BOHR International Journal of Internet of things, Artificial Intelligence and Machine Learning
2024, Vol. 3, No. 1, pp. 58–66
DOI: 10.54646/bijiam.2024.26
www.bohrpub.com

RESEARCH

Stock price prediction through an artificial intelligence model using basic, technical, and macroeconomic indicators

Eui-Bang Lee[1] and Heon Baek[2*]

[1]Department of Data Science, Hwasung Medi-Science (HSMU) University, Hwaseong-si, Republic of Korea
[2]Graduate School of Business Administration, Sogang University, Seoul, Republic of Korea

[*]**Correspondence:**
Heon Baek,
huny01017@hanmail.net

Received: 26 September 2024; **Accepted:** 06 October 2024; **Published:** 22 December 2024

This study aims to analyze and predict future stock values more accurately. We proposed a stock price prediction model based on an artificial intelligence model using basic stock price data and technological and macroeconomic indicators. As a result of the experiment, the model using the features of adding technical indicators to the actual stock index (basic indicators) has better performance than the model using basic indicators and forecast performance by adding basic, technical, and macroeconomic indicators. Comparing artificial intelligence algorithms, the LightGBM model performed better than the deep neural network model and random forest model.

Keywords: stock price prediction, macroeconomic indicator, technical indicator, LightGBM, random forest

1. Introduction

Predicting the stock market is a primary concern for stock traders, individual investors, and portfolio managers. Due to the COVID-19 pandemic, stock markets around the world have plummeted. The stock markets of major countries began to recover after a sharp decline in 2020 compared to 2019. This is because liquidity in the stock market increased as the share of the IT industry in the stock market expanded (1). In 2021, the recovery accelerated sharply, and in 2022, a sharp return to pre-COVID-19 pandemic levels began. However, it is not easy to predict trading patterns and stock prices in the stock market, and it has become more difficult to determine this direction due to complex international relations and economic indicators.

Recently, research has been conducted using actual stock indices and technical indicators to predict changing stock indices influenced by various market fluctuations. In addition, research is being conducted using macroeconomic indicator data, which are important indicators of national management and economic growth or fluctuations (2, 3). As stock price prediction research is being conducted in various stock markets using artificial intelligence algorithms, a method that is different from data combining methods, the possibility of helping users make various investment decisions with high prediction performance is increasing (4). In particular, the performance of artificial intelligence algorithms is improving as new algorithms are continuously added. Machine learning and deep learning are popular in the stock prediction market due to their efficient modeling methods based on big data with less dependence on prior knowledge (5, 6).

Therefore, in this study, to more accurately predict fluctuations in stock price prediction, basic stock price data, technical indicators derived from stock price prediction, and various macroeconomic indicators are generated, and extended input features are applied to machine learning algorithms to test whether stock price prediction is effective. For the forecast period, we want to compare the closing price prediction performance from the closing price of the next day to that of 5 days later.

2. Literature review

2.1. Stock price prediction

There has been a lot of research on stock price prediction in the stock market, and experiments using various methodologies are still ongoing. In particular, with the advent of artificial intelligence, various techniques are being explored for stock price prediction using machine learning methodologies to predict stock market movements (7). Related studies that have recently been developed in this field are as follows, and a brief overview is shown in **Table 1**. Recognizing the difficulties of forecasting stock prices, tree bagging, random forests, and logit models have been developed under the premise that they are more successful at predicting stock price directions. As a result of performing a performance comparison, it was suggested that tree bagging and random forest are more useful methods for predicting stock prices (8). It created a model to predict the prices of 10 high-yielding stocks in the CSI300 index, and random forests outperformed ANN and logistic regression (9). Sharma and Juneja (10) used LSBoost and showed that a model using technical indicators outperformed support vector regression (SVR). The proposed model achieved better prediction performance than did SVR (10). Bucci (11) developed a model to predict the volatility of the S&P 500 index. Capturing the long-term dependence of volatility through a nonlinear autoregressive model process using LSTM (long

TABLE 1 | Related study summary.

