iBusiness, 2026, 18(3), 159-172
https://www.scirp.org/journal/ib
ISSN Online: 2150-4083
ISSN Print: 2150-4075

Predicting the Nikkei 225 Using Machine Learning: An Empirical Comparison

Chikashi Tsuji

Graduate School of Economics, Chuo University, Tokyo, Japan
Email: mail_sec_low@minos.ocn.ne.jp

How to cite this paper: Tsuji, C. (2026). Predicting the Nikkei 225 Using Machine Learning: An Empirical Comparison. *iBusiness, 18,* 159-172.
https://doi.org/10.4236/ib.2026.183009

Received: June 26, 2026
Accepted: September 17, 2026
Published: September 20, 2026

Copyright © 2026 by author(s) and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/
Open Access

Abstract

This study compares the predictive performance of multiple machine learning models for forecasting the closing price of the Nikkei 225 on the forecast day using daily data. Five models—Random Forest Regression, Decision Tree Regression, XGBoost Regression, LightGBM Regression, and K-Nearest Neighbors Regression—are employed. In addition to baseline models that use lags of the closing price as explanatory variables, three additional testing models are constructed by adding lags of the opening, high, and low prices individually to the explanatory variables of the baseline models. Furthermore, six different forecast target periods, ranging from five to ten years, are considered, and predictive performance is evaluated using the R-squared value, mean absolute error, and root mean squared error. The results show that Random Forest Regression and Decision Tree Regression exhibit relatively high predictive performance, followed by K-Nearest Neighbors Regression. In addition, the results indicate that adding lags of the opening, high, and low prices to the explanatory variables slightly improves predictive performance in some cases. Furthermore, predictive performance tends to improve as the forecast target period becomes longer. These findings suggest that, in forecasting the closing price of the Nikkei 225, predictive performance may be influenced not only by the type of model but also by the composition of explanatory variables and the forecast target period.

Keywords

Nikkei 225, Machine Learning, Stock Price Prediction, Time-Series Forecasting, Comparison of Predictive Performance

1. Introduction

In recent years, artificial intelligence (AI) has attracted growing attention as a promising approach to time-series forecasting, with an increasing number of studies exploring and discussing its applications across various domains (e.g., Chen, 2023; Vishwas et al., 2025; Tsuji, 2026). Stock price prediction is one of the important research topics in financial markets, and a wide range of forecasting methods have been examined, from statistical approaches to machine learning and, more recently, deep learning. Because stock prices and stock returns exhibit complex nonlinearities and temporal dependencies, recent studies have employed machine learning models such as Random Forest, Decision Tree, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM), as well as deep learning models such as Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN), with the aim of capturing these characteristics (e.g., Jiang et al., 2019; Tsuji, 2022, 2025).

Previous studies have examined the effectiveness of deep learning models, including LSTM, as well as the possibility of utilizing information other than prices, such as investor sentiment on social media (e.g., Liu et al., 2023). In addition, research has progressed on addressing the non-stationarity of financial time series and on the generalization performance of models across different markets (e.g., Hong et al., 2023).

On the other hand, when using time-series information from stock prices themselves, there remains room for further investigation into the extent to which different machine learning models exhibit different predictive performance and how predictive performance changes depending on the sample period and the specification of explanatory variables.

This study therefore applies five machine learning models—Random Forest Regression, Decision Tree Regression, XGBoost Regression, LightGBM Regression, and K-Nearest Neighbors Regression—to daily Nikkei 225 data to forecast the closing price on the forecast day and compare their predictive performance. As shown in **Figure 1**, the Nikkei 225 has exhibited a pronounced upward trend in recent years, making it a useful benchmark for evaluating and comparing different forecasting models. In addition, this study uses a model that includes only lags of the closing price as explanatory variables, which serves as the baseline model, and compares its predictive performance with that of models that additionally include lags of the opening, high, and low prices.

