2019 Seventeenth International Conference on ICT and Knowledge Engineering

Stock Closing Price Prediction Using Machine Learning

Pawee Werawithayaset
Department of Computer Engineering
Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
pawee.wer@gmail.com

Suratose Tritilanunt
Department of Computer Engineering
Faculty of Engineering, Mahidol University
Nakorn Pathom, Thailand
suratose.tri@mahidol.ac.th

***Abstract*—this research was prepared to predict the closing price of the stock in the Stock Exchange of Thailand (SET). We are using the Multi-Layer Perceptron model, Support Vector Machine model, and Partial Least Square Classifier to predict the closing price of the stock. In the present, people have more knowledge and understanding of investing in the stock market then the Thai stock market has grown significantly. From the statistical data, we can find the movement of stock prices in that stock market move in a cycle. Form this point; we have the idea that if we can predict the stock price nearby real price. We can be investing at the right time and help investors to reduce investment risks. The experimental result shows that Partial Least Square is the best algorithm of the three algorithms to predict the stock closing price.**

***Keywords*—The Stock Exchange of Thailand (SET), stock, machine learning, Multi-Layer Perceptron (MLP), Partial Least Square (PLS), Sequential Minimal Optimization (SMO), EMA, MACD**

I. INTRODUCTION

The Stock Exchange of Thailand acts as a trading center for listed securities and develop various systems necessary to facilitate the trading of securities [1]. In the present, people have more knowledge and understanding of investing in the stock market. The Thai stock market has grown significantly because many new investors are starting to invest in the Stock Exchange of Thailand (SET). Meanwhile, the variance of the stock market relating to many factors such as the number of people in the stock exchange, foreign exchange rates, economic situation, and capital flow. From the factors mentioned above, therefore, the stock market may have a high risk of investment. Investors without knowledge of the investment may cause damage to the money invested. However, from statistical data, it can be found that the movement of stock prices in that stock market moves in a cycle. At this point we have the idea that if we catch the timing signal, the movement of the stock price has been in range then we can invest at that time. [2]

For selected the timing of the stock price movement correctly. We need to be able to predict the price of the stock in the future. If we cannot predict the future price, there is no way to know when to invest. In order to predict the price of the stock in the future is possible, we use statistical analysis to help predict, because of the price of the stock itself is already a cycle. It is possible for us to catch the rhythm that, at present, the price of the stock is within any part of the cycle. In order to know what direction the stock price go in the future.

II. PREDICTION METHODS

For selecting the algorithms, we choose 3 algorithms to test the ability to predict stock closing prices. We use the same training set and the same test set. The three algorithms such as Multi-layer Perceptron, Sequential Minimal Optimizations, and Partial Least Square.

A. *Multi-Layer Perceptron (MLP)*

Multi-layer Perceptron is a mathematical model that researchers use to process data based on the imitation of the human nervous system. [3] Multi-layer Perceptron is a form of artificial neural network with a multi-layered structure. The Multi-layer Perceptron used for complex tasks with good results with a supervise process and using Backpropagation for training. The Backpropagation consists of 2 sub-sections: Forward Pass and Backward Pass. In the Forward Pass, data pass through the artificial neural network into the data input layer and pass from another layer to another layer until reaching the data output layer. For Backward Pass, the connection weight value adjusted according to the value Error-Correction. Error-Correction is the difference of the actual response with the target response is an error signal that this error is sent back into the artificial neural network in the opposite direction to the connection. The weight of the connection be adjusted until the actual response approaches the target response.

Multi-layer Perceptron consists of 3 significant parts, including the input layer, the hidden layer, the output layer. The input layer is the layer that is used to enter the data that we want to analyze into the input layer. The data in the input layer is forward to the hidden layer. The hidden layers try to convert the data into that layer to be linearly separable before the data sent to the output layer. In the final step, the Hidden layer data must be linearly separable. It may need to use a hidden layer of more than 1 layer to convert the data into a linearly separable form. The activation function of each layer is not necessarily a thresholding function and does not need to be the same function.

978-1-7281-3209-9/19/$31.00 ©2019 IEEE

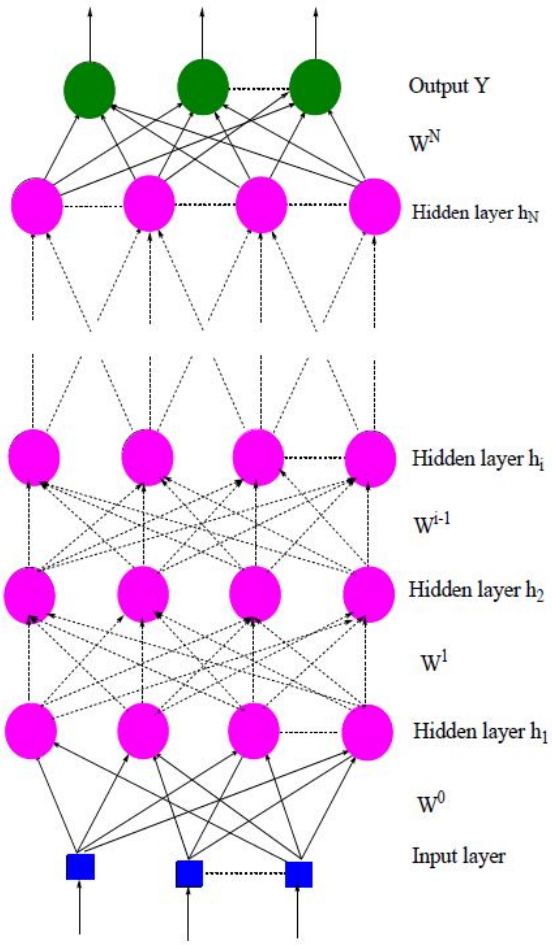

Figure 1. Multi-Layer Perceptron Model. [3]

