Predicting S&P 500 Index ETF (SPY) During COVID-19 via K-Nearest Neighbors (KNN) Algorithm

Wenguang Lin
Western Connecticut State University

Shengxiong Wu
Texas Wesleyan University

In this paper, the daily adjusted closing price of SPY (SPDR S&P 500 ETF Trust) is predicted by using the High-Low prices of SPY, DIA (SPDR Dow Jones Industrial Average ETF Trust), and QQQ (Invesco NASDAQ-100 ETF Trust) via the KNN method during the COVID-19 pandemic period. Results show that applying the KNN method, a simple, intuitive, and explainable machine learning method, is feasible and effective in SPY price prediction and corresponding trade decisions during the COVID-19 pandemic. Experiments also indicate that adding information on High-Low prices from DIA (a value tilt ETF) and QQQ (a growth tilt ETF) cannot improve the accuracy of both SPY price prediction and trading decisions. Results are consistent with previous findings based on the portfolio approach that value spread does not help predict stock market returns.

Keywords: K-Nearest Neighbors (KNN), nonparametric method, S&P 500 index ETF, forecasting, High-Low price

INTRODUCTION

The stock market plays an essential role in the world's economy since it is one of the key platforms for firms to raise capital. Lo and MacKinley (1988) show that stock returns are predictable, while the efficient market hypothesis (EMH) was widely accepted in the academy (Fama, 1970). In recent years, various machine learning approaches, such as Artificial Neural Networks (ANN), Decision Trees, KNN and Support Vector Machine (SVM), have become prevalent and have been employed for stock market prediction because of their edges in dealing with nonlinearity complexity, high-dimensional challenges, and hidden linkages among different factors in the stock market (Kumar et al., 2018). Gu et al. (2020) demonstrate that machine learning can effectively transform its predictor data set into a successful stock return forecasting model which outperforms traditional asset pricing approaches. Although Patel et al. (2015) indicate that the price of stock and index can be predicted with appropriate machine learning algorithms when the price information is cleaned efficiently, the prediction is still challenging (Henrique et al., 2019; Long et al., 2019). The stock market has high volatility and is embedded with many unmeasurable factors, such as the economy, politics, etc. (Oncharoen and Vateekul, 2018).

The machine learning approaches can be summarized into three categories: Supervised Learning, Unsupervised Learning, and Reinforcement Learning. This paper focuses on the KNN algorithm since it is intuitive and interpretable. It is one of the simplest supervised machine learning approaches, and its underlying mechanism is based on the idea that observations near each other should have similar characteristics. Previous research has demonstrated that the KNN method can generate better prediction accuracy than the time series method in economics, such as the linear AR-GARCH model (Meade, 2002). It has also been shown that the method performs better than other machine learning approaches, such as Logistic Regression, Linear Regression, Lasso, Elastic Net and Decision Tree Regression (Subha and Nambi, 2012; Toai et al., 2021). Subasi et al. (2021) find that within seven classifiers, including Random Forest, Bagging, AdaBoost, Decision Trees, SVM, KNN and ANN, KNN has the second highest accuracy (54%) in prediction for the NASDAQ dataset and has the highest accuracy for NIKKEI (56%) and FTSE (54%). Toai et al. (2021) show that 51% of the stock market movement direction is correctly predicted by using the KNN algorithm to predict VN INDEX value from Ho Chi Minh Stock Exchange between 2013 and 2019. Kumbure et al. (2022) also highlight the success of using KNN for stock market prediction conducted by Cao et al. (2019) and Zhang et al. (2017).

Although previous research shows that KNN is effective in predicting the stock price in various periods, it is interesting to know whether it works during the recent COVID-19 pandemic period, especially for finance practitioners. This article aims to investigate the prediction performance of SPY price movement via the KNN approach and its corresponding trading decision during the COVID-19 pandemic. Specifically, we use the High-Low prices of SPY, DIA and QQQ as the predictors in the KNN machine learning algorithm to predict the next day adjusted closing price of SPY, and the daily trading decision is made based on the predicted price. The main reason for choosing DIA and QQQ is due to their difference: DIA, tracking the Dow Jones Industrial Average composed of 30 "Blue-Chip" Stocks, is value tilted while QQQ, tracking the Nasdaq 100 Index composed of big tech firms such as FAANG (Formerly Facebook, Apple, Amazon, Netflix, Google), is growth tilted.