Author	Method	BI	TI	MI	Contribution
Sadorsky (8)	Logit, Random Forest		✓		Propose a Random forest model in stock price prediction
Sharma and Juneja (10)	LSboost, SVR		✓		Random Forest using LSboost perform significantly better than SVR.
Bucci (11)	LSTM, NARX neural networks		✓	✓	LSTM can outperform traditional econometric methods.
Guo et al. (12)	LSTM, LightGBM	✓			Propose the hybrid LSTM-LGBM model
Yong et al. (13)	DNN	✓			The DNN algorithm has high performance, so it was used in a trading system and showed promising results with high profits.
Montenegro and Molina (14)	DNN	✓			High performance was achieved using the DNN model and sliding window technique
Wu et al. (15)	k-means, AprioriAll algorithm	✓	✓		The algorithm proposes a new way to perform stock trend forecasting with sequential charttterns. It can make financial forecasts and yield excess profit value.
Kimand Kim (17)	Granger Causality, Impulse Response Function			✓	KOSPI index has predictive power for producer prices, consumer prices, and the won-dollar exchange rate.

Basic indicators (BI), technical indicators (TI), and macroeconomic indicators (MI).

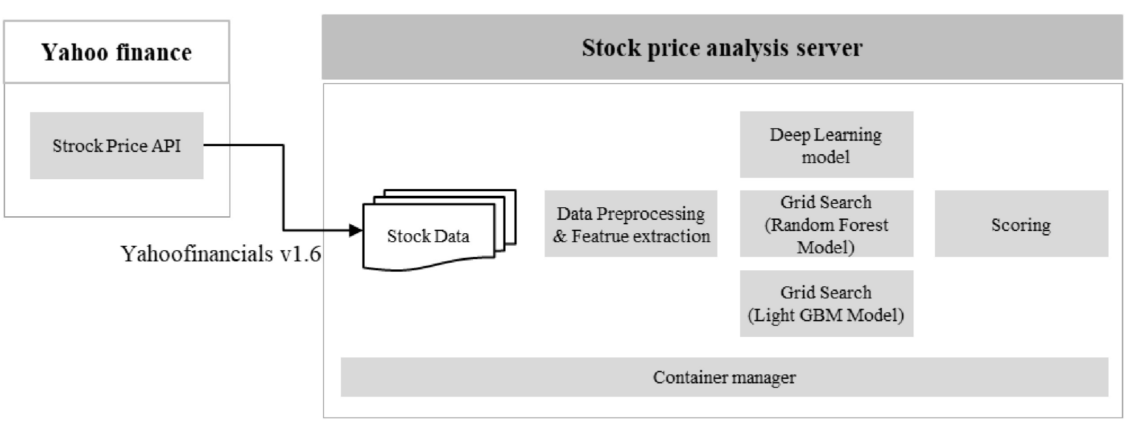

FIGURE 1 | Analysis environment.

TABLE 2 | Prediction performance of the KOSPI.

DNN		T1	T2	T3	T4	T5
BI (Basic indicator)	MAE	24.5008	34.7575	47.6143	64.1523	87.5041
	MSE	985.9465	2023.7133	3518.4627	6050.5581	10532.7693
	MAPE	0.8775	1.2509	1.6970	2.2821	3.1097
BTI (Technical Indicator)	MAE	24.8440	34.3957	44.4298	59.2416	59.1598
	MSE	986.1738	1946.8969	3156.8255	5297.1598	5645.9781
	MAPE	0.8894	1.2357	1.5996	2.0892	2.1413
BTMI (Macroeconomic Indicator)	MAE	25.9114	38.7801	44.5759	53.7806	62.3855
	MSE	1080.6870	2382.9919	3329.7593	4810.0118	6051.5418
	MAPE	0.9322	1.3987	1.5919	1.9310	2.2497
RF		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	25.0548	39.4760	58.5224	59.3093	105.7611
	MSE	1003.4892	2426.2614	5015.0607	5332.0961	15172.8562
	MAPE	0.8950	1.4014	2.0591	2.0987	3.6630
BTI	MAE	26.7357	40.2452	53.2740	65.7834	95.2669
	MSE	1130.4302	2539.7718	4289.4109	6441.8174	12525.6653
	MAPE	0.9516	1.4267	1.8820	2.3179	3.3123
BTMI	MAE	27.9472	40.6793	52.2615	62.9202	75.7710
	MSE	1223.8435	2567.3550	4097.4388	5866.8487	8144.4465
	MAPE	0.9903	1.4393	1.8488	2.2216	2.6651
LightGBM		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	24.7092	34.4849	43.8719	49.7584	52.4897
	MSE	984.1829	1940.0050	3070.8416	3984.5340	4617.2874
	MAPE	0.8834	1.2368	1.5705	1.7852	1.8907
BTI	MAE	24.3030	34.9023	43.4052	49.0996	54.6748
	MSE	960.6301	1972.1393	3026.7709	3922.9062	5000.3464
	MAPE	0.8705	1.2534	1.5590	1.7667	1.9756
BTMI	MAE	25.7689	40.1558	48.2130	63.7915	66.4421
	MSE	1052.6615	2509.5719	3581.6482	5952.7456	6551.1847
	MAPE	0.9195	1.4327	1.7231	2.2732	2.3737

