine learning models can take several forms, including weighted averaging, voting, bagging, boosting, stacking, and hybrid optimization. In a weighted averaging approach, predictions from several models are combined according to their relative accuracy. In stacking, the predictions of base models are entered into a meta-model that learns how to combine them optimally. Hybrid optimization methods may also use evolutionary or swarm intelligence algorithms to improve parameter selection and model performance. These approaches are relevant to stock price prediction because market behavior is rarely explained by one dominant pattern. Instead, price movement may reflect a combination of short-term momentum, medium-term trend, firm-specific fundamentals, liquidity conditions, and broader market sentiment. Integrated models can therefore provide more stable predictions by allowing different algorithms to contribute complementary information [11, 14, 15].

Another important development in stock price forecasting is the incorporation of technical indicators into machine learning models. Technical indicators such as moving averages, exponential moving averages, momentum, relative strength index, moving average convergence divergence, rate of change, rolling volatility, and average true range are commonly used to represent historical price behavior, trend direction, overbought or oversold conditions, and market risk. These indicators are not merely descriptive; when transformed into predictive features, they allow algorithms to identify recurring price patterns and market reactions. Research has shown that models trained on technical indicators can often improve the prediction of closing prices and price movement direction, especially when these indicators are combined with raw market variables such as price and volume [4, 5, 7, 8].

In addition to technical indicators, accounting-based and firm-level financial variables are also important for stock price prediction. Stock prices are influenced by investors' perceptions of profitability, growth potential, financial risk, liquidity, and operational efficiency. Variables such as earnings per share, return on assets, return on equity, price-to-earnings ratio, current ratio, debt-to-assets ratio, and asset turnover can provide information about a company's financial strength and market valuation. Although short-term price movements may be strongly affected by trading behavior and sentiment, medium-term and long-term stock values remain connected to firm fundamentals. Therefore, prediction models that rely only on historical price data may overlook important information embedded in financial statements. Combining technical, market, and accounting variables can provide a richer feature space and increase the ability of machine learning algorithms to detect both short-term and structural determinants of stock prices [2, 3, 9].

Recent research has also emphasized the role of textual information and investor sentiment in market prediction. Financial news, corporate announcements, analyst reports, and social media content may affect investor behavior and create immediate reactions in stock prices. News-based prediction models use text mining, natural language processing, and sentiment analysis to extract information from unstructured textual sources and convert it into numerical predictors. Although the present study focuses mainly on market, technical, and financial variables, the broader literature demonstrates that machine learning has expanded stock price prediction beyond conventional numerical data. The integration of sentiment scores with machine learning models has shown that news and textual signals can improve market forecasting, particularly when price movements are influenced by public expectations and information shocks [16, 17].

The expanding literature on artificial intelligence in financial markets shows that machine learning has become a central component of modern investment analytics. Bibliometric evidence indicates a rapid increase in studies using artificial intelligence, machine learning, and deep learning in finance, reflecting the growing importance of computational methods in risk management, portfolio optimization, financial forecasting, and algorithmic trading. This development is not limited to mature financial markets; emerging and frontier markets have also become important contexts for testing predictive models because they often show higher volatility, lower informational efficiency, and stronger behavioral effects. In such markets, machine learning may offer particular advantages because it can identify nonlinear relationships that are not easily captured by traditional financial models [1, 15, 18].

Deep learning has further advanced stock price prediction by enabling models to learn hierarchical and temporal representations from complex datasets. Recurrent neural networks, long short-term memory networks, convolutional neural networks, and hybrid deep learning architectures have been applied to financial time series because they can process sequential data and detect dependencies across time. Deep learning methods are especially relevant when historical price sequences contain delayed effects, trend reversals, momentum persistence, or volatility clustering. However, deep learning models also require careful design, sufficient data, computational resources, and appropriate validation procedures. Overfitting remains a major concern, particularly in financial prediction, where high in-sample accuracy may not translate into reliable out-of-sample performance. Therefore, model evaluation must rely on chronological train-validation-test separation and robust error metrics rather than random splitting alone [3, 6, 19].

Volatility is another core issue in stock market forecasting. Stock prices do not move smoothly; they fluctuate in response to uncertainty, market shocks, liquidity pressure, and changes in investor expectations. Volatility not only affects risk but also influences the predictability of prices. Periods of high volatility may produce larger prediction errors because price movements become more abrupt and less dependent on recent historical patterns. Machine learning models can incorporate volatility indicators such as rolling standard deviation, average true range, and historical variance to improve the modeling of risk-sensitive price behavior. Recent evidence suggests that machine learning can also be used to forecast and simulate stock market volatility, thereby supporting both price prediction and risk assessment [19, 20].

The predictive value of machine learning has also been examined in relation to specialized financial contexts, including innovative small and medium-sized enterprises, ESG stock markets, clean energy prices, and sector-specific banking stocks. These studies demonstrate that the usefulness of machine learning depends not only on the algorithm but also on the characteristics of the market, industry, and data structure. For example, companies with greater innovation intensity, firms operating in environmentally sensitive sectors, or stocks influenced by clean energy prices may display distinct patterns that require flexible modeling techniques. Similarly, sector-specific models may perform differently because industries vary in liquidity, volatility, regulatory exposure, and sensitivity to macroeconomic conditions. These findings support the need to examine machine learning models within specific market contexts rather than assuming that one model will perform equally well across all markets and sectors [9, 10, 20].

The Tehran Stock Exchange provides a meaningful context for developing and testing an integrated machine learning model for stock price prediction. As an emerging market, it is characterized by periods of high volatility, structural changes, sectoral concentration, macroeconomic uncertainty, and sensitivity to policy and currency conditions. These characteristics make stock price prediction both difficult and practically important. Investors, analysts, portfolio managers, and policymakers require models that can process large amounts of historical and financial data and provide more accurate forecasts. However, many existing studies have either relied on a limited number of algorithms, focused only on technical indicators, or evaluated models without sufficient integration of market, technical, and firm-level variables. In addition, the predictive performance of integrated models in emerging market settings remains an important research gap. Systematic reviews have emphasized that although machine learning has produced promising results in stock prediction, further studies are needed to compare algorithms, combine models, and evaluate their performance across different financial environments [2, 3, 21].

Designing an integrated stock price prediction model requires methodological attention to feature engineering, data preprocessing, model selection, and validation strategy. Financial datasets often contain missing values, extreme observations, nonstationary behavior, scale differences, and temporal dependencies. Therefore, data must be cleaned, normalized, and transformed into a supervised learning structure before model training. Lagged variables and technical indicators must be created in a way that prevents data leakage. Moreover, the evaluation of predictive accuracy should be conducted on unseen future observations rather than randomly mixed samples. Appropriate metrics such as mean absolute error, root mean square error, mean absolute percentage error, coefficient of determination, and directional accuracy can provide complementary information about the predictive quality of models. These methodological considerations are essential because stock price prediction is useful only when a model performs reliably outside the training sample [4-7].