B. *Sequential Minimal Optimization (SMO)*

Sequential Minimal Optimization Algorithm is an effective method for training Support Vector Machines (SVMs) on classification tasks as well as regression problems. [4] Support Vector Machines are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. The SVMs models generate the regress function by applying a set of high dimensional linear functions. The SVMs are employing to estimate the nonlinear behavior of the forecasting data set because Gaussian kernels tend to give excellent performance under general smoothness assumptions.

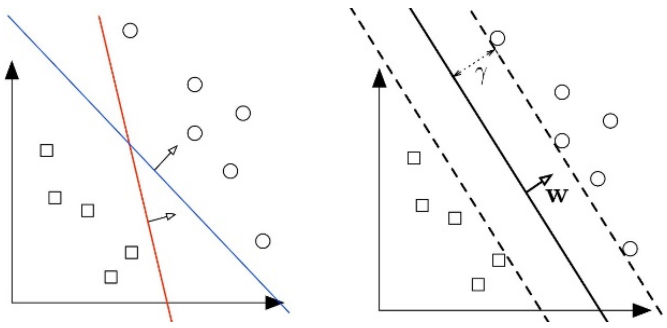

Figure 2. Support Vector Machine Model.[5]

Creating an SVM model selects the linear model that has the best margin distance. It does not choose a linear model that is too close to each vector because it may cause overfitting. If there are new data that are too close together, it makes the predictions error. There are many advantages of SVM, such as the kernel function that can change the dimension of the data to a higher dimension.

C. *Partial Least Square (PLS)*

Partial Least Squares Method is a method for analyzing multiple data in the creation of a relationship model between the observed population groups using the latent variable. [6] PLS has used algorithms in Management of Path Modeling Problems in Economics.

PLS is a method for analyzing data that has received much attention from statisticians and researchers because PLS is able to handle vast amounts of data at the moment with a minimal sample size.

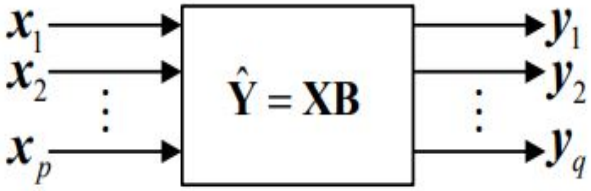

Figure 3. Partial Least Square Model. [7]

Partial Least Square was used to find the relationship between the two-dimension X and Y. The PLS model attempts to find a multi-dimensional direction in the X area, which explains the maximum multi-dimensional variance in the Y area. PLS is especially suitable when the matrix of the predictor has more variables than observation, and when there are many correlations between the X values.

Partial Least Square is one of the Regression techniques. Initially, the Partial Least Square was developed to be able to deal with the problem of regression analysis in the creation of Chemometrics and Spectrometric variables. In the present, Partial Least Square has attracted many statisticians as PLS can handle large amounts of data while there are very few examples. It makes PLS more flexible and able to solve many problems.

III. RESEARCH METHODOLOGY

This research is designed to test the ability to predict stocks. We are using the same training set. We use three algorithms to test the predictability of Multi-layer Perceptron (MLP), sequential minimal optimization algorithm (SMO), and Partial Least Squares Classifier (PLS Classifier).

A. *Data Preparation*

1) Data Selection: In this research, we focus on the trials of 50 stocks form SET50. [8] Therefore, we have to pull out the data of all 50 stocks first. By starting, we have brought the information directly from the Stock Exchange of Thailand (SET) website, which has information on every stock price

that is in the stock market for the past 1 year. We use 100 stocks in the SET100 that do not use the stock in the same duration of a test set. After we get 100 stocks, we combined all 100 stocks for creating the 1 training set. [9]

The raw data format that we have obtained is as follow:

Date	Open	High	Low	Close	Change	%Change	Total Volume (Shares)	Total Value ('000 Baht)
01-12-58	170.00	171.00	169.00	169.00	-2.00	-1.17	3526309.00	599361.48
02-12-58	169.00	171.00	166.00	167.00	-2.00	-1.18	9279658.00	1566102.67
03-12-58	166.00	168.50	166.00	167.00	0.00	0.00	6771799.00	1131959.40
04-12-58	167.00	168.00	166.00	166.00	-1.00	-0.60	11265959.00	1877607.03
08-12-58	165.00	165.50	160.50	160.50	-5.50	-3.31	10213926.00	1657496.39
09-12-58	161.00	161.50	157.00	157.50	-3.00	-1.87	12700080.00	2022292.17
11-12-58	157.00	157.50	153.00	153.50	-4.00	-2.54	8367550.00	1295804.04
14-12-58	150.50	153.00	149.00	152.00	-1.50	-0.98	12570140.00	1896549.39
15-12-58	152.50	154.50	150.50	154.50	2.50	1.64	14404084.00	2201190.99
16-12-58	155.50	157.00	153.50	154.00	-0.50	-0.32	12677956.00	1966075.46
17-12-58	154.00	155.00	153.00	153.00	-1.00	-0.65	21074299.00	3240618.53
18-12-58	152.00	152.50	149.50	150.50	-2.50	-1.63	16811211.00	2536501.04
21-12-58	149.00	152.50	149.00	151.50	1.00	0.66	12697692.00	1920959.48
22-12-58	152.00	153.00	147.00	147.00	-4.50	-2.97	11111842.00	1655141.72
23-12-58	148.00	151.00	148.00	151.00	4.00	2.72	6679497.00	999145.15
24-12-58	152.50	153.00	151.00	151.50	0.50	0.33	4646879.00	704948.30
25-12-58	151.50	152.00	150.00	150.50	-1.00	-0.66	2394769.00	360789.03
28-12-58	151.00	153.50	151.00	152.50	2.00	1.33	4259984.00	649533.03
29-12-58	152.00	152.00	151.00	151.00	-1.50	-0.98	4007675.00	607170.89
30-12-58	152.00	152.50	150.50	150.50	-0.50	-0.33	22749168.00	3436814.43
04-01-59	150.00	150.00	146.00	147.50	-3.00	-1.99	8014150.00	1183980.07
05-01-59	147.50	148.00	145.00	146.00	-1.50	-1.02	8318354.00	1215149.05
06-01-59	146.00	149.00	145.50	148.00	2.00	1.37	5618840.00	828625.22
07-01-59	145.50	146.50	144.00	144.00	-4.00	-2.70	8934136.00	1296899.06
08-01-59	145.00	147.00	145.00	146.50	2.50	1.74	6500285.00	949406.39
11-01-59	145.00	147.50	144.00	147.50	1.00	0.68	8982753.00	1305931.61
12-01-59	148.50	155.00	148.50	154.50	7.00	4.75	14792147.00	2252074.02
13-01-59	158.00	163.00	157.00	163.00	8.50	5.50	22677857.00	3641258.36
14-01-59	160.50	162.50	159.50	159.50	-3.50	-2.15	12313394.00	1980970.47

Figure 4. Raw data from the Stock Exchange of Thailand.

2) Data Cleaning: When we have collected the data, we have to check that the data collected have missing data or not have missing data. For example, it may be due to the fact that the stock has an unusual price movement. Causing the SET to prohibit the trading of that stock on that day or maybe banned for sale the next day, causing the gap of stock price data. We, therefore, need to fill that gap. The method used to fill the gap is to use the average value during the day that the data is missing. After we check the missing value, we proceed to the next step.

3) Data Reduction: After we have finished checking the data, we will use 3 attributes of the data to be used were the highest price, the lowest price, and the closing price. We use only 3 attributes because of :

a) In this research, we focus on the closing price. We do not want the opening price, and we have arranged the data by date, then the date attribute is not required to be used again.

b) Attribute change value and percent change; it a value that tells us is how much today price different from yesterday's price. We have the closing price that arranged by the date. We do not want these attributes in this work.

c) In the attribute, the number of shares traded (stock volume). In this experiment, we focus on the combined model creating from SET100. The model can use to predict every stock. That means we cannot use attributes volume in our research because each of the stocks has a different volume. The amount of shares traded daily is also different too.

d) In the total price value, because each stock price and the volume of each stock are different, therefore not be necessary to use in this work.

4) Data Transformation: After we select 3 attributes, We use these 3 attributes to create other attributes by using the equation to create additional information. Additional attributes that we need are as follows.

a) Pivot Point: Pivot is the average of the day calculated from the average of the highest price, the lowest price, and the closing price. [10] Pivot Point can calculate form following formula:

$$\text{Pivot} = \frac{\text{High}+\text{Low}+\text{Close}}{3}. \quad (1)$$

High is the highestprice of the day. Low is the lowest price of the day.Close is the closing price of the day.

b) Moving Average: Moving average is a calculation of the average price of a stock in a certain period of time. A Moving average that we use is EMA (Exponential Moving Average) [11] following formula:

$$\text{EMA} = [\text{MA}(n-1) + \text{SF}(\text{Pn} - \text{EMA}(n-1)]. \quad (2)$$

MA is a simple moving average, n is a number of days to calculate, Pn is the closing price of that day, and SF is a Smoothing Factor following formula:

$$\text{SF} = 2/(n+1). \quad (3)$$

c) MACD (Moving Average with Convergence Divergence): MACD is used to measure the distance between the Short-term EMA and Long-term EMA to see if the Short-term value is greater than the Long-term value result is a positive value. [12] If the Long-term value is greater than the Short-term value result is a negative value.If the MACD value of the current day is greater than yesterday's value, indicating that the volume is increasing. On the other hand, if the MACD value of the present day is less than yesterday's value, indicating that trading volume has decreased. The formula for finding values is as follows.

$$\text{MACD} = \text{EMA}(\text{Short}) - \text{EMA}(\text{Long}). \quad (4)$$

Short and long means the day that we select to use for calculated. In the standard of MACD, EMA calculated from 12 and 26 days, but if the investor has a trading style that buys quickly, sells quickly, or the period of shares is not long, then can adjust the number of days to a shakedown. In this research, we use 5 and 25 days because we want to predict the short-term closing price. We, therefore, need to choose the least number of days. Then we choose 5 days from the shortest short term and select 25 days from the shortest medium term.