Furthermore, six different forecast target periods, ranging from five to ten years, are considered to examine the effects of differences in the forecast target period on the prediction results. The objective of this study is to empirically clarify, through these comparisons, the relative predictive performance of the respective machine learning models in forecasting the closing price of the Nikkei 225 and the differences in predictive performance resulting from the specification of explanatory variables and forecast target periods.

The results of the analysis show broadly similar tendencies across different forecast target periods for each of the three evaluation measures, the R-squared value (R2), mean absolute error (MAE), and root mean squared error (RMSE). In particular, Random Forest Regression and Decision Tree Regression exhibit relatively high predictive performance, followed by K-Nearest Neighbors Regression.

In addition, adding lags of the opening, high, and low prices to the lags of the closing price as explanatory variables is found to slightly improve predictive performance in some cases. Furthermore, comparison across different forecast target periods indicates that predictive performance tends to improve as the period is extended from five to ten years. These findings are significant contributions of this study.

The remainder of this paper is organized as follows. Section 2 reviews the previous literature, and Section 3 describes the data and models. Section 4 presents the empirical results, followed by the conclusion in Section 5.

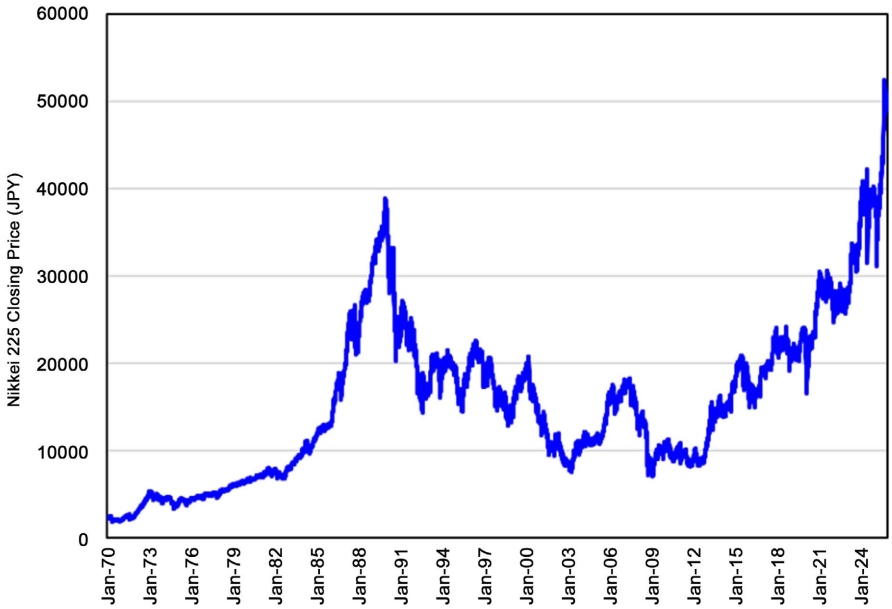

Figure 1. Evolution of the Nikkei 225 closing prices: January 5, 1970, to December 30, 2025.

2. Previous Studies

In recent years, research on stock price prediction has progressed through the use of machine learning and deep learning in addition to conventional statistical methods. An increasing number of studies have employed neural networks such as LSTM and CNN to capture the complex nonlinearities and time-series characteristics of stock prices and stock returns. This section reviews previous studies on stock price prediction using AI in chronological order.

Roondiwala et al. (2017) applied LSTM to stock index prediction and demonstrated the usefulness of Recurrent Neural Networks (RNNs) and LSTM for time-series data such as stock prices. This study is important in that it indicated the potential effectiveness of LSTM, which can learn long-term dependencies, for predicting future price movements using historical stock price information.

Subsequently, Jiang et al. (2019) presented an approach for classifying stock price increases and decreases by combining large-scale data processing using

Spark with machine learning models such as Random Forest and Decision Tree. This demonstrated the potential for applying not only deep learning but also conventional machine learning models to large-scale financial data.