Different from comparing the KNN approach with other machine learning approaches in previous research, this paper focuses on comparing the effects of integrating different predictors into the KNN approach. The predictors include not only High-Low prices from SPY but also High-Low prices from other indices such as DIA and QQQ. Due to the belief that prices should reflect all available information in the market, the predictors used in this paper only include prices (no other index characteristics). Unlike most previous research using daily close price as a predictor, this article focuses on using High-Low prices to predict the next day's adjusted closing price since daily High-Low prices contain useful information for predicting return volatility (Parkinson, 1980). High-Low prices are also commonly used in industry as technical indicator candlesticks.

This paper contributes to the literature in the following area: First, it directly provides timely empirical evidence that applying the KNN machine learning algorithm in SPY price prediction is feasible and effective using a small data set during the COVID-19 pandemic. Second, via a machine learning approach, it confirms that information from the value tilted index (DIA) and growth tilted index (QQQ) does not enhance price prediction abilities for the SPY index. The finding is consistent with previous research that value spread does not have the power to predict stock market returns (Liu and Zhang 2008; Michou 2009).

DATA, METHODS, AND PERFORMANCE EVALUATION

Data

The historical daily highest price (High), the lowest stock price (Low), and the adjusted closing price from 03/11/2020 ~ 06/11/2022 are collected from Yahoo! Finance. The data is split into two datasets where the training data (or in-sample data) is from the period between March 11th, 2020 and December 30th, 2021, and the testing data (or out-of-sample) is the remaining data from the year 2022. Python libraries such as Pandas and NumPy are used to process the data, and Sklearn is used for applying the KNN algorithm.

K-Nearest Neighbors (KNN) Algorithm

The KNN algorithm is one of the most straightforward supervised machine learning approaches without parameters. For a KNN algorithm predicting a continuous value, like the adjusted closing price of SPY in this paper, the prediction is the average of the corresponding outcome variable of the nearest k neighbors to the data point to be predicted. The algorithm uses the data points observed in the training data set to predict the testing data points. Although there are different ways to measure the distance, Euclidean distance is one of the most popular metrics to define the k-nearest neighbors. The paper uses Euclidean distance, which can be calculated by using the following formula:

$$E.D. \ = \sqrt{(x_{1,t} - x_{1,t-p})^2 + (x_{2,t} - x_{2,t-p})^2 + \cdots + (x_{m,t} - x_{\mathrm{m},t-p})^2} \tag{1}$$

where $x_{1,t}$, $x_{2,t}$ and $x_{m,t}$ are predictors of data point at time t, and $x_{1,t-p}$, $x_{2,t-p}$ and $x_{m,t-p}$ are predictors of data point at time t-p.

The paper uses SPY to proxy for the S&P 500 index, DIA for the Dow & Jones Index, and QQQ for the NASDAQ index. Since High-Low prices are only available at the end of the day, the paper uses High-Low prices of the SPY, DIA, and QQQ to forecast the next day's adjusted closing price of SPY. In this paper, the KNN algorithms (models) include 1) HL_SPY model using daily High-Low prices of SPY as the predictors; 2) HL_SPY_DIA model using High-Low prices of SPY and DIA; 3) HL_SPY_QQQ model using High-Low prices of SPY and QQQ; 4) HL_SPY_DIA_QQQ model using High-Low prices of SPY, DIA and QQQ. The K is identified by using GridSearchCV from the Sklearn.

Performance Evaluation of Price Prediction and Trading Strategy

To evaluate the performance of price prediction of the KNN model, different metrics, including the Mean Absolute Deviation (MAD), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), and R-squared (R^2), are used.

Mean absolute deviation (MAD) is to measure the average of the absolute difference between the actual and predicted values:

$$MAD = \frac{\sum_{i=1}^{n} |A_i - P_i|}{n} \tag{2}$$

where A_i is the actual value of data point i, P_i is the forecast value for data point i, and n is the number of data points predicted.