The number of days is the primary day for investors used. In the short term, they are using 5,10,15,20 days. For the medium term, they are using 25,30,50,75 days, and in the long term, they are using 100,200 and 400 days.

d) Signal line : Signal Line is an Exponential Moving Average of MACD. The purpose of the Signal Line line is to specify that the MACD is likely to be an upward trend or downward trend.If the MACD value is greater than the Signal line value, then the MACD is likely to be an upward trend. On the other hand, if the MACD value is less than the Signal line value, the MACD is likely to be a downward trend. The formula for finding values is as follows.

$$\text{Signal line} = \text{EMA (MACD)} \tag{5}$$

Signal line in this research is calculated from EMA of 10 day MACD data.

5) Data Integration: After we convert the data into many attributes, then we have to bring each attribute to add as a set of data and assign a class for analysis. In the data section of the training set for creating the model, we have to combine the data of all 100 stocks from SET100 into one set. To do this, we need to follow the steps of the Data Preparation as well as preparing data for the test set as well.

6) Evaluation: When we get the prediction result to compare which algorithms can predict stock closing price precision. We use Mean Absolute Error (MAE) [13] to test the algorithms, the MAE equation as follows:

$$\text{MAE} = \frac{1}{N}\sum_{i=0}^{N}\left|\bar{y}_i - y_i\right|. \tag{6}$$

We use Price Level Error (PLE) to test the algorithms, the PLE equation as follows:

$$\text{PLE} = \frac{\text{MAE}}{\text{MPL}}. \tag{7}$$

MPL is a Market Price Level. [14]

B. *Algorithm for prediction*

This research uses the Multi-Layer Perceptron (MLP) and Sequential Minimal Optimization (SMO) Module of the Weka program in data classification [15] by setting the parameters as follows.

1) Multi-Layer Perceptron (MLP): To create an MLP model, we define a hidden layer equal to the number of all attributes and class. Give the learning rate at0.3, let the momentum be 0.2 and let training time at 500, which is the basic value for classification. Creating a model, we train with cross-validation use the folds at 10.

2) Sequential Minimal Optimization (SMO): In the creation of the SMO model, we configure the kernel as a Polynomial Kernel, which is the basic kernel for data classification and requires the Filter Type to be normalized.

3) Partial Least Squares Classifier (PLS Classifier): To create the PLS model, we utilizing the PLS Filter's ability to perform predictions.

IV. EXPERIMENT AND RESULT

We create models using training data based on 100 stock data from SET100. We combine all 100 stock data to creating 1 training set. For combining the stock data, we sort stocks with similar prices. If there was a gap between the stock prices, we must fill that gap first. For example, shares A, the final price is 20 baht, and shares B, the final price is 25 baht. According to the Market Price Level principle, at this price level, the share price increase or decrease by 0.1 baht per level. We must start to fill the missing data between both. For the first, we must average the price between both shares prices and added the data by sort the prices from those 2 shares. The price that increases or decreases per price level must more than 0.1 baht per price level. The price added must increase or decrease not higher than 30% of the yesterday price according to the rules of the Stock Exchange of Thailand (SET). We have 31 attributes such as Close Price 5 attributes, Pivot 5 attributes, EMA (5) 5 attributes, EMA (25) 5 attributes, MACD 5 attributes, Signal line 5 attributes, and class 1 attributes.

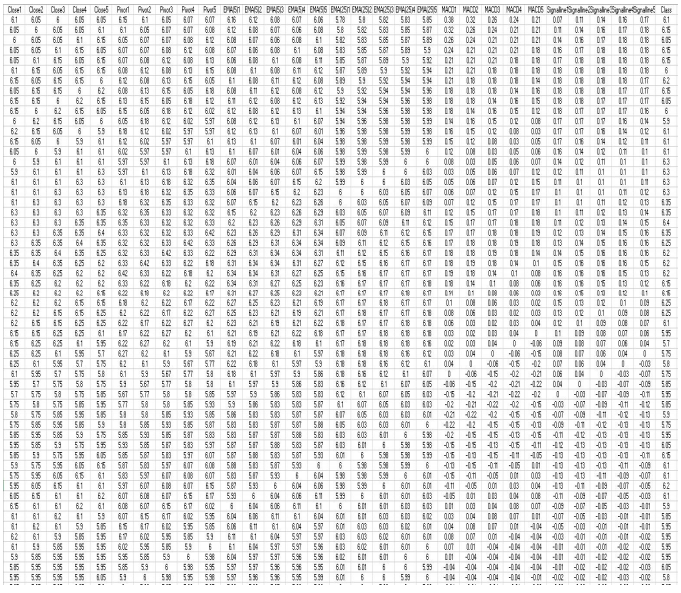

Figure 5. Data for training.