Moreover, Moghar and Hamiche (2020) examined stock time-series forecasting using LSTM and found that the number of training iterations and the period of data used affected predictive performance. These studies indicate that, in stock price prediction, not only the choice of models that appropriately capture the characteristics of time-series data but also training conditions and data handling have important effects on predictive accuracy.

Later, Mokhtari et al. (2021) examined the potential of utilizing AI and machine learning in stock market analysis by incorporating sentiment analysis of posts on Twitter in addition to technical analysis based on historical stock price data. In this way, the information used for stock price prediction has expanded from market data such as prices and trading volume to unstructured data containing investors' opinions and emotions.

Furthermore, Liu et al. (2023) applied Natural Language Processing (NLP) to a large volume of social media data and quantified investor sentiment to analyze the relationship between sentiment and stock prices. Their results showed a significant positive synergy between investor sentiment and stock prices, while also indicating that this relationship could reverse or disappear depending on the period. In addition, the study showed that investor sentiment on social media reflects expectations regarding future market trends, suggesting the potential usefulness of sentiment information for stock price prediction.

Moreover, Hong et al. (2023) focused on the non-stationarity of financial time series and proposed a forecasting model for non-stationary financial time series using meta-learning. In this model, CNN is used as the predictor and LSTM as the meta-learner, and long-term non-stationary time series are divided into multiple shorter subsequences for forecasting. The experiments showed that the proposed model achieved better predictive performance than conventional CNN and autoregressive (AR) models, demonstrating the effectiveness of an adaptive forecasting approach for non-stationary financial data.

Subsequently, Ghosh et al. (2024) compared multiple neural network architectures, including Multilayer Perceptron (MLP), RNN, LSTM, and CNN, examined their respective stock price forecasting performance, and analyzed the applicability of forecasting models to different markets. This indicates that, in addition to accuracy for a particular market or dataset, the generalization performance of models is also an important research issue. Furthermore, Abir et al. (2025) applied CNN, LSTM, and hybrid models combining these approaches to financial markets in the BRICS countries, thereby extending the application of deep learning models to multiple countries and markets.

These studies indicate that research on stock price prediction has evolved from forecasting based on a single time series toward forecasting in non-stationary market environments and different markets, as well as forecasting using combinations of multiple deep learning models. However, relatively few studies have carefully examined well-established models by changing settings such as forecast target periods and explanatory variables. Therefore, we believe that research such as the present study fills the research gap left by existing studies and thus is highly meaningful.

3. Data and Models

This section describes the data used in the analysis and the forecasting models. This study employs the five machine learning models, Random Forest Regression, Decision Tree Regression, XGBoost Regression, LightGBM Regression, and K-Nearest Neighbors Regression, for model comparisons.

The analysis uses daily data for the Nikkei 225. Specifically, the four daily price variables—opening price (Open), high price (High), low price (Low), and closing price (Close)—are used. The overall sample period used in the analysis is from January 5, 1970, to December 30, 2025. To clarify the differences by forecasting periods, six forecast target periods are considered: five years from January 4, 2021, to December 30, 2025; six years from January 6, 2020, to December 30, 2025; seven years from January 4, 2019, to December 30, 2025; eight years from January 4, 2018, to December 30, 2025; nine years from January 4, 2017, to December 30, 2025; and ten years from January 4, 2016, to December 30, 2025. Each of these six periods is analyzed separately for the five machine learning models.

Accordingly, in model training and prediction, for each of the six period settings, the observations outside the respective forecast target period within the overall sample period constitute the training period, while the respective forecast target period constitutes the test period.

The prediction target is the closing price (Close) of the Nikkei 225. First, taking into account the temporal relationship with the closing price, a baseline model is specified using lags 1 through 3 of the closing price as explanatory variables. Moreover, three additional six-variable models are constructed by adding lags 0 through 2 of the opening price (Open), lags 1 through 3 of the high price (High), and lags 1 through 3 of the low price (Low), respectively, to the explanatory variables of the baseline model.