Mean square error (MSE) measures the average of the square error. Different from the MAD, it penalizes larger errors:

$$MSE = \frac{\sum_{i=1}^{n} (A_i - P_i)^2}{n} \tag{3}$$

Root mean square error (RMSE) is the square root of MSE. Different from MSE, the RMSE has the same unit as the original data.

Mean absolute percentage error (MAPE) is the percentage average of absolute errors divided by actual values.

$$MAPE = \frac{\sum_{i=1}^{n} |\frac{A_i - P_i}{A_i}|}{n} \times 100 \tag{4}$$

R-squared is to measure the goodness fit of the model:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (A_i - P_i)^2}{\sum_{i=1}^{n} (A_i - \overline{P_i})^2} \tag{5}$$

Where $\overline{P_i}$ is the average of the predicted values.

The paper also considers a daily trading model based on the predicted price. If the predicted price of SPY the next day is higher than or equal to the adjusted closing price of SPY today, a buy signal is activated, and the trading decision is to buy the stock. In contrast, a sell signal is activated when the expected price is less than today's adjusted closing price. Essentially, a trading decision is a binary classifier. The following confusion matrix is used to understand the performance of the trading strategy:

	Real higher/equal price	Real lower price
Predicted higher/equal price	True Positive (TP)	False Positive (FP)
Predicted lower price	False Negative (FN)	True Negative (TN)

The performance of the binary trading strategy can be measured by accuracy, precision, recall and specificity:

$$\text{Accuracy} = (TP + TN)/(TP+TN+FP+FN) \tag{6}$$

$$\text{Precision} = TP/(TP+FP) \tag{7}$$

$$\text{Recall} = TP/(TP+FN) \tag{8}$$

$$\text{Specificity} = TN/(TN+FP) \tag{9}$$

where TP, TN, FP, and FN are the number of true higher/equal prices, true lower prices, false higher/equal prices, and false lower prices, respectively.

RESULTS AND DISCUSSION

Descriptive Statistics

Table 1 shows that during 3/11/2020 ~ 6/11/2022 High-Low prices of SPY have the highest mean and standard deviation when compared with corresponding values from DIA and QQQ. At the same time, the High-Low prices of DIA are close to corresponding values from QQQ but with a lower standard deviation than QQQ. The results also show that High prices of SPY, DIA and QQQ have relatively low volatilities compared with their corresponding Low prices. Interestingly, Gorenc et al. (2016) also found that the High price has less volatility than the close price.

TABLE 1
DESCRIPTIVE STATISTICS OF DAILY PRICE

03/11/2020~06/11/2022 Final Version

	SPY		DIA		QQQ	
	High	Low	High	Low	High	Low
Mean	392.11	386.59	315.57	311.43	318.33	312.47
Standard Error	2.49	2.52	1.70	1.74	2.27	2.28
Median	407.24	403.38	331.93	326.82	326.46	318.26
Standard Deviation	59.44	60.12	40.63	41.49	54.21	54.39
Kurtosis	-0.72	-0.55	-0.47	-0.25	-0.31	-0.24
Skewness	-0.57	-0.62	-0.79	-0.84	-0.59	-0.60
Range	250.30	257.80	178.30	185.11	234.50	237.65
Minimum	229.68	218.26	191.20	182.10	174.21	164.93
Maximum	479.98	476.06	369.50	367.21	408.71	402.58

Price Prediction and Its Performance Evaluation

Figure 1 and Table 2 show that during 3/11/2020 ~ 6/11/2022 the predicted price from using the HL_SPY model is generally closer to the actual SPY price when compared to using the HL_SPY_QQQ model. Although the standard deviation of the predicted price from the HL_SPY_QQQ model is similar to that from the HL_SPY model, its average predicted price, with a mean of 420.75 and a median of 423.16, is much lower than that (with a mean of 426.94 and a median of 430.54) from the HL_SPY model.

Figure 1 also shows that the predicted price from using the HL_SPY_DIA model is very close to that of using the HL_SPY model. On the other hand, the predicted price from using the HL_SPY_DIA_QQQ model is also very close to that of using the HL_SPY_QQQ model. These results imply that incremental information provided by DIA for the prediction of SPY is minimal.