The integration of machine learning models is therefore a promising strategy for improving stock price prediction in the Tehran Stock Exchange. By combining the outputs of algorithms such as random forest, support vector regression, gradient boosting, extreme gradient boosting, and long short-term memory networks, it becomes possible to design a model that captures nonlinear interactions, sequential dependencies, market liquidity effects, and financial performance signals simultaneously. Such a model may reduce prediction error, improve directional accuracy, and provide a more stable forecasting framework for investors and researchers. The importance of this approach is reinforced by recent studies showing that advanced machine learning, hybrid modeling, and artificial intelligence-based trading systems can improve financial prediction when they are carefully trained, validated, and evaluated [13, 14, 18, 19].

The aim of this study was to design and evaluate an integrated model for predicting the stock prices of companies listed on the Tehran Stock Exchange based on the combination of machine learning algorithms and market, technical, and financial predictors.

2. Methodology

The present study was applied in terms of objective and quantitative, longitudinal, and predictive-modeling in terms of method. The study was designed to develop and evaluate an integrated machine learning model for predicting stock prices of companies listed on the Tehran Stock Exchange. Since the purpose of the research was to identify predictive patterns in historical market and financial data, the study used archival data and did not involve human participants. The statistical population consisted of all companies listed on the Tehran Stock Exchange during the period from 2014 to 2024. After applying the inclusion and exclusion criteria, 148 companies were selected as the final sample. The inclusion criteria were continuous listing on the Tehran Stock Exchange during the study period, availability of complete stock price and trading data, availability of audited financial statements, absence of major changes in fiscal year, and no prolonged trading suspension during the period under investigation. Companies belonging to banks, insurance firms, investment companies, financial intermediaries, and holding companies were excluded because their financial structures and reporting frameworks differ substantially from those of other listed firms. Based on these criteria, the final dataset included 148 listed companies across different industries, and the data structure was organized as a longitudinal panel of market and firm-level observations. The target variable of the study was the future stock price, while the predictor variables included historical price indicators, trading variables, return-based measures, volatility indicators, and selected financial ratios extracted from the published statements of the companies.

The data collection procedure was based on a structured archival extraction checklist developed by the researchers in accordance with the objectives of the study. This checklist was used to record and organize all required variables, including daily opening price, closing price, highest price, lowest price, adjusted price, trading volume, trading value, number of transactions, market capitalization, and lagged return indicators. In addition, firm-level financial data were extracted from annual financial statements and included profitability, liquidity, leverage, efficiency, and valuation ratios. The market data were obtained from the official trading records of the Tehran Stock Exchange and the Tehran Securities Exchange Technology Management system, while financial statement data were collected from the official disclosure system of listed companies and verified through financial information databases. To increase the accuracy of the dataset, all extracted data were checked for duplicate records, missing values, abnormal observations, and inconsistencies between price series and corporate events such as capital increases, dividends, and stock splits. Adjusted prices were used in the modeling process in order to reduce distortions caused by corporate actions and to improve the comparability of price series over time.

In addition to raw market and financial variables, several technical indicators were calculated and added to the dataset as predictive features. These indicators included moving averages over different time windows, exponential moving averages, price momentum, relative strength index, moving average convergence divergence, rate of change, rolling volatility, average true range, and lagged returns. These variables were selected because they can represent short-term and medium-term price behavior, trend direction, market momentum, and risk fluctuations. The final feature set therefore combined historical price information, trading behavior, technical indicators, and accounting-based financial ratios. Data preparation, feature engineering, model training, and statistical evaluation were conducted using Python programming language. The main computational libraries used in the analysis included Pandas and NumPy for data cleaning and numerical processing, Scikit-learn for conventional machine learning algorithms and evaluation metrics, XGBoost for gradient boosting modeling, and TensorFlow/Keras for deep learning implementation. The use of these tools made it possible to process large-scale time-series and panel data, construct multiple predictive models, and integrate their outputs within a unified modeling framework.

Data analysis was conducted in several sequential stages. First, the raw data were cleaned and preprocessed. Missing values were examined and, where appropriate, handled using interpolation or deletion depending on the nature and frequency of missingness. Outliers were identified through statistical screening and visual inspection of abnormal price and volume movements. Since stock market data are highly sensitive to scale differences, all numerical predictor variables were normalized or standardized before entering the machine learning models. The dataset was then transformed into a supervised learning structure by creating lagged variables and defining the prediction horizon. In this study, the model was designed to predict the next-period adjusted closing price based on previous market, technical, and financial information. To avoid data leakage, the dataset was divided chronologically rather than randomly. The earlier observations were used for model training, the middle segment was used for validation and hyperparameter tuning, and the most recent observations were reserved for final testing. This time-based division ensured that the model was evaluated under conditions similar to real stock price forecasting, where future information is not available during model training.

Several machine learning models were first estimated separately in order to compare their individual predictive performance. These models included multiple linear regression as a baseline model, decision tree regression, random forest regression, support vector regression, gradient boosting regression, extreme gradient boosting, and long short-term memory neural network. Each model was trained using the same training data and evaluated on the validation and test datasets. Hyperparameters were optimized through systematic search procedures, including grid search and validation-based tuning. After estimating the individual models, an integrated predictive model was developed by combining the outputs of the best-performing algorithms. The integration strategy was based on ensemble learning, in which the predictions of several models were combined through weighted averaging and stacking. In the stacking procedure, the predictions generated by the base models were entered into a meta-learner, which learned the optimal combination of model outputs and produced the final predicted stock price. This approach was used because different machine learning algorithms may capture different aspects of market behavior; for example, tree-based models are effective in modeling nonlinear relationships and feature interactions, while recurrent neural networks are suitable for learning sequential dependencies in time-series data.

The performance of the individual and integrated models was assessed using standard prediction error and goodness-of-fit indices. The main evaluation criteria included mean absolute error, root mean square error, mean absolute percentage error, and coefficient of determination. Lower values of mean absolute error, root mean square error, and mean absolute percentage error indicated higher prediction accuracy, while a higher coefficient of determination indicated stronger explanatory and predictive power. The final integrated model was compared with the individual models to determine whether combining machine learning methods improved the accuracy of stock price prediction. Feature importance analysis was also conducted for tree-based models in order to identify the most influential predictors in stock price forecasting. The results of the model comparison were interpreted based on both predictive accuracy and model stability across the validation and test datasets. All analyses were performed at a quantitative and computational level, and the final model was selected according to its ability to minimize prediction error and provide stable performance on unseen data.

3. Findings and Results

The initial descriptive examination of the sample showed that the final dataset consisted of 148 companies listed on the Tehran Stock Exchange. Because the present study was conducted on listed companies rather than human participants, the demographic profile of the sample was reported in terms of firm-level characteristics, industry composition, and data availability. The sampled companies were distributed across several major industries, including chemical and petrochemical products, basic metals, food products, cement and construction materials, pharmaceuticals, automobile and parts manufacturing, machinery and equipment, mining and mineral products, non-metallic minerals, computer and electronic products, textiles, and other manufacturing and service-related industries. In terms of industry composition, 24 companies belonged to the chemical and petrochemical sector, 18 companies to basic metals, 15 companies to food products, 13 companies to cement and construction materials, 12 companies to pharmaceuticals, 11 companies to automobile and parts manufacturing, 10 companies to machinery and equipment, 9 companies to mining and mineral products, 8 companies to non-metallic minerals, 7 companies to computer and electronic products, 5 companies to textiles, and 16 companies to other manufacturing and service-related industries. The average listing age of the companies was 18.74 years, indicating that most firms had a sufficient trading history for time-series modeling. After applying data-cleaning procedures, excluding incomplete records, removing observations affected by prolonged trading suspension, and generating lagged and technical indicators, the final modeling dataset included 301,284 firm-day observations. Of these observations, 213,842 were used for training, 37,209 for validation, and 50,233 for final testing. This chronological division of the dataset made it possible to evaluate the predictive models under realistic forecasting conditions, in which the models were trained only on past observations and tested on later market data.

