MJoSHT
Malaysian Journal of Science, Health & Technology
[mjosht.usim.edu.my]

Article

Malaysian Daily Stock Prediction Analysis Using Supervised Learning Algorithms

Hazirah Halul and Karmila Hanim Kamil

Faculty of Science and Technology, Universiti Sains Islam Malaysia, Bandar Baru Nilai, 71800, Malaysia.

Correspondence should be addressed to:
Karmila Hanim Kamil; karmila@usim.edu.my

Article Info
Article history:
Received: 19 July 2021
Accepted: 1 April 2022
Published:2 June 2022

Academic Editor:
Shahrina Ismail

Malaysian Journal of Science, Health & Technology

MJoSHT2022, Volume 8, Issue No. 2
eISSN: 2601-0003

https://doi.org/10.33102/2022229

***Abstract*— Nowadays, Machine Learning (ML) plays a significant role in the economy, especially in the stock trading strategy. However, there is an inadequate extensive data analysis using various ML methods. Previous findings usually focus on the forecasting stock index or selecting a limited number of stocks with restricted features. Therefore, the contribution of this paper focused on evaluating different supervised learning algorithms, namely Logistic Regression (LR), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB), on a big dataset from 28 stocks in Bursa Malaysia. By setting their parameter along and using Walk-Forward Analysis (WFA) method, the trading signal was evaluated based on Accuracy Rate, Precision Rate, Recall Rate, and F1 Score. For stock trading strategies in Malaysia in particular, the findings of this study show that SVM has a better performance compared to LR and XGB in time series forecasting. The ML algorithms have values ranging from 53% to 66% for Accuracy Rate (AR), Recall Rate (RR), and F1 Score (F1). In addition, SVM has the highest Precision Rate (PR) of 73% among the ML algorithms.**

***Keywords*— Machine Learning; Supervised Learning Classifier; Walk-Forward Analysis; time series forecasting.**

I. INTRODUCTION

Despite the claim in [1] that the Covid-19 pandemic was the worst financial crisis after the 1930s Great Depression crisis, an American survey in [2] showed that the number of the retail investing crowd has started to rise during the pandemic. Bursa Malaysia has also indicated an increase in local retail participation standing from 24% in mid-June 2019 to 41.44% in mid-July 2020 [3]. The rise in participation of retail investors has improved the Financial Times Stock Exchange (FTSE) Bursa Malaysia KLCI Index, which helped in recovering the losses from the sell-off in March 2020 due to the Movement Control Order (MCO) implemented by the Malaysian government [4]. Subsequently, the ongoing efforts in increasing stock market vibrancy and liquidity during the challenging pandemic situation had shown a positive impact on the stock market when Bursa Malaysia recorded a 62.0% increase in its Profit After Tax and Minority Interest of RM151.0 million from RM93.2 million in 2019, the highest first-half financial performance since its listing in 2005 [5].

Commenting on the latest overwhelming trading volume that hit the highest point of 27.8 billion on 11 August 2020 [6], Bursa Malaysia, however, expressed their concern over the retail investors, particularly the beginners. A proper assessment and analysis are essential for first-time investors from a basic and technical perspective so that they can make well-informed investing decisions and understand that their investments meet their risk appetite and investment purposes. Although getting retail investors into the stock market has always been a priority for the local bourse, investor's education and literacy is similarly vital agenda to be elevated based on the current fact that 78% of Central Depository System (CDS) accounts opened in the first seven months of 2020 were investors aged 45 years and below. Thus, the emergence of millennial investors in the stock market needs to be acknowledged by expanding the range of digital trading platforms, galvanising various types of decision-making and analytical Artificial Intelligence (AI) powered tools, as well as leveraging digital technology to enhance the efficiency in distributing information to a broader public [7]. These efforts could encourage retail investors to fully utilise the resources available in ensuring a sustainable market in the long term.

There are different ways in making a good investment decision. Technical analysis is one of the methods in investment decision-making. It uses technical indicators, which is a mathematical calculation based on historical information such as price, volume, or open interest that seeks to predict the financial market direction [8]. Traders can take note of any information that they can get from technical indicators and justify them so that they can make a greater profit and lessen the risk. Therefore, it is beneficial to apply technical analysis indicators as a component of trading strategy [9]. Traders typically bring together several technical indicators with further technical analysis in subjective forms, for example, considering the chart patterns to come up with trade ideas. Since there are several of them, many researchers, not only traders, study and analyze their efficiency and performance.