short-term memory) and NARX (exogenous input) neural network models was shown to improve forecasting accuracy even during periods of high volatility (11). Guo et al. (12) presented that a hybrid model combining the LSTM model and the LightGBM model showed high performance.

In addition, studies using data processed to suit the characteristics of time series data were also conducted. First, Yong et al. (13) created a deep neural network (DNN) model to predict the FTSE Strait Time Index t days in the future and proposed a trading system with a 70% profitability (13). Montenegro and Molina (14) presented a sliding window technique based on S&P 500 index data. This has contributed to providing scholars with experience with new approaches and supporting investment decisions in the stock market (14). Wu et al. (15) combined the k-means and AprioriAll algorithms to create a model and showed that it can be used for a long period of time to make profits in real markets (15). Wen et al. (16) presented results showing that models using augmentation methods for time series classification models had low accuracy. However, it was confirmed that the performance slightly improved when the scaling method was applied to the DNN algorithm, and in this study, we developed a model by applying the scaling method to machine learning and deep learning algorithms.

Like the previous studies discussed above, stock prediction research uses a variety of methodologies to test predictability. These studies used ML algorithms and technical indicators. However, the stock price can be affected by macroeconomic factors in addition to information about the company that issued the stock, transaction price, and volume. Kim and Kim (17) analyzed the correlation between macroeconomic variables and stock prices. Granger causality analysis results showed that many variables were correlated. In this study, various technical and economic indicators are used, and stock price prediction is performed using various and latest artificial intelligence algorithms.

TABLE 3 | Prediction performance of the S&P 500 Index.

DNN		T1	T2	T3	T4	T5
BI	MAE	41.3556	73.6123	70.3912	83.4992	103.7099
	MSE	2897.3169	7721.7055	7884.9035	10822.5500	15571.8391
	MAPE	0.9992	1.7855	1.6954	2.0156	2.5171
BTI	MAE	37.7948	85.1625	67.3318	73.7846	103.5708
	MSE	2495.1603	9920.3262	7689.3335	9492.3187	15608.2077
	MAPE	0.9189	2.0603	1.6409	1.7991	2.5047
BTMI	MAE	113.8185	91.6274	63.3331	135.0529	104.6674
	MSE	16058.0708	11249.0806	7088.8953	23695.5208	16516.4189
	MAPE	2.7021	2.2149	1.5403	3.2533	2.5516
RF		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	39.3644	61.3276	63.2502	75.2512	103.0486
	MSE	2607.7551	5717.3074	7171.0530	9656.5952	15477.2388
	MAPE	0.9548	1.4891	1.5419	1.8361	2.5120
BTI	MAE	43.4755	55.9162	56.8351	74.9439	100.3898
	MSE	2963.5362	4873.4760	5617.8814	9020.1066	14330.2345
	MAPE	1.0507	1.3551	1.3838	1.8213	2.4358
BTMI	MAE	49.4815	65.7736	62.5904	118.8608	194.1196
	MSE	3586.4561	6128.7731	6307.9984	18519.0023	45679.6548
	MAPE	1.1930	1.5895	1.5211	2.8693	4.6746
LightGBM		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	36.4766	53.1851	62.6028	72.4293	82.2169
	MSE	2361.2409	4805.6890	7032.5101	9321.1833	11482.9991
	MAPE	0.8851	1.2939	1.5252	1.7671	2.0042
BTI	MAE	38.8309	48.3687	50.5224	67.1159	91.8713
	MSE	2541.0364	3960.0368	4752.5545	7896.6451	12775.6854
	MAPE	0.9411	1.1743	1.2301	1.6323	2.2304
BTMI	MAE	40.2679	58.5128	64.6482	81.4467	106.2063
	MSE	2668.1570	5041.8017	6682.8277	10280.7556	16117.6712
	MAPE	0.9755	1.4192	1.5737	1.9770	2.5762