Table 1. Descriptive statistics of market, technical, and financial variables used in the stock price prediction model

Variable	N	Mean	Standard Deviation	Minimum	Q1	Median	Q3	Maximum
Opening price	301,284	12,486.72	9,315.48	1,145.00	5,381.25	9,872.00	16,740.50	83,950.00
Highest daily price	301,284	12,982.65	9,706.33	1,178.00	5,604.00	10,238.50	17,321.75	87,410.00
Lowest daily price	301,284	12,047.38	8,991.26	1,101.00	5,192.00	9,512.00	16,104.25	80,290.00
Closing price	301,284	12,544.91	9,402.17	1,130.00	5,402.00	9,921.00	16,881.00	85,760.00
Adjusted closing price	301,284	11,936.28	8,842.54	1,021.00	5,110.00	9,415.50	16,042.25	79,630.00
One-day lagged adjusted price	301,284	11,902.46	8,815.67	1,019.00	5,096.00	9,398.00	16,006.00	79,420.00
Five-day moving average	301,284	11,881.94	8,774.61	1,035.20	5,104.40	9,376.80	15,975.60	78,944.40
Ten-day moving average	301,284	11,843.27	8,721.35	1,049.90	5,121.10	9,328.40	15,902.20	78,213.70
Twenty-day moving average	301,284	11,771.36	8,648.09	1,068.60	5,136.70	9,276.30	15,785.90	77,805.40
Twelve-day exponential moving average	301,284	11,856.82	8,733.52	1,041.60	5,115.80	9,344.10	15,931.40	78,377.80
Twenty-six-day exponential moving average	301,284	11,719.45	8,589.76	1,075.80	5,141.40	9,217.20	15,689.70	77,402.60

Daily return (%)	301,284	0.086	2.914	-11.73	-1.42	0.04	1.56	12.38
Log return	301,284	0.001	0.029	-0.124	-0.014	0.000	0.015	0.117
Ten-day momentum	301,284	214.36	1,382.71	-8,920.00	-386.00	48.00	621.00	12,480.00
Ten-day rate of change (%)	301,284	0.91	7.84	-31.62	-3.48	0.37	4.76	42.88
Relative strength index	301,284	53.47	14.92	12.34	42.16	53.82	64.95	91.28
MACD	301,284	82.71	624.38	-4,218.60	-174.25	23.40	296.15	5,364.90
MACD signal line	301,284	77.64	589.13	-3,982.40	-163.70	19.80	274.60	5,022.10
Twenty-day rolling volatility	301,284	0.031	0.018	0.004	0.019	0.027	0.039	0.142
Average true range	301,284	518.42	406.79	42.60	226.30	403.40	693.80	4,812.70
Trading volume	301,284	9,845,216.35	18,630,547.21	10,420.00	1,184,350.00	3,917,860.00	10,284,400.00	284,710,000.00
Trading value	301,284	128,460,000,000.00	249,870,000,000.00	82,000,000.00	12,680,000,000.00	41,930,000,000.00	136,750,000,000.00	4,890,000,000,000.00
Number of transactions	301,284	1,284.73	2,016.55	8.00	214.00	607.00	1,521.00	31,870.00
Market capitalization	301,284	37,842,000,000,000.00	64,985,000,000,000.00	481,000,000,000.00	7,314,000,000,000.00	16,982,000,000,000.00	41,625,000,000,000.00	892,600,000,000,000.00
Beta coefficient	301,284	0.94	0.31	0.18	0.71	0.92	1.14	2.28
Earnings per share	301,284	1,428.36	2,104.75	-4,860.00	312.00	826.00	1,942.00	19,740.00
Price-to-earnings ratio	301,284	9.64	6.92	1.12	5.18	7.86	11.94	58.20
Return on assets (%)	301,284	11.36	8.74	-18.42	5.24	10.18	16.27	48.63
Return on equity (%)	301,284	24.71	19.63	-41.28	10.49	21.36	35.84	112.50
Current ratio	301,284	1.68	0.74	0.31	1.12	1.51	2.04	6.91
Debt-to-assets ratio	301,284	0.57	0.21	0.08	0.42	0.58	0.72	1.18

Asset turnover ratio	301,284	0.86	0.41	0.09	0.55	0.78	1.09	3.84
Gross profit margin (%)	301,284	28.43	14.58	-12.30	18.76	27.14	36.88	78.42

The descriptive results presented in Table 1 show that the sampled companies had considerable variation in market price, trading activity, firm size, and financial performance. The mean adjusted closing price was 11,936.28 Iranian rials, with a standard deviation of 8,842.54, indicating a high degree of dispersion in stock prices across companies and over time. This dispersion was expected because the dataset included firms from different industries with different capitalization levels, liquidity conditions, and growth patterns. The difference between the minimum and maximum adjusted closing prices also shows that the data contained both low-priced and high-priced stocks, making it necessary to normalize the variables before model estimation. Trading volume and trading value were highly dispersed, suggesting that liquidity was not homogeneous among the sampled companies. Some stocks were traded with relatively limited daily volume, whereas others had very high transaction intensity. The technical indicators also showed meaningful variation. For example, the relative strength index had a mean of 53.47, suggesting that, on average, the stocks were located near the neutral-to-slightly-positive momentum range, although the wide minimum and maximum values indicated the presence of both oversold and overbought conditions in different periods. The financial ratios also reflected substantial heterogeneity. Return on assets, return on equity, debt-to-assets ratio, and price-to-earnings ratio varied notably across the firm-day observations, confirming that the inclusion of accounting-based predictors could provide additional information beyond historical price and trading variables. Overall, the descriptive statistics supported the suitability of the dataset for machine learning analysis because the variables contained sufficient variation for the algorithms to learn nonlinear and time-dependent patterns in stock price behavior.