FIGURE 1
SPY ACUTAL AND PREDICTED PRICES

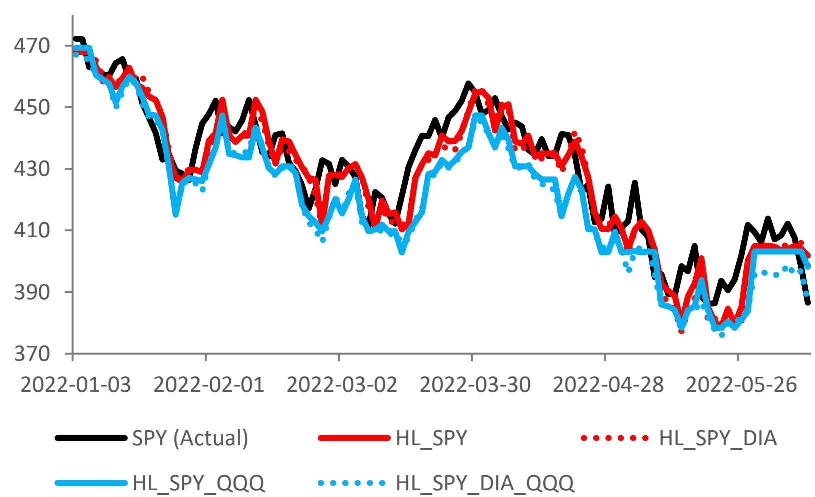

TABLE 2
DESCRIPTIVE STATISTICS OF THE SPY ACTUAL AND PREDICTED PRICES

		Prediction			
	SPY	HL_SPY	HL_SPY_DIA	HL_SPY_QQQ	HL_SPY_DIA_QQQ
Mean	429.04	426.94	426.18	420.75	419.52
Standard Error	2.04	2.14	2.22	2.12	2.17
Median	432.74	430.54	430.18	423.16	422.57
Standard Deviation	21.44	22.51	23.41	22.33	22.84
Kurtosis	-0.71	-0.63	-0.57	-0.47	-0.68
Skewness	-0.23	-0.32	-0.35	0.02	-0.02
Range	86.04	89.24	92.09	90.88	91.45
Minimum	386.20	378.72	377.15	378.39	375.61
Maximum	472.24	467.96	469.24	469.27	467.06

TABLE 3
COMPARATIVE ANALYSIS OF MAD, RMSE, MAPE AND R^2
FOR PRICE PREDICTION OF SPY

Model	MAD	RMSE	MAPE (%)	R^2
HL_SPY	6.6867	8.1783	1.57	0.8529
HL_SPY_DIA	6.9319	8.6846	1.63	0.8381
HL_SPY_QQQ	9.4163	11.6519	2.20	0.7087
HL_SPY_DIA_QQQ	9.9290	12.2097	2.32	0.6801

Table 3 provides information on performance evaluation for price prediction of different KNN models during 2022/01/03 ~ 2022/06/11. The forecasting error of the HL_SPY model is relatively low compared with other KNN models, as indicated by both MAD (6.6867) and RMSE (8.1783). The MAPE is 1.57% for the HL_SPY model, which indicates an average of 1.57% of the price difference between the actual movement of the SPY's daily adjusted closing price and the predicted price movement. The MAPE, in this case, is lower than 3.01% for the prediction of NYA, an NYSE Composite, via the KNN model from Lin et al. (2012). Table 3 shows that the HL_SPY has the highest price prediction accuracy compared with all other KNN models when using the metrics of MAD, RMSE or MAPE. With additional information on High-Low prices from the DIA, the MAPE increases from 1.57% for the HL_SPY model to 1.63% for the HL_SPY_DIA model, while MAD and RMSE increase to 6.9319 and 8.6846, respectively. At the same time, the HL_SPY_QQQ model has much higher MAD (9.4163) and RMSE (11.6519) compared with the corresponding MAD and RMSE from the HL_SPY model. For the HL_SPY_QQQ model, the MAPE increases to 2.22%. A combination of High-Low prices from SPY, DIA, and QQQ results in a MAPE of 2.32%. The result shows that adding information on High-Low prices from DIA and QQQ would increase the MAPE, while for QQQ, the increase of MAPE is much more significant. The result can be due to higher volatility for QQQ compared to that of DIA. Table 1 shows that the standard deviation of DIA's High (Low) price is only 40.63 (41.49), while QQQ has a standard deviation of 54.21 (54.39) for High (Low) price during the same period.