2.2. Algorithm

2.2.1. Random forest

It is an algorithm that performs machine learning based on bagging (bootstrap aggregating), an ensemble learning method. RF draws a final conclusion by synthesizing the results of multiple decision trees obtained from subdatasets in a unique original training set by bagging (18, 19). The random forest algorithm further reduces prediction error by decreasing the correlation between decision trees. This robustness is not significantly affected by noise or outliers, and as the number of trees increases, the overfitting problem of decision trees can be overcome. In addition, since several predictors are randomly combined during the random node optimization process, stable results are derived even when there are many predictors. It has the advantage of being able to analyze the influence of various predictors without being influenced by the influence of variables (20).

2.2.2. LightGBM

Gradient boosting decision tree model developed by Microsoft in 2017. Compared with other existing machine learning algorithms, this approach reduces computational speed and memory consumption and optimizes parallel learning (21). While conventional tree-based algorithms use a level-by-level partitioning method, LightGBM uses a leaf-by-leaf tree partitioning method.

2.2.3. Deep neural network

A DNN uses more layers than the existing artificial neural network and adds features to greatly increase the expressive power of specialized data; the greater the number of hidden layers is, the greater the feature extraction function can be (22). The DNN is designed as a back-propagation

TABLE 4 | Prediction performance of the FTSE.

DNN		T1	T2	T3	T4	T5
BI	MAE	28.2533	46.2579	51.1611	112.8457	59.3123
	MSE	1462.6030	3465.9105	4374.2597	16172.2407	6033.2441
	MAPE	0.7105	1.1640	1.2914	2.8310	1.4974
BTI	MAE	27.2785	41.6835	55.3362	91.2142	153.7829
	MSE	1423.4344	3034.2613	4873.2789	11290.0227	28311.0770
	MAPE	0.6868	1.0509	1.3970	2.2926	3.8512
BTMI	MAE	27.3948	56.8762	114.4614	78.5782	146.1426
	MSE	1457.1140	5036.3237	17942.8319	9254.7578	26483.6857
	MAPE	0.6906	1.4323	2.8513	1.9842	3.6638
RF		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	27.4093	39.4044	49.5524	56.3955	62.2777
	MSE	1440.6212	2870.5768	4399.8963	5737.1284	6940.4254
	MAPE	0.6899	0.9942	1.2538	1.4274	1.5803
BTI	MAE	27.4100	38.9585	47.6122	54.0669	59.3116
	MSE	1429.4259	2794.1546	3967.9547	5025.9390	6044.0032
	MAPE	0.6897	0.9817	1.2008	1.3623	1.4956
BTMI	MAE	44.6823	60.3915	94.5423	114.9104	133.7460
	MSE	2785.8075	5150.0037	11515.9562	16580.3082	21942.4106
	MAPE	1.1191	1.5113	2.3618	2.8712	3.3412
LightGBM		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	27.3556	39.3305	48.4258	55.3602	59.2077
	MSE	1420.5494	2830.1171	4133.0911	5206.9024	5999.2161
	MAPE	0.6884	0.9915	1.2225	1.3986	1.4981
BTI	MAE	27.4126	39.2390	48.2906	54.7206	58.3820
	MSE	1416.9654	2792.0287	4108.2263	5230.8084	5941.4417
	MAPE	0.6895	0.9889	1.2176	1.3796	1.4755
BTMI	MAE	31.6406	54.6038	71.3116	91.9512	112.8690
	MSE	1698.2694	4434.2430	7302.1872	11467.5344	16574.8608
	MAPE	0.7953	1.3719	1.7921	2.3112	2.8381

algorithm (23). To improve the learning performance of the built model, optimal results can be derived by appropriately applying various techniques that can solve the problem of the model.

3. Methodology

The analysis environment of this study is shown in **Figure 1**. The analysis data was stored using network-attached storage. The server was operated using a laptop and tablet PC. Data preprocessing includes null value replacement according to the standard and generation of technical indicators. Feature extraction is performed for final use and data is saved to a file. The training ratio is 80% and the test ratio is 20%. After that, the data was converted to a dataset for min-max scaling and sliding windows, and then principal component analysis was performed conditionally. After creating a model using various algorithms, it is synthesized through scoring.