Table 2. Correlation matrix among the main predictors and adjusted closing stock price

Variable	Adjusted Price	Daily Return	Trading Volume	Trading Value	No. of Transactions	Market Cap.	MA20	RSI	MACD	Volatility	ROA	Debt-to-Assets	P/E	EPS
Adjusted Price	1.000	0.218	0.264	0.581	0.397	0.742	0.981	0.312	0.436	-0.184	0.286	-0.173	0.241	0.522
Daily Return	0.218	1.000	0.164	0.209	0.188	0.091	0.196	0.417	0.382	0.284	0.073	-0.048	0.062	0.106
Trading Volume	0.264	0.164	1.000	0.713	0.692	0.386	0.251	0.116	0.158	0.241	0.058	0.037	-0.021	0.137
Trading Value	0.581	0.209	0.713	1.000	0.744	0.628	0.562	0.181	0.274	0.148	0.143	-0.056	0.084	0.329
No. of Transactions	0.397	0.188	0.692	0.744	1.000	0.502	0.384	0.156	0.246	0.194	0.109	-0.028	0.057	0.261
Market Cap.	0.742	0.091	0.386	0.628	0.502	1.000	0.725	0.128	0.218	-0.226	0.317	-0.214	0.176	0.611
MA20	0.981	0.196	0.251	0.562	0.384	0.725	1.000	0.287	0.391	-0.201	0.279	-0.168	0.232	0.509
RSI	0.312	0.417	0.116	0.181	0.156	0.128	0.287	1.000	0.548	0.096	0.081	-0.034	0.072	0.144

MACD	0.436	0.382	0.158	0.274	0.246	0.218	0.391	0.548	1.000	0.121	0.096	-0.046	0.089	0.208
Volatility	-0.184	0.284	0.241	0.148	0.194	-0.226	-0.201	0.096	0.121	1.000	-0.137	0.192	-0.081	-0.162
ROA	0.286	0.073	0.058	0.143	0.109	0.317	0.279	0.081	0.096	-0.137	1.000	-0.426	0.194	0.468
Debt-to-Assets	-0.173	-0.048	0.037	-0.056	-0.028	-0.214	-0.168	-0.034	-0.046	0.192	-0.426	1.000	-0.117	-0.286
P/E	0.241	0.062	-0.021	0.084	0.057	0.176	0.232	0.072	0.089	-0.081	0.194	-0.117	1.000	0.173
EPS	0.522	0.106	0.137	0.329	0.261	0.611	0.509	0.144	0.208	-0.162	0.468	-0.286	0.173	1.000

The correlation results in Table 2 indicate that the adjusted closing price had the strongest positive correlation with the twenty-day moving average. This result was expected because moving averages are directly derived from historical prices and reflect the persistence of price trends over short and medium horizons. Market capitalization also showed a strong positive relationship with adjusted price, suggesting that larger firms generally had higher share prices or more stable price levels during the study period. Trading value, earnings per share, and MACD were also positively associated with adjusted price, indicating that both liquidity-related variables and profitability-related variables were relevant to price behavior. The correlation between adjusted price and volatility was negative, suggesting that stocks with higher instability tended to have lower price stability and weaker price levels. Debt-to-assets ratio was also negatively correlated with adjusted price, showing that firms with higher financial leverage were generally associated with lower stock price levels. The correlation matrix also shows that none of the independent variables, except mechanically related price-based indicators such as moving averages, had perfect or near-perfect multicollinearity. Therefore, the selected variables were considered suitable for machine learning modeling. In addition, the moderate correlations among technical indicators, trading variables, and financial ratios suggested that each group of variables contributed different informational content to the prediction task.

Table 3. Comparative performance of individual and integrated machine learning models on the test dataset

Model	Main Input Structure	MAE	RMSE	MAPE (%)	R^2	Directional Accuracy (%)	Theil's U	Training Time (seconds)	Performance Rank
Multiple linear regression	Market, technical, and financial variables	842.60	1,328.42	12.84	0.781	57.30	0.624	18.42	9
Ridge regression	Market, technical, and financial variables	805.13	1,271.65	12.21	0.794	58.41	0.603	21.36	8
Decision tree regression	Market, technical, and financial variables	715.18	1,110.34	10.91	0.838	61.08	0.552	42.77	7
Support vector regression	Scaled market, technical, and financial variables	589.32	910.25	8.96	0.891	64.42	0.451	684.19	6
Random forest regression	Market, technical, and financial variables	496.77	782.64	7.72	0.918	67.83	0.385	326.55	5
Gradient boosting regression	Market, technical, and financial variables	461.84	720.15	7.10	0.932	70.25	0.354	412.63	4
Extreme gradient boosting	Market, technical, and financial variables	398.26	642.37	6.18	0.948	72.61	0.315	389.74	3

Long short-term memory neural network	Sequential lagged price and technical indicators	421.91	681.74	6.51	0.942	73.08	0.329	1,846.28	4
Weighted averaging ensemble	Combined predictions of random forest, XGBoost, gradient boosting, and LSTM	364.20	596.58	5.71	0.956	74.82	0.294	2,184.36	2
Stacking ensemble model	Meta-learning based on XGBoost, random forest, gradient boosting, SVR, and LSTM outputs	331.44	548.63	5.08	0.964	76.39	0.271	2,376.92	1

The comparative model results in Table 3 demonstrate that all nonlinear machine learning models outperformed the baseline linear models. The multiple linear regression model produced the weakest predictive performance, with a mean absolute error of 842.60, root mean square error of 1,328.42, mean absolute percentage error of 12.84%, and coefficient of determination of 0.781. Although ridge regression slightly improved prediction accuracy compared with ordinary linear regression, its performance remained weaker than nonlinear algorithms. The decision tree model improved the results by capturing nonlinear splits in the data, but its error level remained relatively high compared with ensemble-based methods. Random forest, gradient boosting, and extreme gradient boosting showed substantially stronger performance, confirming that tree-based ensemble algorithms were able to identify nonlinear relationships, variable interactions, and heterogeneous patterns across companies and industries. Among the individual models, extreme gradient boosting achieved the best overall result, with an MAE of 398.26, RMSE of 642.37, MAPE of 6.18%, and R^2 of 0.948. The LSTM model also performed strongly, especially in directional accuracy, showing that sequential price patterns and lagged technical indicators were useful for learning temporal dependencies. However, the best performance was obtained by the stacking ensemble model. This integrated model reduced the RMSE from 642.37 in the best individual model to 548.63 and reduced the MAPE from 6.18% to 5.08%. It also achieved the highest coefficient of determination, 0.964, and the highest directional accuracy, 76.39%. These findings indicate that integrating several machine learning models produced a more accurate and stable stock price prediction system than relying on a single algorithm.

Table 4. Prediction performance of the final stacking ensemble model by industry group

Industry Group	Number of Companies	Test Observations	Mean Actual Adjusted Price	Mean Predicted Adjusted Price	MAE	RMSE	MAPE (%)	R^2	Directional Accuracy (%)
Chemical and petrochemical products	24	8,205	13,842.71	13,796.54	349.82	571.43	4.76	0.967	77.84
Basic metals	18	6,084	14,376.29	14,291.63	362.74	588.26	4.91	0.964	76.95
Food products	15	5,120	10,284.65	10,318.47	317.51	526.18	5.24	0.961	75.62
Cement and construction materials	13	4,362	8,963.42	8,928.36	286.43	491.74	5.37	0.958	74.88
Pharmaceuticals	12	4,107	15,708.39	15,641.92	374.69	602.55	4.83	0.966	78.16
Automobile and parts manufacturing	11	3,714	7,846.58	7,782.44	341.82	579.60	6.18	0.951	72.49
Machinery and equipment	10	3,395	9,624.36	9,681.70	329.64	540.31	5.76	0.955	74.23