These four explanatory-variable specifications are then applied to the five forecasting models described above for comparative analysis. For lag 0 of Open, it is assumed that the opening price of the day is known at the time of prediction. In other words, the analysis assumes that the closing price of the day is predicted once the opening price of that day becomes known.

4. Results

This section compares and examines the results obtained from each forecasting model. The five models—Random Forest Regression, Decision Tree Regression, XGBoost Regression, LightGBM Regression, and K-Nearest Neighbors Regression—are applied to the price series of the Nikkei 225 and compared across six forecast target periods of five, six, seven, eight, nine, and ten years. The results for the predictive accuracy of each period are presented in Tables 1-6, respectively. For all forecasting models, model specifications were determined based solely on data from the training period, and observations from the forecast target period were not used for model fitting.

Each of the six tables presents the results of forecasting the closing price on the forecast day using past 3-day closing prices as explanatory variables in Panel A; the results of forecasting using past 3-day closing prices and same-day and past 2-day opening prices as explanatory variables in Panel B; the results of forecasting using past 3-day closing prices and past 3-day high prices as explanatory variables in Panel C; and the results of forecasting using past 3-day closing prices and past 3-day low prices as explanatory variables in Panel D.

First, across the three evaluation measures—R2, MAE, and RMSE—shown in Tables 1-6, broadly similar tendencies are observed with respect to the predictive accuracy of the daily closing price of the Nikkei 225 across the six forecast target periods. In particular, Random Forest Regression and Decision Tree Regression exhibit relatively high predictive accuracy, followed by K-Nearest Neighbors Regression. In contrast, XGBoost Regression and LightGBM Regression underperform the other three models.

Furthermore, as shown in Tables 1-6, adding lags of Open, High, or Low as explanatory variables to the baseline models, which use only lags of Close as explanatory variables, slightly improves predictive accuracy in some cases. At the same time, predictive accuracy tends to improve as the forecast target period is extended from five to ten years. As the values of R2, MAE, and RMSE in Tables 1-6 indicate, these tendencies remain highly consistent across the six forecast target periods.

Table 1. Results of 5-year forecasts of the Nikkei 225 daily closing prices.

Panel A. Forecasting using Past 3-day Close Prices

Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.81558	1090.59975	2699.75107
Decision Tree Regression	0.81534	1158.90021	2701.51411
XGBoost Regression	0.80763	1145.36361	2757.31333
LightGBM Regression	0.77703	1325.91695	2968.57030
K-Nearest Neighbors Regression	0.81390	1106.53505	2712.00109

Panel B. Forecasting using Past 3-day Close Prices and Same-day and Past 2-day Open Prices

Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.81518	1036.34562	2702.69431

Continued

Decision Tree Regression	0.81772	1080.98340	2684.01892
XGBoost Regression	0.80336	1124.99080	2787.77702
LightGBM Regression	0.77843	1281.49296	2959.21338
K-Nearest Neighbors Regression	0.81506	1066.78127	2703.54347
Panel C. Forecasting using Past 3-day Close Prices and Past 3-day High Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.81481	1094.83399	2705.39066
Decision Tree Regression	0.81566	1139.68042	2699.16368
XGBoost Regression	0.80884	1137.45201	2748.64161
LightGBM Regression	0.78029	1310.28690	2946.79702
K-Nearest Neighbors Regression	0.81357	1111.38107	2714.43792
Panel D. Forecasting using Past 3-day Close Prices and Past 3-day Low Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.81428	1098.52381	2709.27923
Decision Tree Regression	0.81575	1146.51763	2698.49230
XGBoost Regression	0.79907	1187.41044	2817.98405
LightGBM Regression	0.77660	1331.12153	2971.37893
K-Nearest Neighbors Regression	0.81374	1112.94727	2713.16110

Notes. Our full sample period is from January 5, 1970, to December 30, 2025, and the test period is from January 4, 2021, to December 30, 2025. R2 denotes the R-squared value, MAE denotes mean absolute error, and RMSE denotes root mean squared error.