While there are those that still use traditional Technical Analysis methods to examine trends and patterns, more recent studies have attracted the attention of Machine Learning (ML) approaches, a subset of AI. This has been proven true as it is beneficial when combining technical indicators with computational intelligence since it gives better results [10]. Human traders cannot possibly accomplish to explore a large number of data sources instantaneously. However, ML algorithms can. Real-time news and trade results are monitored using mathematical models, and patterns that can influence stock prices to go up or down are identified. Based on its calculations, it then takes initiatives to sell, hold, or buy stocks and given the trading operations are in vast volumes, even a small advantage is equivalent to significant profits.

Nonetheless, insufficient research development and limited broad data analysis of different ML methods has put constraints on the usefulness of ML as a stock market prediction method. Previous findings generally focus on forecasting stock index price [11] or choosing a small number of stocks with restricted features based on their own preferences [12], or using a very short backtesting period [13].

Given all the statements above, the main objective of this study is to evaluate the efficiency of the ML algorithms used in trading strategy by assessing the evaluation indicators. Additionally, this study determines the best performance in forecasting from the ML algorithms used. The organization of the paper is organized as follows; Section I-A provides the background of ML algorithms, and Section I-B outline the recent works on ML algorithms in financial time series forecasting. Following, Section II gives details on data and methodology. Next, Section III offers the discussion on the findings, and the final part concludes the results and offers future recommendations.

A. Background of ML Algorithms

Machine Learning (ML), a subset of AI, is a concept in which the machine is granted to learn and train from models and experience without being designed too explicitly. Hence, algorithmic issues are pivotal. The algorithms for ML are developed to perform the learning tasks and are concerned with their computational effectiveness [14]. Therefore, the programmer, in this case, the trader, would write the code based on given data so the machine could make sense of the data and build logic or asset price prediction.

One subgroup of ML algorithms is supervised learning, also known as traditional ML. By definition, supervised learning is when the machine learns and make decisions based on the guidance of coding created by the programmer [15]. The dataset will act as an instructor and train the machine. A few examples of supervised learning are Random Forest (RF) and Super Vector Machine (SVM). Classification is one of the techniques used in supervised learning. It is defined as identifying the data to which a set of groupings belongs with the basis that the category of the training dataset is recognized.

This research focuses on three supervised learning classifiers.

1) Logistic Regression (LR): this model is designed for binary classification. It is a mathematical model used in statistics to calculate the approximate probability of an event occurring or not having been given prior data.

2) Support Vector Machine (SVM): it uses probabilistic methods to generate non-overlapping partitions and utilise the features [16]. The model estimates the optimal margin that clearly categorises the data points. Furthermore, it supports the vector to maximise the model fit's margin. There are several sorts of kernels that can be applied in SVM, such as linear, polynomial, and radial basis kernels.

3) Extreme Gradient Boosting (XGB): It can be used to analyse the data in vector form for classification and regression that generates a forecast model typically in decision tree form and simplifies them by granting the optimization of an arbitrary differentiable loss function [17]. This will give an output of probability when using logistic regression for binary classification. The framework is like gradient boosting, but it is more efficient.

B. Related Works

To compare the performance of Logistic Regression (LR), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB), the evaluation indicators are assessed. Most papers commonly use accuracy metrics as evaluation indicators to analyze the efficiency of the model in classification, which includes Accuracy Rate (AR) or F1 Score (F1). Other indicators include error metrics, which include Mean Absolute Error (MAE) or Mean Squared Error (MSE). Nevertheless, the way the data is interpreted matters most because it shows how important the findings are.

Based on previous papers, there are three findings in comparing the efficiency of the ML algorithms. The first finding shows that SVM performs better than the other two ML algorithms. Reference [18] shows that traditional ML works no better in directional evaluation indicators in contrasted to Deep Neural Network (DNN). However, DNN performs better with transaction costs. The paper shows that SVM has better efficiency in stock forecast among the selected ML algorithms in this study for the Chinese stock market, followed by XGB and LR. Another paper by [19] stated that it is useful to use SVM in time series forecasting as it has a small MSE value. SVM is also efficient in training the data; consequently, it has a short training time.