We have 16,922 data for training from SET100, all of 100 stock duration 1 year. After we had received the training set, we create a test set based on 50 stocks from SET50, with 1 stock having 1 test set. In summary, we have 50 test set after we have prepared the data. In the 1 test set, we have 295 data duration for 1 year. We created a model by using a training set with 3 algorithms and get 3 models. After that, we test all 50 test sets with 3 models. Therefore, we get 150 experiment results.

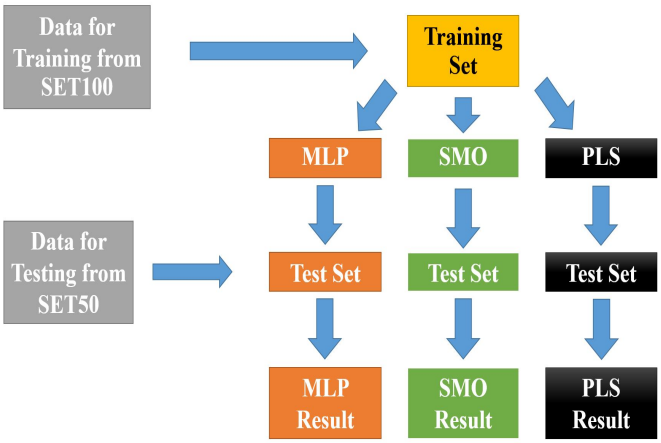

Figure 6. data for training from SET100 and data for testing from SET50.

In the Stock Exchange of Thailand (SET), it has a Market Price Level (MPL). Market Price Level is a Minimum price movement, as prescribed by the SET for securities trading, vary according to each market price level. If we want to know our predictions on how much the error value in the price level is different. We need to calculate the stock price level according to the following in TABLE I.

TABLE I. MARKET PRICE LEVEL

Market Price Level (THB)	Tick sizes (THB) (effective since March 30, 2009)
Less than 2	0.01
2 up to less than 5	0.02
5 up to less than 10	0.05
10 up to less than 25	0.10
25 up to less than 100	0.25
100 up to less than 200	0.50
200 up to less than 400	1.00
400 up	2.00

From the TABLE I., It shows that the price range of each stock is an uneven price movement. The price range change according to the market price level. Therefore we have to divide the price level from the MAE value that can let us know the stock Price Level Error (PLE) how much changed.

Figure 7. Bid and Offer, Bid for the people who want to buy stock and Offer for people who want to sell stock.

From the figure 7 shows that the current price is equal to 2 baht. If the stock price increases and greater than 2 baht, there be tick sizes at 0.02 baht per level, resulting in the price increase to be 2.02, 2.04, 2.06 baht, respectively. On the other hand, if the stock's price decreases and less than 2 baht, there be tick sizes at 0.01 baht per level, resulting in the price decreasing to be 1.99, 1.98, 1.97 baht, respectively. At this point, we use the Price Level Error to show how many deviations from the current price. For example, if the result of the error value is 3 PLE, then the possible price error of that stock be graded from 1.97, 1.98, 1.99, 2.00, 2.02, 2.04, and 2.06. The results of the Prediction shown in TABLE II.