On the other hand, with additional information from DIA and QQQ, the R^2 decreases from 0.8529 for the HL_SPY model to 0.8381, 0.7087 and 0.6801 for HL_SPY_DIA, HL_SPY_QQQ and HL_SPY_DIA_QQQ models, respectively. The HL_SPY_QQQ model has a much lower R^2 compared with the HL_SPY model. The results can be due to the relatively high volatility from QQQ (compared with the volatility from DIA) incorporated into the HL_SPY_QQQ model.

In summary, integrating High-Low price information from DIA and QQQ has not improved the price prediction accuracy of SPY. The HL_SPY model has the best price prediction performance, measured by MAD, RMSE, MAPE and R^2, compared to the HL_SPY_DIA, HL_SPY_QQQ and HL_SPY_DIA_QQQ models.

Trading Decision and Its Performance Evaluation

A daily trading decision based on the predicted prices from different KNN models is also considered in this paper. If the predicted price the next day is higher than or equal to the adjusted closing price today, a buy signal is activated, and the trading decision is to buy the stock. In contrast, a sell signal is activated when the predicted price is less than today's adjusted closing price.

TABLE 4
PREDICTION PERFORMANCE ON THE DECISIONS OF TRADING

Model	Accuracy	Recall	Precision	Specificity
HL_SPY	0.53	0.41	0.47	0.63
HL_SPY_DIA	0.52	0.38	0.45	0.63
HL_SPY_QQQ	0.54	0.14	0.47	0.87
HL_SPY_DIA_QQQ	0.55	0.12	0.50	0.90

Table 4 shows that during 2022/01/03 ~ 2022/06/11 the HL_SPY model has an accuracy of 0.53 for the trading decisions, while the HL_SPY_DIA model has an accuracy of 0.52. The recall, precision, and specificity also decrease from 0.41, 0.47 and 0.63 for the HL_SPY model to 0.38, 0.45 and 0.63 for the HL_SPY_DIA model, respectively. As for the HL_SPY_QQQ model, the accuracy for the trading decision is equal to 0.54, while the recall decreases from 0.41 to 0.14. At the same time, the specificity increases from 0.63 to 0.87. This indicates that the prediction from the HL_SPY_QQQ model tilts into lower prices compared with using the HL_SPY model. Table 4 shows that the HL_SPY_DIA_QQQ model has a prediction accuracy of 0.55, which is very close to that of the HL_SPY model. In summary results show that all these four KNN models have an accuracy larger than 50% for the corresponding trading decision. Results also show that adding High-Low price information from DIA and QQQ in the algorithm cannot significantly enhance the accuracy of SPY trading decisions.