3.1. Data source and descriptive analysis

The selection of major indices used the indices of major countries used in previous studies, and included exchange rate information and digital currency between major countries. Futures included major raw materials and food. For analysis, we extract 12 years of stock price information data from finance.yahoo.com. Two of those years are used to generate technical variables and are not used for training or testing. The types of variables used after the null processing process are stock indices such as KOSPI, KOSDAQ, FTSE, S&P500, DowJones, NASDAQ, Nikkei, and HSI. Exchange rate information is Korea, UK, EUR, Japan, China, and Bitcoin. Bond is 10 years, 2 years. Future is Gold, Copper,

TABLE 5 | Prediction performance of the Nikkei Index.

DNN		T1	T2	T3	T4	T5
BI	MAE	272.8249	459.9882	483.6888	519.7405	645.9728
	MSE	123625.8031	337442.5134	370674.5315	455738.7106	659186.6588
	MAPE	0.9818	1.6586	1.7410	1.8735	2.3214
BTI	MAE	272.6831	472.0456	461.7705	958.1761	623.1839
	MSE	124781.8098	353483.8472	343363.9390	1314677.8670	617394.1366
	MAPE	0.9817	1.7027	1.6626	3.4476	2.2375
BTMI	MAE	299.0890	812.0159	526.4737	686.3744	647.3483
	MSE	140776.5156	903754.1832	449004.0024	743275.7972	669755.0259
	MAPE	1.0714	2.9192	1.8991	2.4920	2.3262
RF		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	268.6062	385.2419	481.8372	534.4075	591.9769
	MSE	117659.4718	235168.0319	368914.8487	472506.8908	580801.4891
	MAPE	0.9661	1.3863	1.7343	1.9242	2.1312
BTI	MAE	339.3918	572.1002	782.7323	1091.0303	1180.9983
	MSE	174023.8749	460260.3959	835784.9498	1524030.1342	1793061.3865
	MAPE	1.2161	2.0479	2.7976	3.8969	4.2173
BTMI	MAE	261.6831	413.0317	562.8869	858.6085	825.9138
	MSE	113351.8090	262352.7236	468983.7774	1002402.0803	985442.4951
	MAPE	0.9402	1.4811	2.0166	3.0614	2.9447
LightGBM		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	275.4296	402.8143	517.9181	556.1161	597.8837
	MSE	122079.8065	253862.8035	409077.4070	496352.6452	588104.3572
	MAPE	0.9897	1.4471	1.8605	1.9998	2.1516
BTI	MAE	280.2054	433.6003	542.5604	607.0885	577.6836
	MSE	125987.0646	288962.1810	444960.7317	570194.1272	530745.4374
	MAPE	1.0069	1.5559	1.9490	2.1801	2.0760
BTMI	MAE	251.0820	388.3674	513.4836	517.2289	553.7928
	MSE	103154.3604	234627.8430	402514.4198	438156.8333	506071.0285
	MAPE	0.9028	1.3965	1.8440	1.8627	1.9894

Crude Oil, Natural Gas Corn, Rough Rice, Soybean Meal, and Coffee. The technical variables are used by extracting data from Yahoo Financial. That is simple moving average (SMA), weighted moving average (WMA), Relative Strength Index (RSI), Stochastic K% D%, Moving Average Convergence Divergence (MACD).

3.2. Model optimization

The model training and optimization process was performed as follows. First, by dividing the training set, various parameters were tested using the 5-fold cross-validation technique. This is a method of calculating the performance of a test set by selecting the best parameter according to the criteria and learning the best parameter by combining the five divided training sets. This method was used for model training in the random forest and LightGBM models. Second, early stopping, a method mainly used to prevent overfitting of the DNN model, is a method for stopping learning if there is no improvement in the performance of the validation set through loss calculation while learning the training set. In this study, the training set and the validation set were classified as 8:2, and when the loss reduction in the validation set was less than 0.00001 while learning the training set, early stopping was used to stop the learning.

4. Analysis results and discussion

4.1. Results

The following steps are taken for prediction: (1) Extract 12 years of data. (2) The first 2 years are used only for technical

TABLE 6 | Prediction performance of the Hang Seng Index.