But there is also inconsistency from other papers that indicated SVM is underperforming compared to other ML models. Reference [10] discovered that compared to SVM, their proposed model, a classification model using the computational efficient functional link artificial neural network (CEFLANN), gives a greater profit percentage. Results from [12] found that RF has higher accuracy than SVM in stock price prediction. The discrepancies between the RF and SVM are almost 30%, with SVM at 68%. This suggests that the application of the SVM algorithm is less fitting for the five Malaysian stocks used in the paper. Similarly, the recent paper by [20] stated that RF has higher accuracy than SVM in predicting asset prices on Malaysian commodities prices and microstructure market variables. It is also found that the effectiveness of the algorithm used depends on the size of the dataset of the study.

The second finding is that LR performs better than other ML algorithms. Reference [21] found that LR, NB, SVM, and Gaussian Discriminant Analysis (GDA) can deliver predictability to some degree. Their findings demonstrate that LR is more accurate in forecasting than SVM and others. It also recommends that SVM with RBF kernel has better standing compared to other kernels that they used in predicting market trends, though it does not show its strong points before choosing the parameters. Recently, it has been found that LR has a percentage of 86% for correctly classified stock market movement for one Malaysian stock [13]. This shows that LR has an acceptable percentage of accuracy in price prediction to be applied to the Malaysian trading strategy.

Third and the last finding is that XGB performs better than the other two ML algorithms. Reference [22] studied on US stocks stated that RF and XGB can attain high accuracies for long-term predictions. This result is consistent with the outcomes from [23] that demonstrates XGB performs better, followed by SVM and LR. Similarly, among three selected ML classifiers in [18] for American stocks, XGB works better in forecasting, followed by LR and SVM. However, reference [24] showed that among the tree-based ML algorithm used in the paper, the Adaboost algorithm had generated the highest model fitting ability by having the most accurate results based on their error metrics, followed by XGB, which is placed second, having the longest average runtime per sample.

In summary, based on previous papers in asset price prediction, which focus on three supervised learning algorithms, the findings are divided into three which are (1) SVM has a better performance compared to LR and XGB, (2) LR is better than the other two algorithms, and (3) XGB has better performance than SVM and LR. However, there are also contradictory findings. Fig. 1 illustrates a compilation of evaluation indicators on supervised learning algorithms of related papers. Based on Fig. 1, it can be seen that LR is the best algorithm in stock prediction, which has the highest Prediction Rate (PR), Recall Rate (RR), and F1 values, followed by XGB has the highest AR value and SVM. Note that each outcome has a different number of the dataset, vary in terms of features and techniques in data analysis and applies different evaluation indicator. There are many different types of ML algorithms available that are being used in forecasting. Each has its advantages and disadvantages in forecasting. Therefore, this paper will study the accuracy of stock price prediction, specifically on Malaysian stocks, to assess the efficiency of the ML algorithms used in the trading strategy.

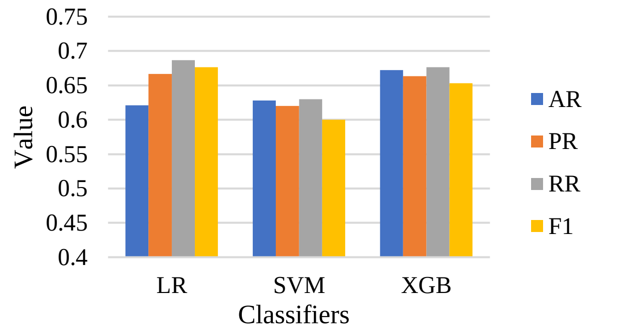

Fig. 1 Comparison of Evaluation Indicators on Supervised Learning Classifiers of Previous Papers.

II. DATA AND METHODOLOGY

A. *Data Collection*

The data of daily stock is taken from Yahoo! Finance, which includes open, close, high, low, adjusted close prices and volume using "getSymbols()" in the quantmod package used in the R language in the RStudio software [25]. Initially, thirty big companies in Bursa were chosen as training dataset (D) for this study. However, only twenty-eight stocks were used. Another two stocks were exempted because of a lack of data.

The study has a total of 1,250 trading days taken before the date February 1st, 2018. By having a big dataset for each stock, forecasting is more reliable and valid since it is in a longer period to ascertain the pattern.