Table 2. Results of 6-year forecasts of the Nikkei 225 daily closing prices.

Panel A. Forecasting using Past 3-day Close Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.87709	953.59991	2469.86650
Decision Tree Regression	0.87660	1015.37587	2474.73816
XGBoost Regression	0.86773	1018.05453	2562.19506
LightGBM Regression	0.85150	1148.36613	2714.82237
K-Nearest Neighbors Regression	0.87586	971.34155	2482.21023
Panel B. Forecasting using Past 3-day Close Prices and Same-day and Past 2-day Open Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.87701	892.75023	2470.65750

Continued

Decision Tree Regression	0.87836	957.24588	2457.11892
XGBoost Regression	0.86824	975.42374	2557.23526
LightGBM Regression	0.85204	1115.51635	2709.90466
K-Nearest Neighbors Regression	0.87681	927.37058	2472.69466
Panel C. Forecasting using Past 3-day Close Prices and Past 3-day High Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.87644	958.91135	2476.38175
Decision Tree Regression	0.87678	1009.06895	2472.95892
XGBoost Regression	0.87026	999.86801	2537.58512
LightGBM Regression	0.85397	1132.79038	2692.17424
K-Nearest Neighbors Regression	0.87566	973.01612	2484.19373
Panel D. Forecasting using Past 3-day Close Prices and Past 3-day Low Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.87614	960.36123	2479.36198
Decision Tree Regression	0.87678	1011.04136	2472.99860
XGBoost Regression	0.86593	1022.77689	2579.53995
LightGBM Regression	0.85138	1147.95385	2715.91793
K-Nearest Neighbors Regression	0.87572	978.16133	2483.63770

Notes. Our full sample period is from January 5, 1970, to December 30, 2025, and the test period is from January 6, 2020, to December 30, 2025. R2 denotes the R-squared value, MAE denotes mean absolute error, and RMSE denotes root mean squared error.

Table 3. Results of 7-year forecasts of the Nikkei 225 daily closing prices.

Panel A. Forecasting using Past 3-day Close Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.90405	844.08330	2290.88025
Decision Tree Regression	0.90394	902.13725	2292.22219
XGBoost Regression	0.89615	903.88740	2383.34565
LightGBM Regression	0.88434	1012.03101	2515.24671
K-Nearest Neighbors Regression	0.90314	857.97267	2301.71220
Panel B. Forecasting using Past 3-day Close Prices and Same-day and Past 2-day Open Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.90413	781.53646	2289.97819

Continued

Decision Tree Regression	0.90498	851.60917	2279.81889
XGBoost Regression	0.89849	845.16203	2356.29358
LightGBM Regression	0.88499	969.74541	2508.12516
K-Nearest Neighbors Regression	0.90393	814.64290	2292.37096
Panel C. Forecasting using Past 3-day Close Prices and Past 3-day High Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.90369	845.48256	2295.23088
Decision Tree Regression	0.90415	891.99150	2289.68040
XGBoost Regression	0.89831	884.69800	2358.42492
LightGBM Regression	0.88627	996.49204	2494.20091
K-Nearest Neighbors Regression	0.90298	858.82289	2303.62538
Panel D. Forecasting using Past 3-day Close Prices and Past 3-day Low Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.90361	845.97694	2296.21487
Decision Tree Regression	0.90369	905.85974	2295.14861
XGBoost Regression	0.89890	883.09870	2351.59734
LightGBM Regression	0.88557	990.26859	2501.81596
K-Nearest Neighbors Regression	0.90303	864.07701	2303.10238

Notes. Our full sample period is from January 5, 1970, to December 30, 2025, and the test period is from January 4, 2019, to December 30, 2025. R2 denotes the R-squared value, MAE denotes mean absolute error, and RMSE denotes root mean squared error.