Mining and mineral products	9	3,014	16,294.82	16,188.63	386.27	621.44	4.68	0.969	78.71
Non-metallic minerals	8	2,706	8,415.73	8,461.95	309.85	518.37	5.92	0.954	73.86
Computer and electronic products	7	2,368	11,736.28	11,802.51	348.91	563.72	5.41	0.959	75.04
Textiles	5	1,710	6,928.17	6,861.39	372.56	619.28	7.14	0.943	70.96
Other manufacturing and service-related industries	16	5,448	9,982.64	10,031.26	336.72	552.81	5.59	0.957	74.77
Total sample	148	50,233	11,936.28	11,914.73	331.44	548.63	5.08	0.964	76.39

The industry-level results in Table 4 show that the final stacking ensemble model performed consistently across different sectors of the Tehran Stock Exchange. The highest predictive accuracy was observed in mining and mineral products, pharmaceuticals, chemical and petrochemical products, and basic metals. These industries had relatively lower MAPE values and higher coefficients of determination, indicating that their stock price behavior was more learnable from historical market, technical, and financial variables. This may be due to the stronger relationship between firm fundamentals, commodity-related trends, market capitalization, and stock price movements in these sectors. The model also performed acceptably in food products, cement and construction materials, machinery and equipment, computer and electronic products, and non-metallic minerals, with R^2 values above 0.95. The weakest performance was observed in the textile sector, where the MAPE reached 7.14% and the directional accuracy decreased to 70.96%. This result suggests that smaller sectors with fewer companies and lower trading liquidity may contain more irregular price movements and less stable predictive patterns. The automobile and parts manufacturing sector also showed a relatively higher prediction error than most other industries, which may reflect stronger sensitivity to policy changes, production uncertainty, and market speculation. Nevertheless, the total-sample results confirmed the overall strength of the integrated model, as the mean predicted adjusted price was very close to the mean actual adjusted price, and the model achieved an overall MAPE of 5.08%, R^2 of 0.964, and directional accuracy of 76.39%.

Table 5. Relative importance of predictors in the final integrated stock price prediction model

Rank	Predictor	Predictor Category	Relative Importance (%)	Cumulative Importance (%)	Direction of Relationship with Predicted Price	Interpretation
1	One-day lagged adjusted price	Historical price	15.84	15.84	Positive	The previous adjusted price was the strongest predictor, showing that short-term price continuity was central to forecasting.
2	Twenty-day moving average	Technical trend	12.67	28.51	Positive	The medium-term trend strongly explained the next-period price level.
3	Five-day moving average	Technical trend	10.91	39.42	Positive	Recent price trend contained strong short-term predictive information.
4	Market capitalization	Firm size and market value	8.74	48.16	Positive	Larger firms generally showed more stable and predictable price behavior.
5	Trading value	Liquidity	7.26	55.42	Positive	Higher transaction value improved the model's ability to predict price movements.

6	Earnings per share	Financial performance	6.43	61.85	Positive	More profitable companies tended to have higher predicted stock prices.
7	MACD	Technical momentum	5.91	67.76	Positive	Momentum-based price changes contributed meaningfully to prediction.
8	Relative strength index	Technical momentum	5.18	72.94	Positive	RSI helped the model identify overbought, oversold, and momentum conditions.
9	Trading volume	Liquidity	4.76	77.70	Positive	Volume captured market attention and trading intensity.
10	Twenty-day rolling volatility	Risk indicator	4.29	81.99	Negative	Higher volatility was associated with greater prediction error and lower price stability.
11	Return on equity	Financial performance	3.92	85.91	Positive	Firms with stronger shareholder returns tended to have higher price expectations.
12	Price-to-earnings ratio	Valuation	3.44	89.35	Positive	Valuation level provided additional information about investor expectations.
13	Debt-to-assets ratio	Financial leverage	2.97	92.32	Negative	Greater leverage generally reduced predicted price stability.
14	Average true range	Risk and price range	2.51	94.83	Negative	Wide daily price ranges increased uncertainty in prediction.
15	Number of transactions	Liquidity and market activity	1.84	96.67	Positive	Transaction frequency added limited but useful liquidity information.
16	Current ratio	Liquidity ratio	1.28	97.95	Positive	Better short-term solvency had a weak positive contribution.
17	Asset turnover ratio	Operating efficiency	1.07	99.02	Positive	Operating efficiency provided limited incremental predictive value.
18	Gross profit margin	Profitability	0.98	100.00	Positive	Profit margin had the smallest contribution after controlling for stronger market and technical variables.

The feature-importance results in Table 5 show that historical and technical price variables played the dominant role in stock price prediction. The one-day lagged adjusted price had the highest relative importance, accounting for 15.84% of the total predictive contribution. This finding indicates that the stock prices of Tehran Stock Exchange companies had a strong short-term continuity pattern, meaning that the most recent price level was highly informative for predicting the next-period price. The twenty-day and five-day moving averages ranked second and third, confirming the importance of both medium-term and short-term price trends. Market capitalization, trading value, and trading volume were also among the most important predictors, showing that firm size and market liquidity were essential for improving prediction accuracy. Among the financial variables, earnings per share and return on equity were the most influential, suggesting that profitability information added meaningful explanatory power beyond technical indicators. The negative direction of volatility, average true range, and debt-to-assets ratio shows that risk and financial leverage weakened price stability and increased prediction difficulty. Overall, the results indicate that the most accurate prediction model was not based on a single type of information, but rather on the integration of historical price behavior, technical trend indicators, liquidity measures, and firm-level financial characteristics.