B. *ML Algorithms Selection*

There are three supervised learning algorithms chosen to foresee the direction of the stock prices. The three models are Logistic Regression (LR), Support Vector Machine (SVM), and Extreme Gradient Boosting (XGB).

C. *Parameter Used in ML Algorithms*

The features or the parameters have consisted of well-known Technical Indicators. In this study, the number of features or Technical Indicators, n is set to be twenty for the data of each trading day. The chosen Technical Indicators are selected based on the readings and rationale of past papers, which includes Moving Average Convergence Divergence (MACD), On-Balance-Volume (OBV), Relative Strength Index (RSI), William %R index (WR), etc.

The Technical Indicators were computed using their equations based on the adjusted stock prices and volume. All the values have been normalized and are in the decimal point between 0 to 1, which also can be interpreted as a percentage of 0% to 100%. Unfortunately, the weight distribution of each parameter cannot be calculated because ML models have a semi-black box property meaning the input and output data is observable, but the internal mechanism is unknown.

D. Walk-Forward Analysis (WFA) Method

WFA is a process of optimizing a trading structure using a constrained set of parameters and then testing out-of-sample data with the best-optimized parameter set. Instead of working with the past data, the model which uses recent data is trained to carry out the forecast for the testing set, which is the future out-of-sample data. Next, a new training set is trained for the next round, and the previous training data set will walk forward in one step. This will enhance the robustness and the conviction of the real-time trading strategy. strategy.

For this study which follows a method similar to [18], the training and the testing can be seen illustrated in Fig. 2. The input format for each ML algorithm used is Matrix (m, n), which denotes a matrix of m rows and n columns. In the study, the data of past trading days, m = 250 is sampled as training set in each series of Walking-Forward Analysis (WFA) and the n = 20 is the technical indicator explained in Section II-C. For each step, the past 250 days (one year) of stock data is used as the training set, and the testing set is set based on the data for the next 5 days (one week). For every stock, each includes 1,250 trading days of 20 parameters of data, so it will take (1,250-250)/5 = 200 training sessions that will generate 1,000 predictions of trading signals for daily trading strategy.

E. Evaluation Indicators

The output of the ML algorithms will be an indicator to foretell the direction of the stock. The sequences of sets {UP, DOWN} is set as the actual label of the dataset. For this study, the profit of trading strategies is considered to be the "UP" label. Hence, the loss is considered a "DOWN" label. As a result, four classifications of actual and predicted labels are produced, as shown in Table II, which are TU, TD, FU, and FD. TU is the number of UP that both the actual and predicted labels are UP; TD is the number of DOWN that both the actual and the predicted labels are DOWN; FU is the number of UP that the predicted labels are UP, but actual is DOWN; FD is the number of DOWN that the predicted label values are DOWN however actual is UP.

TABLE I
THE MATRIX OF TWO CATEGORIZATION RESULTS OF SUPERVISED LEARNING ALGORITHMS

		Predicted Labels	
		UP	DOWN
Actual Labels	UP	TU	FD
	DOWN	FU	TD

There are four evaluation indicators used to evaluate which ML algorithms work best in forecasting based on the table above. The indicators are Accuracy Rate (AR), Precision Rate (PR), Recall Rate (RR) and F1 Score (F1).

AR is the number of predictions that is true to the total number of predictions ratio shown in Eq.1.

$$AR = \frac{(TU + TD)}{(TU + FD + FU + TD)}. \qquad (1)$$

PR is the number of "UP" that is correctly predicted to all predicted "UP" ratios shown in Eq.2. A high PR value indicates that ML algorithms focus on "UP" compared to "DOWN".

$$PR = \frac{TU}{(TU + FU)}. \qquad (2)$$

RR is the number of correctly predicted "UP" to the actual labelled "UP" ratio shown in Eq.3. High RR indicates ML algorithms effectively identify and obtain a large number of "UP". In reality, it is tricky to produce an algorithm that has a high PR and RR at one time.

$$RR = \frac{TU}{(TU + FD)}. \qquad (3)$$

Hence, it is important to calculate the ML algorithm ability in categorization that has both PR with RR using a new evaluation indicator. In this case, F1 Score becomes the balanced average of PR and RR. Therefore, F1 is a much more thorough evaluation indicator, as shown in Eq.4. It is important to note that a high F1 score corresponds to a good result.