DNN		T1	T2	T3	T4	T5
BI	MAE	386.0887	393.6255	481.7207	557.6865	640.1345
	MSE	225596.2185	261161.4649	388542.0657	508337.8891	660413.7932
	MAPE	1.6094	1.6841	2.0516	2.4097	2.7652
BTI	MAE	285.2739	403.9066	481.4583	550.2282	634.7134
	MSE	144838.3849	277841.4669	388207.7483	489906.1453	651353.1176
	MAPE	1.2215	1.7097	2.0516	2.3639	2.7212
BTMI	MAE	797.7902	463.2472	480.5088	1159.5745	648.9619
	MSE	785357.6592	349599.0867	395901.5458	1813313.7631	663333.2212
	MAPE	3.3641	1.9872	2.0411	4.6434	2.7472
RF		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	279.1902	405.9005	501.8807	565.7887	645.4762
	MSE	135382.8406	273177.2318	409668.4214	514477.1732	664246.0897
	MAPE	1.2047	1.7620	2.1729	2.4548	2.8066
BTI	MAE	277.1096	403.8132	497.3962	563.2439	634.7536
	MSE	133848.9141	266901.7599	401415.1343	508232.6735	637343.8114
	MAPE	1.1927	1.7444	2.1502	2.4378	2.7459
BTMI	MAE	279.5876	445.2553	538.8628	799.3714	969.0903
	MSE	135369.7900	326730.8722	475461.6596	940061.4798	1343494.1854
	MAPE	1.1842	1.8657	2.2527	3.2847	3.9499
LightGBM		**T1**	**T2**	**T3**	**T4**	**T5**
BI	MAE	271.8605	394.1658	490.1688	551.3574	626.1597
	MSE	131019.2188	259942.0345	391997.0294	492595.7041	630199.1883
	MAPE	1.1610	1.6950	2.1061	2.3701	2.6952
BTI	MAE	274.9458	398.3284	489.9944	553.9011	624.6578
	MSE	132307.0813	264164.9780	395803.8520	496613.1145	624535.3257
	MAPE	1.1780	1.7118	2.1092	2.3891	2.6871
BTMI	MAE	262.3098	399.4820	479.9398	617.7644	753.1060
	MSE	124830.0025	269326.4243	385732.2030	592919.1614	838628.4767
	MAPE	1.1143	1.6924	2.0364	2.5899	3.1093

indicator generation. (3) 8 years of data out of 10 years are used for training/validation. (4) Provide 2 years of test results.

For the test set results, see Tables 2–6. The data collection period was 12 years before the reference date (FTAS 2022/11/03, GSPC 2022/10/31, HSI 2022/11/04, KS11 2022/10/28, and N225 2022/11/02), and 10 years were used for training/validation/testing. That is, the remaining two years are used for feature calculation of the moving average. The training/validation set and test set division should measure time intervals that are not included in the training/validation set; 8 years should be used for training/validation, and 2 years should be used for testing. That is, data from 2020 to 2022 were used (KOSPI test set: 2020/10/29 to 2022/10/28; training set: 2012/10/29 to 2020/10/28). Therefore, for training/validation, various parameters were used to improve the performance because the pattern of the stock market changes greatly due to a sharp rise in 2020 and a rise in late 2021 and 2022 after a decline in 2019.

4.2. Discussion

See summary (Table 7). We applied several models to predict the composite stock indices of major countries and selected the optimal model. The optimal model learns according to the optimal parameter selection method presented above. For each type of feature, one of nine features is selected as the DNN, RF, or LGBM for each type of basic, technological, overall index, or algorithm. The optimal model has 15 technical indicators and 5 basic/total indicators for each type of feature; in the algorithm, 16 LGBMs, 5 DNNs, and 4 RFs are used. Therefore, the features that should be applied first for stock price prediction include basic indicators and technical indicators (SMA and WMA of 5, 10, 20, 60, and

TABLE 7 | The best model for predicting each index.