FIGURE 2
CUMULATIVE RETURNS

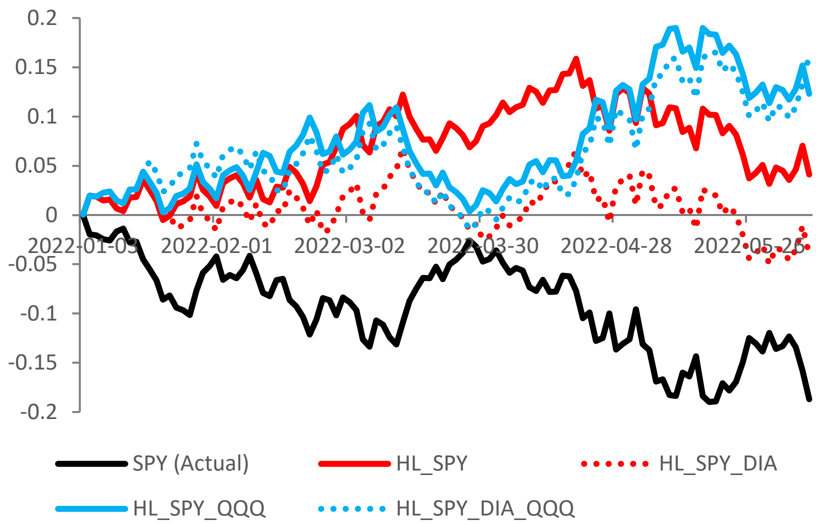

Figure 2 compares cumulative returns from buy-and-hold and other trading strategies during 2022/01/03 ~ 2022/06/11. It shows that all these trading strategies based on price prediction from the KNN models outperform the buy-and-hold strategy. And the HL_SPY_QQQ model performs better than the HL_SPY model during 2022/01/03 ~ 2022/06/11, while the HL_SPY model performs better than the HL_SPY_DIA model during the same period. At the end of the period, HL_SPY_QQQ has a cumulative return of 12.28%, which outperforms the HL_SPY model (4.12%) and actual buy-and-hold strategy (-18.70%). At the same time, figure 2 shows that it is more volatile to use the HL_SPY_QQQ model (with a standard deviation of 5.29%), compared with the HL_SPY model (with a standard deviation of 4.14%). For example, on 04/02/2022, the cumulative return for the HL_SPY_QQQ model only is 5.50%, which is much lower compared with the cumulative return of 15.86% for the HL_SPY model on the same day. Interestingly, based on the cumulative return, the HL_SPY_DIA and HL_SPY_DIA_QQQ models are dominated by HL_SPY and HL_SPY_QQQ models, respectively, indicating that adding information from DIA does not improve any information efficiency from the measurement of cumulative return for the trading decisions.

Robustness Check

Table 5 provides additional information on performance evaluation for price prediction of different KNN models during 2022/01/03 ~ 2022/09/11. Except for the HL_SPY_DIA model having a little higher R^2 than the HL_SPY model during this period, the results are quite consistent with the previous findings, showing that it is effective to use only the High-Low prices of SPY for its price prediction, and that there is no significant value added when incorporating High-Low prices from DIA and QQQ.

TABLE 5
COMPARATIVE ANALYSIS OF MAD, RMSE, MAPE AND R^2 FOR PRICE PREDICTION OF SPY DURING 2022/01/03 ~ 2022/09/11

Model	MAD	RMSE	MAPE (%)	R^2
HL_SPY	6.6473	8.1783	1.56	0.8532
HL_SPY_DIA	7.8928	9.6289	1.92	0.8608
HL_SPY_QQQ	10.9779	13.5430	2.67	0.7246
HL_SPY_DIA_QQQ	11.5593	13.8759	2.29	0.7109

Table 6 shows that the accuracy of the trading decision based on the predicted price is also larger than 50% for different KNN models during 2022/01/03 ~ 2022/09/11 and that the accuracy for trading decision does not improve when incorporating High-Low prices from DIA and QQQ.

TABLE 6
PREDICTION PERFORMANCE ON THE DECISIONS OF TRADING DURING 2022/01/03 ~ 2022/09/11

Model	Accuracy	Recall	Precision	Specificity
HL_SPY	0.54	0.41	0.48	0.65
HL_SPY_DIA	0.50	0.27	0.42	0.69
HL_SPY_QQQ	0.53	0.09	0.41	0.89
HL_SPY_DIA_QQQ	0.54	0.08	0.46	0.93

In summary, the robustness check confirms that: 1) using KNN is an effective method for the price prediction of PSY during the Covid-19 period; 2) incorporating information from the DIA and QQQ cannot enhance the market price prediction of SPY. The results are also consistent with previous research (Liu and Zhang 2008; Michou 2009), which indicates that value spread does not help predict the market return.