TABLE II. PREDICTION RESULTS

Stock Name	SMO		MLP		PLS		Price Level
	MAE	*PLE*	*MAE*	*PLE*	*MAE*	*PLE*	
ADVANC	1.91	3.81	1.90	3.81	2.00	4.00	0.50
AOT	0.57	2.28	0.64	2.58	0.59	2.37	0.25
BA	0.31	1.22	0.30	1.21	0.31	1.23	0.25
BANPU	0.34	3.41	0.32	3.20	0.35	3.49	0.10
BBL	1.69	1.69	1.74	1.74	1.78	1.78	1.00
BCP	0.42	1.67	0.43	1.72	0.42	1.67	0.25
BDMS	0.22	0.87	0.21	0.86	0.21	0.84	0.25
BEC	0.36	1.44	0.35	1.40	0.35	1.42	0.25
BEM	0.14	1.42	0.17	1.74	0.10	1.03	0.10
BH	2.23	4.47	2.22	4.45	2.26	4.52	0.50
BLA	0.65	2.58	0.67	2.69	0.66	2.63	0.25
BTS	0.12	1.23	0.13	1.27	0.08	0.78	0.10
CBG	1.10	4.41	1.13	4.52	1.18	4.71	0.25
CENTEL	0.53	2.13	0.53	2.13	0.55	2.20	0.25
CK	0.39	1.57	0.40	1.60	0.40	1.59	0.25
CPALL	0.61	2.43	0.65	2.59	0.61	2.44	0.25
CPF	0.46	1.82	0.47	1.86	0.46	1.86	0.25
CPN	0.66	2.65	0.72	2.89	0.69	2.75	0.25
DELTA	1.21	4.83	1.25	5.02	1.25	4.98	0.25
DTAC	0.92	3.69	0.92	3.68	0.93	3.70	0.25
EGCO	1.66	1.66	1.76	1.76	1.69	1.69	1.00
GLOW	0.80	3.20	0.87	3.49	0.85	3.41	0.25
HMPRO	0.16	1.62	0.16	1.59	0.15	1.49	0.10
INTUCH	0.60	2.40	0.66	2.63	0.63	2.51	0.25
IRPC	0.12	2.39	0.20	4.05	0.07	1.31	0.05
ITD	0.14	2.76	0.19	3.84	0.09	1.74	0.05
IVL	0.56	2.25	0.57	2.29	0.57	2.28	0.25
JAS	0.16	3.29	0.21	4.17	0.13	2.69	0.05
KBANK	2.21	4.42	2.24	4.48	2.32	4.64	0.50
KTB	0.20	1.97	0.18	1.83	0.19	1.89	0.10
LH	0.15	1.46	0.14	1.42	0.11	1.14	0.10
M	0.49	1.97	0.58	2.31	0.51	2.05	0.25
MINT	0.48	1.92	0.52	2.06	0.48	1.91	0.25
PTT	4.26	2.13	4.30	2.15	4.46	2.23	2.00
PTTEP	1.24	2.49	1.29	2.58	1.29	2.58	0.50
PTTGC	0.89	3.57	0.92	3.68	0.94	3.77	0.25
ROBIN	0.87	3.49	0.92	3.69	0.92	3.69	0.25
SAWAD	0.67	2.70	0.68	2.73	0.69	2.76	0.25
SCB	1.72	3.44	1.80	3.59	1.76	3.53	0.50
SCC	4.64	2.32	4.87	2.44	4.88	2.44	2.00
SPRC	0.21	2.10	0.19	1.93	0.20	2.00	0.10
TASCO	0.50	2.00	0.48	1.91	0.51	2.05	0.25
TCAP	0.45	1.81	0.51	2.04	0.47	1.86	0.25
TMB	0.11	5.68	0.26	13.14	0.04	2.04	0.02
TOP	0.90	3.60	0.95	3.79	0.92	3.70	0.25
TPIPL	0.11	5.70	0.26	13.00	0.05	2.46	0.02
TTW	0.13	1.29	0.11	1.10	0.10	0.95	0.10
TU	0.24	2.41	0.24	2.42	0.24	2.39	0.10
WHA	0.12	5.86	0.24	12.17	0.05	2.69	0.02
TRUE	0.16	3.25	0.19	3.78	0.13	2.52	0.05
Average	**0.80**	**2.70**	**0.83**	**3.26**	**0.81**	**2.45**	**0.32**

From TABLE II. Mean Absolute Error Results are calculated from 1 year (295 days) historical data.

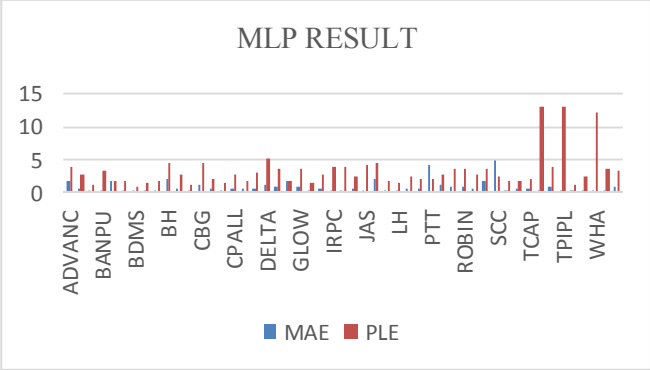

Figure 8. Mean Absolute Error and Price Level Error of Multi-Layer Perceptron results.

1) Multi-layer Perceptron (MLP): Multilayer Perceptron has an average error value of 0.83376, which is considered the average error value of the highest MAE of three algorithms.

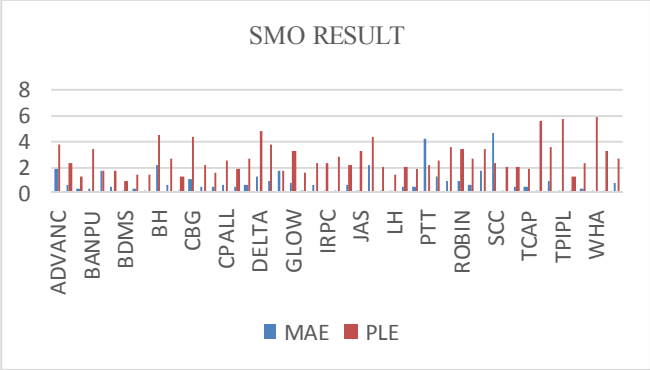

Figure 9. Mean Absolute Error and Price Level Error of Sequential Mimimal Optimization results.

2) sequential minimal optimization (SMO): Sequential Minimal Optimization has an average error of 0.79656, which is considered the least average MAE error value of three algorithms.