		T1	T2	T3	T4	T5
KOSPI	Feature	BTI	BTI	BTI	BTI	BI
	Algorithm	LGBM	DNN	LGBM	LGBM	LGBM
	MAPE	0.8705	1.2357	1.5590	1.7667	1.8907
S&P500	Feature	BI	BTI	BTI	BTI	BI
	Algorithm	LGBM	LGBM	LGBM	LGBM	LGBM
	MAPE	0.8851	1.1743	1.2301	1.6323	2.0042
FTSE	Feature	BTI	BTI	BTI	BTI	BTI
	Algorithm	DNN	RF	RF	RF	LGBM
	MAPE	0.6868	0.9817	1.2008	1.3623	1.4755
Nikkei	Feature	BTMI	BI	BTI	BTMI	BTMI
	Algorithm	LGBM	RF	DNN	LGBM	LGBM
	MAPE	0.9028	1.3863	1.6626	1.8627	1.9894
HANG SENG	Feature	BTMI	BI	BTMI	BTI	BTI
	Algorithm	LGBM	DNN	LGBM	DNN	LGBM
	MAPE	1.1143	1.6841	2.0364	2.3639	2.6871

120 days, RSI, Stochastic K%, Stochastic D%, MACD, signal line, and histogram, respectively) combined. The reason that the performance of the technical indicator data, including basic indicators, is greater than that of the basic indicators when using the fewest indicators and that of the overall indicators (including basic, technical, and macroeconomic indicators) when using the most data is that indices with a high correlation (over 0.995) of T+1 to 5, such as the KOSPI and S&P, are included in the foundation; thus, it was concluded that the inclusion of technical indicators would improve the optimal performance. In addition, an algorithm that should be applied prior to RF or DNN for stock price prediction is proposed as a LightGBM model. The structure is considered to minimize overfitting and maximize performance due to the characteristics of the algorithm.

In previous studies, exchange rates, raw materials, and bonds are mentioned as indicators that affect stock price prediction, and some studies have shown them to be important factors, but the corresponding indicator was effective in Nikkei and Hang Seng. In other words, the results of this study provide important data for reference in the use of various algorithms and indicators depending on the indicator.

5. Conclusion

The COVID-19 pandemic has significantly shaken stock markets around the world. The need for more accurate stock price predictions has grown as stock prices undergo major changes. Various studies have mentioned that the addition of technical and macroeconomic indicators to input features for improving predictive performance leads to substantial performance improvement. This study combines the effective activities presented in these studies to present a model suitable for predicting the global stock market during the COVID-19 pandemic and after. To this end, a model was developed to predict the closing price of the composite stock indices of five major countries 1, 2, 3, 4, and 5 days after each, and the results were compared. As results, the optimal model was the most selected model with input features combining basic and technical index features, and the model using the LGBM algorithm was selected the most.

The implications of this study are as follows. First, to predict stock prices, stock prices must be derived through model optimization and best model selection according to the nationality and forecasting time of the indicator, but there are input indicators that can be used preferentially. In other words, the model using only basic and technical indicators has greater performance or less performance difference than the model added by synthesizing macroeconomic indicators. Therefore, in short-term forecasting, applying a model that uses only basic indicators or technical indicators, including basic indicators, is more efficient for model creation and operation (15 out of 25 basic/technical indicators were selected as the best model). Second, the algorithm that can be used preferentially is LightGBM (16 out of 25 models using LGBM were selected as the best model). A variant of the study using multi-sliding windows is shown (24). For model optimization, it is necessary to identify and compare the performances of various algorithms, which require high amounts of system resources and take a long time. The optimization method and results presented in this study have implications for implementing an efficient system for stock price prediction.

6. Limitations

In this study, various economic indicators and algorithms were used, and the best model was selected and compared using optimization techniques. However, macroeconomic indicators such as interest rates and price indices are also suggested to be important factors in the stock market. This study has the following limitations. First, based on the data provided on the finance.yahoo.com website, interest rates, price indices, real estate price indices, etc., are not considered, and it is difficult to generalize a model using macroeconomic input features. Second, it is difficult to generalize because it is impossible to apply various algorithms by concentrating on a few representative algorithms and presenting comparison results. Third, it was not possible to present a direction for mid- to long-term stock price prediction by generating a model to predict it based on short-term prediction. Finally, it is difficult to generalize because it is not possible to compare and analyze various prediction techniques, such as sliding window and multistep prediction methods, used for time series prediction. It is expected that future research will be able to derive more generalized information if such research is conducted.

Data availability statements

Data derived from public domain resources.

Conflict of Interest

The author has no relevant financial or non- financial interests to disclose. No funds, grants, or other support were received during the preparation of this manuscript.