$$F_1 = 2 * \frac{PR * RR}{(PR + RR)}. \qquad (4)$$

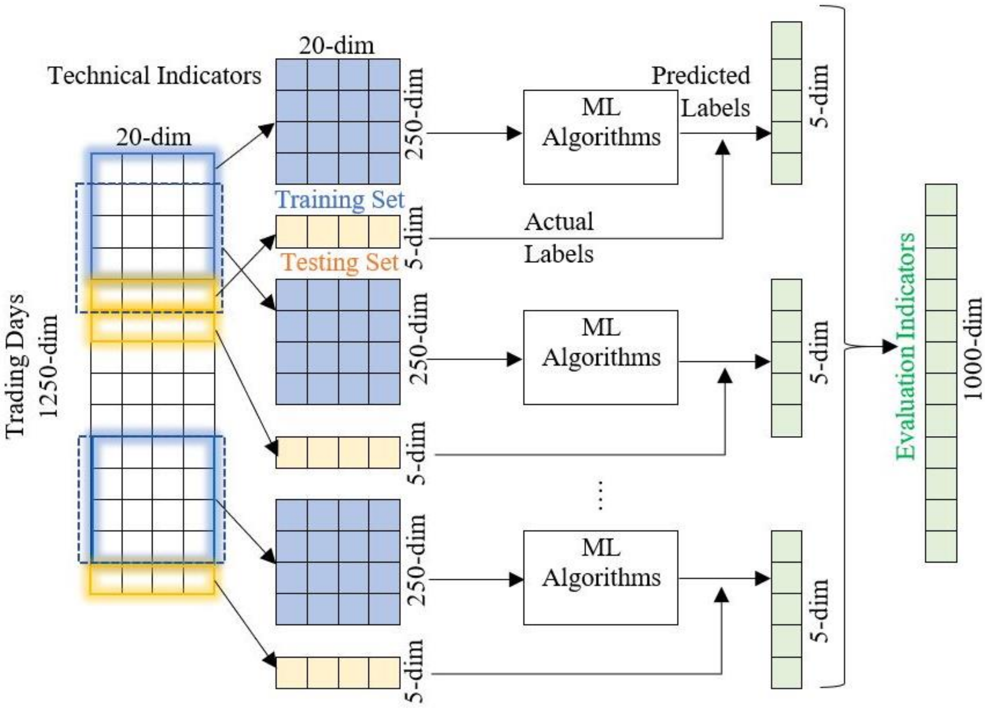

Fig. 2 Schematic diagram of WFA Method.

III. RESULTS AND DISCUSSION

Table II below shows a description of the results after data analysis was conducted in decimal points. The best result metric of the overall algorithm is in boldface. It can be seen that SVM has the best performance in the evaluation indicator, followed by LR and XGB.

TABLE II
EVALUATION INDICATORS USED ON SUPERVISED LEARNING ALGORITHMS

	Logistic Regression (LR)	Support Vector Machine (SVM)	Extreme Gradient Boosting (XGB)
Accuracy Rate (AR)	0.5484	**0.5571**	0.5313
Precision Rate (PR)	0.6603	**0.7317**	0.6052
Recall Rate (RR)	**0.6062**	0.6000	0.6013
F1 Score (F1)	0.6314	**0.6589**	0.6028

It is noteworthy to highlight that the best performance for traditional ML is SVM, followed by LR and XGB for AR, PR and F1. The value differences for PR are quite large, which is more than 10%. For RR, the best performance is LR for traditional ML, followed by XGB and SVM with a thin difference. With this, it shows that LR is the second-best performance in stock prediction after SVM, thus bringing XGB in the last place. Fig. 3 illustrates the comparison of Evaluation indicators on ML Algorithms used for LR, SVM and XGB. The chart represents the values from Table II for a better picture of the numbers.