Table 4. Results of 8-year forecasts of the Nikkei 225 daily closing prices.

Panel A. Forecasting using Past 3-day Close Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.91608	764.73614	2144.23134
Decision Tree Regression	0.91554	827.94652	2151.05330
XGBoost Regression	0.90858	820.99140	2238.01718
LightGBM Regression	0.89983	892.66862	2342.59752
K-Nearest Neighbors Regression	0.91522	778.49900	2155.20648
Panel B. Forecasting using Past 3-day Close Prices and Same-day and Past 2-day Open Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.91623	702.71275	2142.37098

Continued

Decision Tree Regression	0.91627	793.26870	2141.77253
XGBoost Regression	0.91055	763.68722	2213.80165
LightGBM Regression	0.90048	856.20747	2334.99461
K-Nearest Neighbors Regression	0.91599	734.63173	2145.38275
Panel C. Forecasting using Past 3-day Close Prices and Past 3-day High Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.91582	763.99030	2147.49936
Decision Tree Regression	0.91559	818.30517	2150.50203
XGBoost Regression	0.91161	803.66639	2200.58138
LightGBM Regression	0.90138	879.29435	2324.44428
K-Nearest Neighbors Regression	0.91510	780.67689	2156.75410
Panel D. Forecasting using Past 3-day Close Prices and Past 3-day Low Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.91561	767.57620	2150.25795
Decision Tree Regression	0.91549	831.16630	2151.80460
XGBoost Regression	0.91290	800.84445	2184.42320
LightGBM Regression	0.89993	891.74273	2341.41941
K-Nearest Neighbors Regression	0.91513	782.84788	2156.26945

Notes. Our full sample period is from January 5, 1970, to December 30, 2025, and the test period is from January 4, 2018, to December 30, 2025. R2 denotes the R-squared value, MAE denotes mean absolute error, and RMSE denotes root mean squared error.

Table 5. Results of 9-year forecasts of the Nikkei 225 daily closing prices.

Panel A. Forecasting using Past 3-day Close Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.92820	696.13137	2021.66512
Decision Tree Regression	0.92738	768.63354	2033.15572
XGBoost Regression	0.92384	735.73993	2082.04861
LightGBM Regression	0.91517	807.51863	2197.44445
K-Nearest Neighbors Regression	0.92748	708.07887	2031.68122
Panel B. Forecasting using Past 3-day Close Prices and Same-day and Past 2-day Open Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.92834	637.93423	2019.58642

Continued

Decision Tree Regression	0.92838	723.21554	2019.06665
XGBoost Regression	0.92319	700.08930	2090.87839
LightGBM Regression	0.91574	758.77877	2189.93906
K-Nearest Neighbors Regression	0.92818	664.56023	2021.94099
Panel C. Forecasting using Past 3-day Close Prices and Past 3-day High Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.92802	693.53387	2024.12561
Decision Tree Regression	0.92771	752.02628	2028.44275
XGBoost Regression	0.92523	732.34380	2062.94788
LightGBM Regression	0.91608	799.61407	2185.54282
K-Nearest Neighbors Regression	0.92737	710.40317	2033.22502
Panel D. Forecasting using Past 3-day Close Prices and Past 3-day Low Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.92771	698.03494	2028.43264
Decision Tree Regression	0.92778	757.45497	2027.50698
XGBoost Regression	0.92215	745.41652	2104.98440
LightGBM Regression	0.91539	806.53620	2194.49927
K-Nearest Neighbors Regression	0.92739	713.46267	2032.97641

Notes. Our full sample period is from January 5, 1970, to December 30, 2025, and the test period is from January 4, 2017, to December 30, 2025. R2 denotes the R-squared value, MAE denotes mean absolute error, and RMSE denotes root mean squared error.