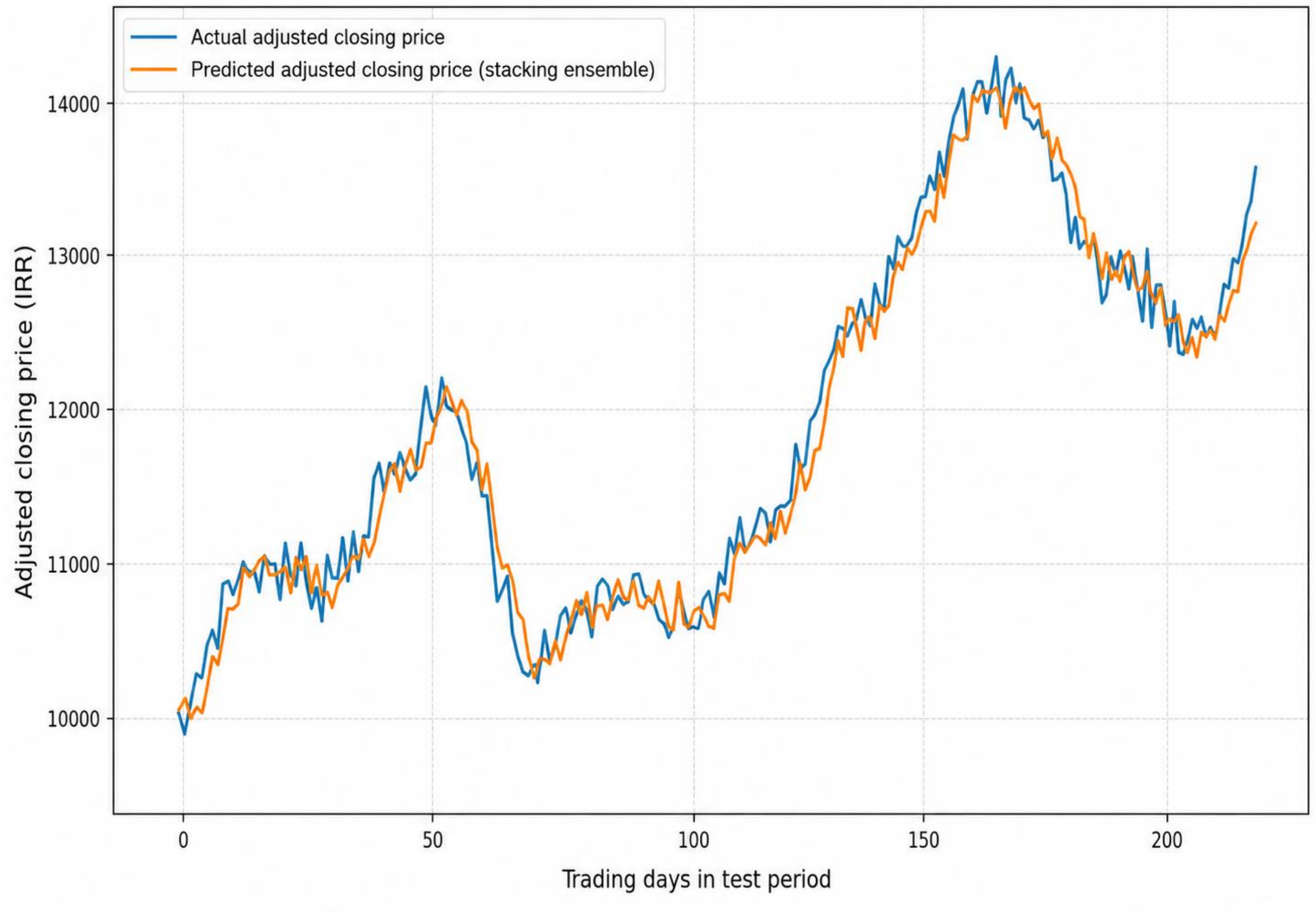


Figure 1. Actual and predicted adjusted closing stock prices based on the final stacking ensemble model

The visual comparison of actual and predicted adjusted closing prices showed that the final stacking ensemble model closely followed the real movement of stock prices in the test period. The predicted series captured both upward and downward movements and showed only limited divergence from the actual price series during highly volatile intervals. The figure indicated that prediction errors were generally smaller during stable market phases and became more visible during sudden price jumps, sharp corrections, and periods of abnormal trading pressure. This pattern is consistent with the numerical findings, where the model achieved strong overall accuracy but still showed some sensitivity to extreme fluctuations. The visual evidence therefore supported the quantitative results and confirmed that the integrated machine learning model was able to reproduce the general direction, level, and trend of stock prices with acceptable precision. In addition, the closeness of the predicted and actual lines demonstrated that combining individual algorithms through stacking reduced the instability of single-model predictions and produced a smoother, more reliable forecasting output.

4. Discussion and Conclusion

The present study was conducted with the aim of designing and evaluating an integrated model for predicting the stock prices of companies listed on the Tehran Stock Exchange based on the combination of machine learning algorithms and market, technical, and financial predictors. The findings showed that the final dataset, consisting of 148 listed companies and 301,284 firm-day observations, had sufficient variability in price, trading, technical, and accounting variables for predictive modeling. The descriptive results indicated substantial heterogeneity among the sampled companies in terms of adjusted closing price, trading volume, trading value, market capitalization, profitability, leverage, and technical indicators. This heterogeneity is important because stock price prediction becomes meaningful when the model is exposed to different price levels, firm sizes, liquidity conditions, and sectoral behaviors. The broad dispersion observed in market and financial variables suggests that the Tehran Stock Exchange contains diverse informational patterns that cannot be adequately modeled through simple linear assumptions. This finding is consistent with the general view that stock markets are nonlinear and multidimensional systems in which price behavior is shaped by the simultaneous interaction of technical, financial, and behavioral factors [1-3].

The correlation analysis showed that the adjusted closing stock price had the strongest relationship with historical price-based variables, especially the twenty-day moving average and lagged adjusted price. This result indicates that short-term and medium-term price continuity played an important role in the prediction of future prices. In financial time-series modeling, this finding is theoretically meaningful because stock prices often contain trend persistence, momentum behavior, and delayed adjustment to information. Although market efficiency theories suggest that price changes should be difficult to predict, empirical studies using machine learning have shown that historical price patterns can still contain useful predictive information, especially when the market has informational frictions, liquidity constraints, or behavioral inefficiencies. The positive association between adjusted price and technical indicators such as moving averages, MACD, and relative strength index supports previous findings that technical variables can improve predictive performance by transforming raw price data into structured signals that represent trend, momentum, and market direction [4, 5, 7, 8].

Another important result was the positive relationship between adjusted stock price and market capitalization, trading value, earnings per share, and profitability indicators. This suggests that the prediction of stock prices cannot be restricted to technical variables alone. Larger firms, firms with higher transaction value, and firms with stronger earnings performance generally produced more stable and informative price patterns. This finding is aligned with the argument that firm-level financial fundamentals remain relevant even in machine learning-based prediction systems. While technical indicators may capture short-term price behavior, accounting and financial variables can reflect the underlying economic quality, profitability, financial health, and valuation position of listed companies. The inclusion of earnings per share, return on assets, return on equity, price-to-earnings ratio, and leverage indicators therefore helped enrich the predictive feature space. This result supports previous studies emphasizing that machine learning models become more informative when they integrate historical market data with financial and firm-level predictors rather than relying exclusively on price sequences [2, 3, 9].

The results of the comparative model evaluation showed that nonlinear machine learning algorithms outperformed traditional linear models. Multiple linear regression and ridge regression had the weakest predictive performance among the tested models, while random forest, gradient boosting, extreme gradient boosting, and LSTM produced substantially lower prediction errors and higher explanatory power. This finding confirms that stock price behavior in the Tehran Stock Exchange is not adequately explained by a purely linear functional structure. The superior performance of nonlinear models indicates that the relationship between predictors and future stock prices includes nonlinear interactions, threshold effects, and time-dependent patterns. This is consistent with prior studies showing that machine learning algorithms are more effective than conventional models in financial forecasting because they are capable of learning complex relationships from high-dimensional market data [5, 6, 18].

Among the individual models, extreme gradient boosting achieved the strongest performance, followed closely by the LSTM model. The strong performance of extreme gradient boosting can be explained by its ability to combine multiple weak learners, reduce prediction error sequentially, handle nonlinear feature interactions, and provide robust performance on structured tabular data. In financial datasets that include price indicators, trading variables, and financial ratios, boosting algorithms are often highly effective because they can model complex interactions without requiring strict distributional assumptions. This finding is consistent with studies that have highlighted the predictive strength of optimized and hybrid machine learning algorithms in stock price forecasting [11, 14]. The strong performance of the LSTM model also indicates that sequential dependencies were present in the stock price series. Since LSTM networks are designed to learn temporal patterns and retain relevant information across time steps, their performance supports the view that historical market sequences contain meaningful information for next-period price prediction [6, 19].