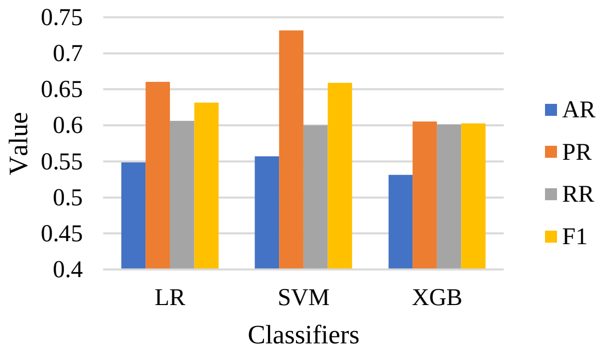

Fig. 3 Comparison of Evaluation Indicators on ML algorithms used.

Based on the findings, the results are consistent with the results from [18] and [19], which demonstrates that SVM has better performance in stock forecast compared to XGB and LR. This implies that Malaysian stocks are more compatible with the application of SVM than the other two ML algorithms in a trading strategy. This also shows that the Malaysian stock market behaves similarly to the Chinese stock market. This is due to the fact that both countries are from Asia, and China plays a major role in the Asian economy. Hence, it contradicts other references and the main finding from the literature review stating other that SVM has better performance. This also indicates that Malaysian stocks are less compatible with the application LR and XGB in a trading strategy. Investors can use LR or XGB in trading strategies if they want, but it is less effective as compared to SVM. SVM is superior to LR because SVM uses regression to find the optimal margin that splits the categories. It helps in reducing the risk of error in the data. Other than that, LR is exposed to the risk of overfitting. SVM is superior to XGB because XGB requires a large dataset or a sufficient number of samples for it to perform itself. Given the outcomes from this study, it can be said that all ML algorithms are able to provide predictability and accuracy to a certain degree in time series forecasting.

IV. CONCLUSIONS

This study is aimed to evaluate different algorithm models, three selected supervised learning algorithms used in stock trading with a total of 1,250 trading days of each data stock with twenty parameters from twenty-eight companies in Bursa Malaysia. The supervised learning models selected are Logistic Regression (LR), Support Vector Machine (SVM), & Extreme Gradient Boosting (XGB) to foresee the direction of the stock prices. By using Walk-Forward Analysis (WFA) method, the trading signal as output was evaluated based on evaluation indicators: Accuracy Rate (AR), Precision Rate (PR), Recall Rate (RR), and F1 Score (F1).

From the perspective of the Malaysian trading strategy, it can be concluded that SVM has a better performance compared to LR and XGB in time series forecasting because it has the highest values in AR, PR, and F1 value at 55.7%, 73.2%, and 65.9%, respectively. Thus, it is consistent with the first finding of the recent related papers and contradicts to second and third findings. On the other hand, LR comes in second as it has the highest RR value at 60.6%, followed by XGB. Note that the value differences for PR are quite large, which is around 13% between SVM and XGB.

There are some limitations to this paper. However, this will provide opportunities for future recommendations. For suggestions, increase the number of stock data because it helps increase the performance of the ML algorithm, as discussed in the paper.

ACKNOWLEDGEMENT

The authors would like to thank the Faculty of Science and Technology, Universiti Sains Islam Malaysia, for this research.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest regarding the publication of this paper.

REFERENCES

[1] BBC News. (2020, April 9). Coronavirus: Worst economic crisis since 1930s depression, IMF says. Retrieved from BBC News Website: https://www.bbc.com/news/business-52236936

[2] Fitzgerald, M. (2021, 8 April). A large chunk of the retail investing crowd started during the pandemic, Schwab survey shows. Retrieved from CNBC Web site: https://www.cnbc.com/2021/04/08/a-large-chunk-of-the-retail-investing-crowd-got-their-start-during-the-pandemic-schwab-survey-shows.html

[3] Tan, V. (2020 , July 16). COVID-19 lockdown stimulates Malaysia's retail investor boom. Retrieved from CNA News Asia Web site: https://www.channelnewsasia.com/news/asia/malaysia-covid-19-lockdown-retail-investor-boom-share-trading-12894640

[4] Lee, Y. N. (2020, August 26). Retail investors with ‘money to play with’ help Malaysian stocks recoup nearly all losses this year. Retrieved from CNBC Website: https://www.cnbc.com/2020/08/26/malaysian-retail-investors-pile-into-stocks-help-market-recoup-losses.html

[5] Bursa Malaysia. (2020). Bursa Malaysia announces RM151.0 million profit-after-tax-and-minority-interest for the first half of 2020 highest first half financial performance since listing in 2005. Retrieved from Bursa Malaysia website: https://www.bursamalaysia.com/about_bursa/media_center/