Table 6. Results of 10-year forecasts of the Nikkei 225 daily closing prices.

Panel A. Forecasting using Past 3-day Close Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.94133	646.52942	1919.57260
Decision Tree Regression	0.94056	719.34984	1931.99684
XGBoost Regression	0.93599	694.16319	2004.87671
LightGBM Regression	0.93060	745.66403	2087.65692
K-Nearest Neighbors Regression	0.94071	658.97482	1929.55019
Panel B. Forecasting using Past 3-day Close Prices and Same-day and Past 2-day Open Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.94149	589.63178	1916.80362

Continued

Decision Tree Regression	0.94151	670.22412	1916.49787
XGBoost Regression	0.93976	629.21047	1944.94714
LightGBM Regression	0.93162	694.33715	2072.24427
K-Nearest Neighbors Regression	0.94132	616.34753	1919.68910
Panel C. Forecasting using Past 3-day Close Prices and Past 3-day High Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.94114	645.79033	1922.53939
Decision Tree Regression	0.94080	707.27427	1928.14411
XGBoost Regression	0.93625	694.92244	2000.89342
LightGBM Regression	0.93128	740.83249	2077.34107
K-Nearest Neighbors Regression	0.94063	661.36332	1930.93430
Panel D. Forecasting using Past 3-day Close Prices and Past 3-day Low Prices			
Model	Forecasting Performance		
	R2	MAE	RMSE
Random Forest Regression	0.94105	647.33798	1924.10647
Decision Tree Regression	0.94083	712.88133	1927.71269
XGBoost Regression	0.93822	678.37639	1969.74899
LightGBM Regression	0.93085	743.93305	2083.95292
K-Nearest Neighbors Regression	0.94064	663.65923	1930.76031

Notes. Our full sample period is from January 5, 1970, to December 30, 2025, and the test period is from January 4, 2016, to December 30, 2025. R2 denotes the R-squared value, MAE denotes mean absolute error, and RMSE denotes root mean squared error.

5. Conclusions

This study used daily data for the Nikkei 225 to forecast its closing price using five machine learning models—Random Forest Regression, Decision Tree Regression, XGBoost Regression, LightGBM Regression, and K-Nearest Neighbors Regression—and compared their predictive performance. The analysis employed baseline models using lags of the closing price as explanatory variables, as well as models that added lags of the opening, high, and low prices, respectively, to the explanatory variables of the baseline models. In addition, comparisons were conducted across six forecast target periods ranging from five to ten years.

The results showed broadly similar tendencies across different forecast target periods for each of the three evaluation measures, R2, MAE, and RMSE. In particular, Random Forest Regression and Decision Tree Regression exhibited relatively high predictive performance, followed by K-Nearest Neighbors Regression.

In addition, adding lags of the opening, high, and low prices to the lags of the closing price as explanatory variables slightly improved predictive performance in some cases. This suggests that incorporating information related to intraday price formation, in addition to historical closing-price information, may contribute to improving predictive performance.

Furthermore, comparison across different forecast target periods indicated that predictive performance tends to improve as the period is extended from five to ten years. This result suggests that, when forecasting financial time series such as the Nikkei 225 using machine learning models, the period of data used in the analysis may affect predictive performance.

Overall, this study demonstrated that, in forecasting the daily closing price of the Nikkei 225, not only the type of model but also the composition of explanatory variables and the forecast target period are important factors to consider when evaluating predictive performance. This clarification is the most significant contribution of this study. Moreover, we found that, among the models compared in this study, Random Forest Regression and Decision Tree Regression performed relatively well. This finding is also an important contribution of this study.

This study is a comparative analysis centered on price data for the Nikkei 225, and whether similar results would be obtained for other markets, or when additional information such as trading volume, news, and investor sentiment is incorporated, remains a subject for future research. Nevertheless, the present study highlights the critical importance of meticulous foundational research. From this perspective, we anticipate that the findings of the present study will make a significant contribution to further research in this area and serve as a valuable basis for further investigation.