The most important finding of the study was that the stacking ensemble model outperformed all individual algorithms. The final integrated model achieved the lowest mean absolute error, root mean square error, and mean absolute percentage error, while also producing the highest coefficient of determination and directional accuracy. Specifically, the stacking model reduced the prediction error compared with the best single model and improved the ability to predict the direction of stock price movement. This result provides strong evidence that integrating machine learning models is more effective than selecting a single algorithm for stock price prediction. The superiority of the stacking approach can be explained by the complementary strengths of the base learners. Tree-based models captured nonlinear cross-sectional relationships among technical, liquidity, and financial variables; boosting models improved error correction; support vector regression contributed nonlinear mapping; and the LSTM model captured sequential dependencies in the price series. The meta-learner then combined these outputs and produced a more stable final prediction. This finding is aligned with previous studies suggesting that hybrid and ensemble models are particularly suitable for financial markets because market data are noisy, nonlinear, and influenced by multiple sources of uncertainty [3, 13-15].

The industry-level analysis showed that the final stacking ensemble model performed consistently across most industry groups, although prediction accuracy varied by sector. The strongest performance was observed in mining and mineral products, pharmaceuticals, chemical and petrochemical products, and basic metals. These sectors had relatively lower prediction errors and higher coefficients of determination. One possible explanation is that these industries are more strongly linked to observable fundamentals, commodity-related trends, market capitalization, and stable sectoral information, allowing the model to extract more reliable predictive patterns. In contrast, the textile sector and automobile and parts manufacturing showed relatively weaker prediction accuracy. These sectors may be more affected by lower liquidity, irregular trading patterns, policy uncertainty, supply-chain instability, and speculative behavior. This sectoral variation is consistent with previous evidence that the performance of stock prediction models may differ across market segments, industries, and institutional contexts [9, 10, 20].

The feature-importance analysis further clarified the internal logic of the final model. The most influential predictors were one-day lagged adjusted price, twenty-day moving average, five-day moving average, market capitalization, trading value, earnings per share, MACD, relative strength index, trading volume, and rolling volatility. This pattern shows that the model relied primarily on historical price continuity and trend indicators, while also incorporating liquidity, firm size, profitability, and risk information. The dominance of lagged price and moving-average variables confirms that price history is a central input in short-term forecasting. At the same time, the importance of trading value and volume suggests that liquidity and market participation affect price predictability. The contribution of earnings per share and profitability indicators shows that firm fundamentals remained relevant even in a machine learning framework. These results are consistent with research emphasizing that the most accurate stock prediction models are those that combine technical indicators, trading information, and firm-level variables in a unified analytical framework [2-4, 7].

The negative role of volatility and leverage in the model is also noteworthy. Rolling volatility, average true range, and debt-to-assets ratio contributed negatively to predicted price stability. This means that higher risk and higher leverage were associated with more unstable price behavior and greater difficulty in prediction. In financial markets, volatility reflects uncertainty, abrupt price movement, and changing expectations; therefore, periods of high volatility often reduce the accuracy of prediction models. Similarly, highly leveraged firms may be more vulnerable to financial pressure, interest-rate changes, and market pessimism. The present findings support the view that volatility is not only a risk indicator but also an important determinant of predictability. This is consistent with recent machine learning-based research showing that volatility modeling is essential for improving the simulation and forecasting of stock market behavior [19, 20].

The findings of this study also contribute to the broader literature on artificial intelligence in financial markets. Previous bibliometric and systematic reviews have shown that artificial intelligence and machine learning are increasingly used in finance, particularly in prediction, risk assessment, portfolio management, and algorithmic trading. The present study extends this literature by showing that an integrated machine learning model can be successfully applied to the Tehran Stock Exchange, an emerging market context characterized by volatility, sectoral heterogeneity, and complex market behavior. The results confirm that machine learning is not only relevant for highly developed markets but can also offer valuable predictive advantages in markets where traditional models may be less effective due to instability and nonlinear dynamics [1, 18, 21].

Although the present study focused on numerical market, technical, and financial predictors, its findings are also compatible with the broader movement toward richer and more heterogeneous financial prediction systems. Recent studies have shown that financial news, text mining, and sentiment analysis can provide additional predictive value, especially when market movements are influenced by information shocks and investor expectations [16, 17]. The relatively strong performance of the integrated model in the present study suggests that even without direct textual sentiment data, combining multiple numerical information sources can produce accurate predictions. However, the remaining prediction error, particularly during volatile periods and in less liquid industries, indicates that future models may benefit from adding unstructured information such as news sentiment, macroeconomic announcements, social media signals, and policy-related variables. Such extensions may further improve predictive accuracy by incorporating dimensions of market psychology and information diffusion that are not fully captured by price and financial variables alone.

Overall, the results support the methodological value of combining multiple machine learning algorithms in stock price prediction. The study showed that no single model was able to fully capture the complexity of stock price behavior, but an integrated stacking model provided the most accurate and stable results. This finding is consistent with the logic of ensemble learning, according to which prediction improves when different models learn different aspects of the data and their outputs are combined through an optimized structure. In practical terms, the results suggest that investors and analysts should avoid relying solely on simple trend indicators or single-algorithm forecasts. Instead, stock price prediction should be approached as a multi-source and multi-model problem in which technical, financial, liquidity, and risk indicators are simultaneously processed through flexible computational methods. This interpretation is aligned with recent advances in machine learning and deep learning for stock market forecasting, which emphasize model integration, careful validation, and context-specific evaluation [3, 5, 6, 12].

The present study had several limitations that should be considered when interpreting the findings. First, the sample was limited to 148 companies listed on the Tehran Stock Exchange, and the results may not be directly generalizable to all listed firms, smaller over-the-counter companies, or other capital markets. Second, although the dataset included a large number of firm-day observations, the predictive variables were limited to historical market data, technical indicators, trading variables, and selected financial ratios. Macroeconomic indicators, political events, exchange-rate shocks, commodity prices, investor sentiment, and textual news data were not included in the final model. Third, the model was evaluated using historical data, and although chronological data splitting was used to reduce data leakage, future market conditions may differ from the patterns observed in the study period. Fourth, the study focused on prediction accuracy and did not directly evaluate trading profitability, transaction costs, portfolio performance, or risk-adjusted investment returns. Finally, machine learning models, especially integrated and deep learning models, may have lower interpretability than traditional statistical models, even when feature-importance analysis is used.

Future research should expand the present study in several directions. Future studies can examine whether the performance of integrated machine learning models improves when macroeconomic variables, exchange-rate indicators, inflation expectations, commodity prices, interest-rate information, and policy uncertainty measures are added to the feature set. Researchers can also integrate financial news, corporate announcements, social media sentiment, and natural language processing-based variables to examine whether textual information improves prediction accuracy in the Tehran Stock Exchange. In addition, future studies should compare different ensemble strategies, including bagging, boosting, weighted averaging, blending, stacking, and deep hybrid architectures, in order to identify the most stable combination method. It is also recommended that future research evaluate the models across different prediction horizons, such as one-day, one-week, one-month, and quarterly forecasts. Finally, future studies can move beyond statistical accuracy and examine whether the predicted prices can be converted into profitable and risk-controlled trading strategies after considering transaction costs, liquidity constraints, and portfolio risk.