[6] The Star. (2020). Bursa to consolidate to 1,500-1,530 next week on prolonged bargain-hunting. Retrieved from The Star Website: https://www.thestar.com.my/aseanplus/aseanplus-news/

[7] Wei, N. S. (2020). Bursa: Retail investors need to analyse, assess companies' fundamentals before investing in market. Retrieved from The Edge Markets Web site: https://www.theedgemarkets.com/article/bursa-retail-investors-need-analyse-assess-companies-fundamentals-investing-stock-market

[8] Scott, G., Carr, M., & Cremonie, M. (2016). Literature Review: Technical Analysis-Modern Perspectives. CFA Institute Research Foundation 2016L-1. ISBN 978-1-944960-11-7.

[9] Murphy, J. J. (1999). Study Guide to Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications. Penguin.

[10] Dash, R., & Dash, P. K. (2016). A hybrid stock trading framework integrating technical analysis with machine learning techniques. The Journal of Finance and Data Science, 2(1), 42-57. doi.org/10.1016/j.jfds.2016.03.002

[11] Sagir, A. M., & Sathasivam, S. (2017). The use of artificial neural network and multiple linear regressions for stock market forecasting. Matematika, 33. doi.org/10.11113/matematika.v33.n1.956

[12] Abd Samad, P. H. D., Mutalib, S., & Abdul-Rahman, S. (2019). Analytics of stock market prices based on machine learning algorithms. Indonesian Journal of Electrical Engineering and Computer Science, 16(2), 1050-1058. doi.org/10.11591/ijeecs.v16.i2.pp1050-1058.

[13] Zaini, B. J., Mansor, R., Yusof, N., & Sang, B. H. Classify Stock Market Movement Based on Technical Analysis Indicators Using Logistic Regression.

[14] Shalev-Shwartz, S., & Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.

[15] Lantz, B. (2015). Discover how to build machine learning algorithms, prepare data, and dig deep into data prediction techniques with R. Birmingham: Packt Publishing Ltd.

[16] Jatain, A., & Ranjan, A. (2017). A review study on big data analysis using R studio. International Journal of Computer Science and Mobile Computing, 6(6), 8-13.

[17] Nishida, K. (2017, March 9). Introduction to Extreme Gradient Boosting in Exploratory. Retrieved from Medium Website: https://blog.exploratory.io/introduction-to-extreme-gradient-boosting-in-exploratory-7bbec554ac7

[18] Lv, D., Yuan, S., Li, M., & Xiang, Y. (2019). An empirical study of machine learning algorithms for stock daily trading strategy. Mathematical problems in engineering, 2019. doi.org/10.1155/2019/7816154

[19] Nadh, V. L., & Prasad, G. S. (2018). Support vector machine in the anticipation of currency markets. Int. J. Eng. Technol, 7(2-7), 66.

[20] Butt, S., Ramakrishnan, S., Chohan, M. A., & Punshi, S. K. Prediction of Malaysian Exchange Rate using microstructure fundamental and commodities prices: A machine learning method. International Journal of Recent Technology and Engineering (IJRTE), 8(2).

[21] Liu, C., Wang, J., Xiao, D., & Liang, Q. (2016). Forecasting s&p 500 stock index using statistical learning models. Open journal of statistics, 6(6), 1067-1075. doi.org/10.4236/ojs.2016.66086.

[22] Basak, S., Kar, S., Saha, S., Khaidem, L., & Dey, S. R. (2019). Predicting the direction of stock market prices using tree-based classifiers. The North American Journal of Economics and Finance, 47, 552-567. doi.org/10.1016/j.najef.2018.06.013.

[23] Dey, S., Kumar, Y., Saha, S., & Basak, S. (2016). Forecasting to Classification: Predicting the direction of stock market price using Xtreme Gradient Boosting. doi.org/10.13140/RG.2.2.15294.48968.

[24] Nabipour, M., Nayyeri, P., Jabani, H., Mosavi, A., & Salwana, E. (2020). Deep learning for stock market prediction. Entropy, 22(8), 840. doi.org/10.3390/e22080840.

[25] R Core Team. (2020). R: A Language and Environment for Statistical Computing. Retrieved from R Foundation for Statistical Computing: https://www.R-project.org/