CONCLUSIONS

Using machine learning approaches to predict the stock price is an exciting topic in asset management. In this paper, we apply the KNN algorithm to predict the daily price of SPY during the COVID-19 pandemic period. The findings show that applying the KNN algorithm to the price prediction of SPY is feasible and effective when using a small data set, i.e., High-Low daily prices during the pandemic period of 03/11/2020 ~ 06/11/2022. The results also show that integrating High-Low daily prices from DIA (a value tilted ETF) and QQQ (a growth tilted ETF) does not improve the accuracy of SPY price prediction. The results are consistent with previous empirical work (Liu and Zhang 2008; Michou 2009), which uses a portfolio approach to show that value spread does not have a predictive ability for aggregate stock returns.

REFERNCES

Cao, H., Lin, T., Li, Y., & Zhang, H. (2019). Stock price pattern prediction based on complex network and machine learning. *Complexity*.

Fama, E.F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance*, *25*(2), 383–417.

Gorenc Novak, M., & Velušček, D. (2016). Prediction of stock price movement based on daily high prices. *Quantitative Finance*, *16*(5), 793–826.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *The Review of Financial Studies*, *33*(5), 2223–2273.

Henrique, B.M., Sobreiro, V.A., & Kimura, H. (2019). Literature review: Machine learning techniques applied to financial market prediction. *Expert Systems With Applications*, *124*, 226–251.

Kumar, I., Dogra, K., Utreja, C., & Yadav, P. (2018, April). A comparative study of supervised machine learning algorithms for stock market trend prediction. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)* (pp. 1003–1007). IEEE.

Kumbure, M.M., Lohrmann, C., Luukka, P., & Porras, J. (2022). Machine learning techniques and data for stock market forecasting: A literature review. *Expert Systems With Applications*, 116659.

Lin, A., Shang, P., Feng, G., & Zhong, B. (2012). Application of empirical mode decomposition combined with k-nearest neighbors approach in financial time series forecasting. *Fluctuation and Noise Letters*, *11*(02), 1250018.

Liu, N., & Zhang, L. (2008). Is the value spread a useful predictor of returns? *Journal of Financial Markets*, *11*(3), 199–227.

Lo, A.W., & MacKinlay, A.C. (1988). Stock market prices do not follow random walks: Evidence from a simple specification test. *The Review of Financial Studies*, *1*(1), 41–66.

Long, W., Lu, Z., & Cui, L. (2019). Deep learning-based feature engineering for stock price movement prediction. *Knowledge-Based Systems*, *164*, 163–173.

Malinda, M., & Chen, J.H. (2022). The forecasting of consumer exchange-traded funds (ETFs) via grey relational analysis (GRA) and artificial neural network (ANN). *Empirical Economics*, *62*(2), 779–823.

Meade, N. (2002). A comparison of the accuracy of short term foreign exchange forecasting methods. *International Journal of Forecasting*, *18*(1), 67–83.

Michou, M. (2009). Is the value spread a good predictor of stock returns? UK evidence. *Journal of Business Finance & Accounting*, *36*(7-8), 925–950.

Oncharoen, P., & Vateekul, P. (2018, August). Deep learning for stock market prediction using event embedding and technical indicators. In *2018 5th international conference on advanced informatics: Concept theory and applications (ICAICTA)* (pp. 19–24). IEEE.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of Business*, pp. 61–65.

Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015). Predicting stock and stock price index movement using trend deterministic data preparation and machine learning techniques. *Expert Systems With Applications*, *42*(1), 259–268.

Subasi, A., Amir, F., Bagedo, K., Shams, A., & Sarirete, A. (2021). Stock Market Prediction Using Machine Learning. *Procedia Computer Science*, *194*, 173–179.

Subha, M.V., & Nambi, S.T. (2012). Classification of Stock Index movement using k-Nearest Neighbours (k-NN) algorithm. *WSEAS Transactions on Information Science & Applications*, *9*(9), 261–270.

Toai, T.K., Thinh, B.T., & Senkerik, R. (2021, September). Predicting VN-Index Value by KNN Algorithm of Machine Learning. In *3rd Annual International Conference on Public and Business Administration (AICoBPA 2020)* (pp. 228–233). Atlantis Press.

Zhang, N., Lin, A., & Shang, P. (2017). Multidimensional k-nearest neighbor model based on EEMD for financial time series forecasting. *Physica A: Statistical Mechanics and its Applications*, *477*, 161–173.