Acknowledgements

The author is grateful for the repeated cordial invitations from Joy Deng, Managing Editor of the Journal. The author also thanks the anonymous reviewers for their supportive and constructive comments on this paper. Additionally, the author appreciates the financial support from the Chuo University Grant for Special Research. Lastly, the author also thanks all the editors of this journal for their kind attention to this paper.

Author Contributions

The author designed and conducted all of the research.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Abir, S. I., Shoha, S., Al Shiam, S. A., Zakaria, R. M., Shimanto, A. H., Shamsul Arefeen, S. M. et al. (2025). Deep Learning for Financial Markets: A Case-Based Analysis of BRICS Nations in the Era of Intelligent Forecasting. *Journal of Economics, Finance and Accounting Studies, 7,* 1-15. https://doi.org/10.32996/jefas.2025.7.1.1

Chen, X. (2023). Stock Price Prediction Using Machine Learning Strategies. *BCP Business*

& Management, 36, 488-497. https://doi.org/10.54691/bcpbm.v36i.3507

Ghosh, B. P., Bhuiyan, M. S., Das, D., Nguyen, T. N., Jewel, R. M., Mia, M. T., Cao, D. M., & Shahid, R. (2024). Deep Learning in Stock Market Forecasting: Comparative Analysis of Neural Network Architectures across NSE and NYSE. *Journal of Computer Science and Technology Studies, 6,* 68-75. https://doi.org/10.32996/jcsts.2024.6.1.8

Hong, A., Gao, M., Gao, Q., & Peng, X. (2023). Non-Stationary Financial Time Series Forecasting Based on Meta-Learning. *Electronics Letters, 59,* e12681. https://doi.org/10.1049/ell2.12681

Jiang, X., Mo, H., & Li, H. (2019). Stock Classification Prediction Based on Spark. *Procedia Computer Science, 162,* 243-250. https://doi.org/10.1016/j.procs.2019.11.281

Liu, Q., Lee, W., Huang, M., & Wu, Q. (2023). Synergy between Stock Prices and Investor Sentiment in Social Media. *Borsa Istanbul Review, 23,* 76-92. https://doi.org/10.1016/j.bir.2022.09.006

Moghar, A., & Hamiche, M. (2020). Stock Market Prediction Using LSTM Recurrent Neural Network. *Procedia Computer Science, 170,* 1168-1173. https://doi.org/10.1016/j.procs.2020.03.049

Mokhtari, S., Yen, K. K., & Liu, J. (2021). Effectiveness of Artificial Intelligence in Stock Market Prediction Based on Machine Learning. *International Journal of Computer Applications, 183,* 1-8. https://doi.org/10.5120/ijca2021921347

Roondiwala, M., Patel, H., & Varma, S. (2017). Predicting Stock Prices Using LSTM. *International Journal of Science and Research (IJSR), 6,* 1754-1756. https://doi.org/10.21275/art20172755

Tsuji, C. (2022). Exchange Rate Forecasting via a Machine Learning Approach. *iBusiness, 14,* 119-126. https://doi.org/10.4236/ib.2022.143009

Tsuji, C. (2025). Forecasting Exchange Rates with Artificial Intelligence: A Study of European Currencies. *International Business Research, 18,* 24-34. https://doi.org/10.5539/ibr.v18n5p24

Tsuji, C. (2026). Volatility Estimation and Crash Risk Forecasting Using Unified GARCH Models: Empirical Evidence from the Russian Invasion of Ukraine. *Quantitative Finance and Economics, 10,* 438-480. https://doi.org/10.3934/qfe.2026018

Vishwas, H. N., Usha, B, A., & Sangeetha, K. N. (2025). Artificial Intelligence-Based Stock Market Price Prediction, a Review. *Mathematical Models in Engineering, 11,* 204-229. https://doi.org/10.21595/mme.2025.24861