From a practical perspective, the findings suggest that investors, analysts, financial institutions, and portfolio managers can benefit from using integrated machine learning models as decision-support tools in stock market analysis. The results show that combining technical indicators, liquidity measures, market capitalization, profitability indicators, leverage measures, and historical price information can produce more accurate predictions than relying on a single class of variables. However, such models should not be used as automatic substitutes for professional judgment. Instead, they should be incorporated into broader investment decision-making frameworks that also consider market conditions, sectoral trends, macroeconomic developments, and risk tolerance. Brokerage firms and investment companies can use similar models to screen stocks, identify probable price trends, monitor prediction errors, and improve timing decisions. Regulatory and market institutions can also benefit from predictive analytics by monitoring abnormal price movements and periods of heightened volatility. For best practice, model outputs should be updated regularly, tested on new data, and interpreted cautiously, especially during unstable market periods.

Authors' Contributions

Authors equally contributed to this article.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

[1] S. Ahmed, M. M. Alshater, A. E. Ammari, and H. Hammami, "Artificial intelligence and machine learning in finance: A bibliometric review," *Research in International Business and Finance,* vol. 61, p. 101646, 2022/10/01/ 2022, doi: 10.1016/j.ribaf.2022.101646.

[2] D. Sheth and M. Shah, "Predicting Stock Market Using Machine Learning: Best and Accurate Way to Know Future Stock Prices," (in English), *International Journal of System Assurance Engineering and Management,* vol. 14, no. 1, pp. 1-18, 2023, doi: 10.1007/s13198-022-01811-1.

[3] G. Sonkavde, D. S. Dharrao, A. M. Bongale, S. T. Deokate, D. Doreswamy, and S. K. Bhat, "Forecasting Stock Market Prices Using Machine Learning and Deep Learning Models: A Systematic Review, Performance Analysis and Discussion of Implications," *International Journal of Financial Studies,* vol. 11, no. 3, p. 94, 2023, doi: 10.3390/ijfs11030094.

[4] L. Liu, B. Peng, and J. Yu, "Stock Price Prediction Based on Machine Learning," pp. 1277-1282, 2022, doi: 10.2991/978-94-6463-036-7_189.

[5] J. Chen, Y. Wen, Y. A. Nanehkaran, M. D. Suzauddola, W. Chen, and D. Zhang, "Machine learning techniques for stock price prediction and graphic signal recognition," *Engineering Applications of Artificial Intelligence,* vol. 121, p. 106038, 2023/05/01/ 2023. [Online]. Available: https://doi.org/10.1016/j.engappai.2023.106038.

[6] V. Chang, Q. A. Xu, A. Chidozie, and H. Wang, "Predicting Economic Trends and Stock Market Prices with Deep Learning and Advanced Machine Learning Techniques," *Electronics,* vol. 13, no. 17, p. 3396, 2024, doi: 10.3390/electronics13173396.

[7] M. Vijh, D. Chandola, V. A. Tikkiwal, and A. Kumar, "Stock Closing Price Prediction Using Machine Learning Techniques," *Procedia Computer Science,* vol. 167, pp. 599-606, 2020, doi: 10.1016/j.procs.2020.03.326.

[8] M. Yang, "Predicting the Direction of Stock Price Movement With Machine Learning Algorithms," *Advances in Economics Management and Political Sciences,* vol. 52, no. 1, pp. 283-291, 2023, doi: 10.54254/2754-1169/52/20230758.

[9] W. Liu, Y. Suzuki, and S. Du, "Forecasting the Stock Price of Listed Innovative SMEs Using Machine Learning Methods Based on Bayesian optimization: Evidence from China," *Computational Economics,* vol. 63, no. 5, pp. 2035-2068, 2024, doi: 10.1007/s10614-023-10393-4.

[10] B. Egüz, F. E. Çorbacı, and T. Kaya, "Stock Price Prediction of Turkish Banks Using Machine Learning Methods," pp. 222-229, 2021, doi: 10.1007/978-3-030-85577-2_26.

[11] S. Das, T. P. Sahu, and R. R. Janghel, "Effective forecasting of stock market price by using extreme learning machine optimized by PSO-based group oriented crow search algorithm," *Neural Computing & Applications,* vol. 34, pp. 555-591, 2022, doi: 10.1007/s00521-021-06403-x.

[12] W. Yao, Y. Zhang, S. Chang, J. Li, Q. Zhao, and F. Ge, "Stock price analysis and forecasting based on machine learning," ed, 2022.

[13] I. Behravan and S. M. Razavi, "Stock price prediction using machine learning and swarm intelligence," *Journal of electrical and computer engineering innovations (JECEI),* vol. 8, no. 1, pp. 31-40, 2020. [Online]. Available: https://jecei.sru.ac.ir/article_1421_6d94cea044bc908eba67a7ad40fe184e.pdf.

[14] A. Khosravi and S. M. J. Amirhosseini, "Providing a stock price prediction method using improved hybrid machine learning algorithms," *Journal of Computing and Distributed Systems,* vol. 6, no. 1, pp. 74-88, 2023.

[15] R. Chowdhury, M. Mahdy, T. Nourin Alam, and G. Al Guaderi, "Predicting the Stock Price Of Frontier Markets Using Modified BlackScholes Option Pricing Model and Machine Learning," *Physica A: Statistical Mechanics and its Applications,* 2020, doi: 10.1016/j.physa.2020.124444.
[16] M. N. Ashtiani and B. Raahemi, "News-based intelligent prediction of financial markets using text mining and machine learning: A systematic literature review," *Expert Systems with Applications,* vol. 217, p. 119509, 2023/05/01/ 2023. [Online]. Available: https://doi.org/10.1016/j.eswa.2023.119509.
[17] J. Maqbool, P. Aggarwal, R. Kaur, A. Mittal, and I. A. Ganaie, "Stock Prediction by Integrating Sentiment Scores of Financial News and MLP-Regressor: A Machine Learning Approach," *Procedia Computer Science,* vol. 218, pp. 1067-1078, 2023/01/01/ 2023. [Online]. Available: https://doi.org/10.1016/j.procs.2023.01.086.
[18] S. Z. Shaikh, K. R. Khan, F. K. Sherwani, and M. Khan, "Smart Trading: Unlocking Artificial Intelligence in Stock Market," *The Business & Management Review,* vol. 15, no. 03, 2025, doi: 10.24052/bmr/v15nu03/art-17.
[19] J. Mansilla-Lopez, "Factors, Forecasts, and Simulations of Volatility in the Stock Market Using Machine Learning," *Journal of Risk and Financial Management,* 2025, doi: 10.3390/jrfm18050227.
[20] F. Ghallabi, B. Souissi, A. M. Du, and S. Ali, "ESG stock markets and clean energy prices prediction: Insights from advanced machine learning," *International Review of Financial Analysis,* vol. 97, no. 1, p. 103889, 2025, doi: 10.1016/j.irfa.2024.103889.
[21] A. Taftiyan, M. Mirzababaei Mazraeh Akhond, and A. S. Oliaei Bajgani, "A systematic review of the machine learning approach in stock price prediction in the form of a time series model," in *The 2nd National Conference on Modern Applied Research in Accounting*, Damghan, 2023. [Online]. Available: https://civilica.com/doc/1964451.