References

1. Koreacapitalmarketinstitute. *Comparison of Global Stock Market Performance After the Spread of COVID-19*. (2022). Available online at: https://www.kcmi.re.kr/publications/pub_detail_view?syear=2020&zcd=002001016&zno=1551&cno=5537 (accessed December 11, 2022).
2. Choi J, Lee O. Correlation analysis among the price of apartments in Seoul, stock market and main economic indicators. *J Digital Convergence.* (2014) 12:45–59.
3. Neifar M, Dhouib S, Bouhamed J, Abdallah B, Islem A, Braiek B, et al. *The Impact of Macroeconomic Variables on Stock Market in United Kingdom. Munich Personal RePEc Archive* (2021). Available online at: https://mpra.ub.uni-muenchen.de/106246/ (accessed February 25, 2021).
4. Phoon K, Koh F. Robo-advisors and wealth management. *J Alternative Investments.* (2017) 20:79–94.
5. Chong E, Han C, Park F. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. *Exp Syst Appl.* (2017) 83:187–205.
6. Song Y, Lee J, Lee J. A study on novel filtering and relationship between input-features and target-vectors in a deep learning model for stock price prediction. *Appl Intell.* (2019) 49:897–911.
7. Soni P, Tewari Y, Krishnan D. Machine learning approaches in stock price prediction: A systematic review. in *Proceedings of the Journal of Physics: Conference Series*. IOP Publishing (2022). 012065
8. Sadorsky P. A random forests approach to predicting clean energy stock prices. *J Risk Financial Manag.* (2021) 14:48.
9. Zhang C, Bai Y. Chinese a share stock ranking with machine learning approach. in *Proceedings of the 2019 6th International Conference on Information Science and Control Engineering (ICISCE)*. Piscataway, NJ: IEEE (2019). 195–9.
10. Sharma N, Juneja A. Combining of random forest estimates using LSboost for stock market index prediction. in *Proceedings of the 2017 2nd International Conference for Convergence in Technology (I2CT)*. Piscataway, NJ: IEEE (2017). p. 1199–202.
11. Bucci A. Realized volatility forecasting with neural networks. *J Financial Econometrics.* (2020) 18:502–31.
12. Guo Y, Li Y, Xu Y. Study on the application of LSTM-LightGBM Model in stock rise and fall prediction. in *Proceedings of the MATEC Web of Conferences, MATEC Web Conf.* (Vol 336), EDP Sciences (2021). 05011
13. Yong B, Abdul Rahim M, Abdullah A. A stock market trading system using deep neural network. in *Proceedings of the Asian Simulation Conference*. Berlin: Springer (2017). 356–64.
14. Montenegro C, Molina M. A DNN approach to improving the short-term investment criteria for S&P500 index stock market. in *Proceedings of the 2019 3rd International Conference on E-commerce, E-Business and E-Government*. ACM (2019). p. 100–4.
15. Wu K, Wu Y, Lee H. Stock trend prediction by using a-means and a priori all algorithm for sequential chart pattern mining. *J Inf Sci Eng.* (2014) 30:653–67.
16. Wen Q, Sun L, Yang F, Song X. Time series data augmentation for deep learning: A survey. *arXiv [preprint]* (2020):doi: 10.24963/ijcai.2021/631
17. Kim J, Kim J. Study on interrelation between stock price and macroeconomic variables. *J CEO Manag Stud.* (2014) 17:163–86.
18. Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32.
19. Yang B, Di X, Han T. Random forests classifier for machine fault diagnosis. *J Mech Sci Technol.* (2008) 22:1716–25.
20. James G, Witten D, Hastie T. *An Introduction to Statistical Learning*. Berlin: Springer (2013).
21. Wang D, Zhang Y, Zhao Y. LightGBM: An effective miRNA classification method in breast cancer patients. in *Proceedings of the 2017 International Conference on Computational Biology and Bioinformatics*. Piscataway, NJ: IEEE (2017). 7–11.
22. Sun Y, Wang X, Tang X. Deep learning face representation from predicting 10,000 classes. in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. Piscataway, NJ: IEEE (2014). 1891–8.
23. Hinton G, Osindero S, Teh Y. A fast learning algorithm for deep belief nets. *Neural Comput.* (2006) 18:1527–54.
24. Baek H, Lee E-B. Feature expansion effect approach for improving stock price prediction performance. *Comput Econ* (2024): 1–26.