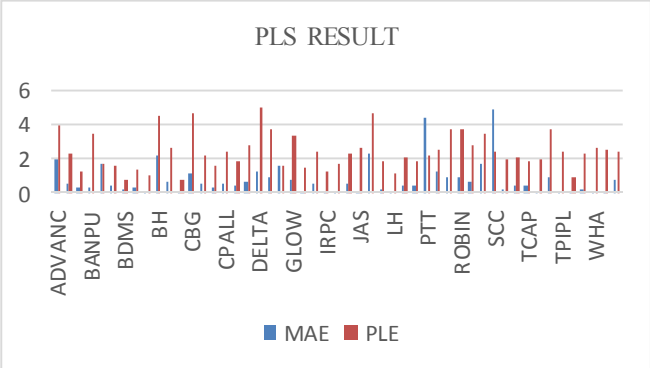

Figure 10. Mean Absolute Error and Price Level Error of Partial Least Square results.

3) Partial Least Squares Classifier (PLS Classifier): The Partial Least Square has an average error of 0.81225, which is considered to be less average MAE error value which is ranked 2nd among 3 algorithms.

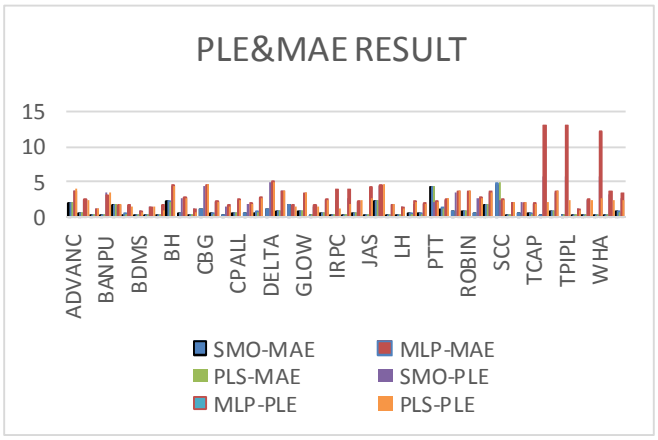

Figure 11. Mean Absolute Error and Price Level Error of three algorithm results.

From Figure. 11, it can see that the result of the Mean Absolute Error from all 3 algorithms shows that the model from SMO gives the least error result. However, if look at the format of the output from the picture shows that the movement of the MAE value from all 3 algorithms moving in a similar direction and tendency.

As for the Price Level Error, show that MLP and SMO have a certain period of very high error values. If we look at the Market Price Level of stocks during the period of high error, it shows that most of them are low-priced stocks with a Market Price Level of only 0.02 baht per level. It shows that MLP and SMO are less able to predict stock with low price range accurately, but PLS does much better in predicting stocks of all price ranges.

V. CONCLUSIONS

This research was prepared to predict the stock closing price of the next day. For helping reduce the investment risk of investors in the Stock Exchange of Thailand (SET). We selected Multi-Layer Perceptron (MLP), Sequential Minimal Optimization (SMO), and Partial Least Square (PLS) algorithms to compare predictive ability. In choosing the appropriate algorithm, we use the error value method to compare the algorithms, such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE). The error values that weka displayed have many values. However, the main thing that we are most interested in is the Mean Absolute Error. The Mean Absolute Error can tell us how much the stock closing price predicted different from the actual stock closing price. After we get the MAE value. We applied the MAE value to create the Price Level Error.

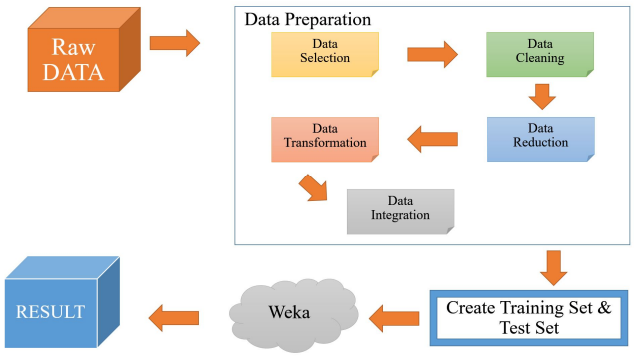

Figure 12. Methodology show that Data Preparation Method.

For the method, we bring the raw data that is the data of the daily stock price — for the first Data Preparation, starting with the Data Selection. We use the stock information from SET100 as a training set and use the data from SET50 as a test set. After that, we start Data Cleaning. Sometimes there are many factors that cause the stock price information to be missing some days. If the stock price has missing data, we must fill that gap before doing the next step. Next, we make a Data Reduction eliminate attributes that are not required or not useful. When we have the only selected attributes, we do the Data Transformation to add the required attributes such as MACD, EMA. Once the desired attributes added, we do Data Integration to include all the attributes arranged to prepare the data to be tested.

The result of prediction by WEKA found that the algorithm it gave the least average error was Sequential Minimal Optimization however when we compare the result of prediction with the stock price level shows that Partial Least Square gives the lowest average error per level.

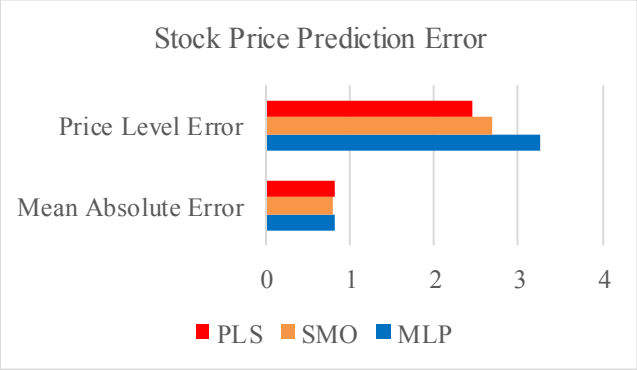

Figure 13. Stock price prediction error results.

By summarizing the results of sorting form ascending as follows:

Price Level Error sort by the least error value.

1. Partial Least Square
2. Sequential Minimal Optimization
3. Multilayer Perceptron

Mean Absolute Error sort by the least error value.

1. Sequential Minimal Optimization
2. Partial Least Square
3. Multilayer Perceptron

For investors, this research can be used to help reduce investment risk by looking at the error value in the stock closing price movement. It is appropriate to invest or not to be invest. This research may help in deciding when to buy the stock on that day. For example, if the predicted value of the closing price is 2 baht. The investor can look at that if the price has gone down to less than 2 baht or it more expensive than 2 baht. The investor should be buying the stock more or do not buy it more.

Based on the results of this experiment, we have seen that when we create a model from the 100 stock price of the SET100 can be used to predict the stock price of all 50 stocks in the SET50, indicating that overall the major stock price have trend and pattern movements that can be used to predict each other.

For future work, we hope that this work can be developed to predict the stock price in the long term. The investor can try using another indicator to test more, such as RSI or stochastic oscillator. To help investors make long-term investment decisions and reduce investor risk.

REFERENCES

[1] The Stock Exchange of Thailand, "History of the Stock Exchange of Thailand", 2008. [Online]. Available: https://www.set.or.th/en/about/overview/history_p1.html [Accessed 21 January 2018].

[2] W. Sirijunyapong, A. Leelasantitham, S. Kiattisin and W. Wongseree, "Predict the stock exchange of Thailand - Set," *The 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE)*, Chiang Rai, 2014, pp. 1-4.

[3] Hassan Ramchoun, Mohammed Amine Janati Idrissi, Youssef Ghanou, Mohamed Ettaouil," Multilayer Perceptron: Architecture Optimization and Training," International Journal of Interactive Multimedia and Artificial Intelligence, Vol. 4, N1

[4] Chan-Yun Yang, Kuo-Ho Su and Gene Eu Jan, "An elaboration of sequential minimal optimization for support vector regression," *2014 IEEE International Conference on System Science and Engineering (ICSSE)*, Shanghai, 2014, pp. 88-93.

[5] Linear Support Vector Machine Model [Online]. Available: https://mubaris.com/posts/svm/ [Accessed 3 Mar 2018].

[6] Z. Zhu, Q. Peng, S. Sun and Z. Li, "Stock trend analysis based on feature rank by partial least squares," *2015 IEEE International Conference on Information and Automation*, Lijiang, 2015, pp. 3003-3008.

[7] Y. Zhang and L. Zhang, "A modified PLS regression model for quality prediction," *Proceeding of the 11th World Congress on Intelligent Control and Automation*, Shenyang, 2014, pp. 1383-1387.

[8] The Stock Exchange of Thailand, "The Stock Data of SET50" [Online]. Available: https://marketdata.set.or.th/mkt/sectorquotation.do?sector=SET50&language=th&country=TH [Accessed 24 January 2018].

[9] The Stock Exchange of Thailand, "The Stock Data of 100" Stock for Create Model [Online]. Available: https://marketdata.set.or.th/mkt/sectorquotation.do?sector=SET100&language=th&country=TH[Accessed 24 January 2018].

[10] Boonpeng and P. Jeatrakul, "Decision support system for investing in stock market by using OAA-Neural Network," *2016 Eighth International Conference on Advanced Computational Intelligence (ICACI)*, Chiang Mai, 2016, pp. 1-6.

[11] T. Gao, Y. Chai and Y. Liu, "Applying long short term memory neural networks for predicting stock closing price," *2017 8th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, 2017, pp. 575-578.

[12] W. Lertyingyod and N. Benjamas, "Stock price trend prediction using Artificial Neural Network techniques: Case study: Thailand stock exchange," *2016 International Computer Science and Engineering Conference (ICSEC)*, Chiang Mai, 2016, pp. 1-6.

[13] Tingwei Gao, Xiu Li, Yueting Chai and Youhua Tang, "Deep learning with stock indicators and two-dimensional principal component analysis for closing price prediction system," *2016 7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, Beijing, 2016, pp. 166-169.

[14] The Stock Exchange of Thailand, "Market Price Level" [Online]. Available: https://www.set.or.th/en/products/trading/equity/tradingsystem_p5.html[Accessed 25 Mar 2018].

[15] V. Mhetre and M. Nagar, "Classification based data mining algorithms to predict slow, average and fast learners in educational system using WEKA," *2017 International Conference on Computing Methodologies and Communication (ICCMC)*, Erode, 2017, pp. 475-479.