SNRU Journal of Science and Technology

Journal home page: snrujst.snru.ac.th

Forecast of Changes in Exchange Rate between Thai Baht and US Dollar Using Data Mining Technique

Sayan Tepdang*, Ratthakorn Pongprasert

Faculty of Business Administration and Information Technology, Rajamangala University of Technology Tawan-Ok, Chakrabongse Bhuvanarth Campus, 122/41 Vipavadi Rd., Dindang, Bangkok, 10400 Thailand

*Corresponding Author: sayan.te@cpc.ac.th

Received: 4 April 2020; **Revised:** 26 June 2020; **Accepted:** 30 June 2020; **Available online:** 1 September 2020

Abstract

Floating exchange rate system and several factors made it hard to forecast changes in exchange rate on a daily basis. However, taking several factors into account can predict changes in exchange rate. Therefore, this study aims to forecast daily changes in exchange rate between Thai Baht and US Dollar by using data mining technique. 9 algorithms were used to forecast: 1) Naive Bayes 2) Generalized Linear Model 3) Logistic Regression 4) Fast Large Margin 5) Deep Learning 6) Decision Tree 7) Random Forest 8) Gradient Boosted Trees, and 9) Support Vector Machine. Each algorithm was tested accuracy by using 10-fold cross validation with train/test ratio following: 90:10, 80:20 to 10:90 respectively.17 factors were used to analyze data for example, exchange rate between Thai Baht to US Dollar, gold prices, US Dollar price index, crude oil price, price stock exchange index in Thailand, USA, Europe, Britain, Japan, and China. Dataset from January 3, 2002 to April 18, 2019 were used to categorize data. The data were collected from the Bank of Thailand, the Federal Reserve Bank of Saint Louis, th.investing.com, and finance.yahoo. The results reported that Logistic Regression was reached the highest accuracy at 64.86% in train/test portion 80:20, Fast Large Margin was reported at 64.66% in train/test portion 80:20 and 90:10, whereas Logistic Regression was exhibited at 64.61% in train/test portion of 70:30. Decision Tree was shown the lowest accuracy at 57.74% in train/test portion of 20:80. Three factors: US Dollar price index, gold price, and Nasdaq price index were respectively reported as the three most significant correlation of changes in exchange rate. The least factor was Nikkei price index. The result shows that the proposed techniques can be used to support exchange risk management and to forecast other foreign exchange rates.

Keywords: Data Mining; Exchange Rate Forecast; Logistic Regression

1. Introduction

Foreign exchange rates, shortly called exchange rates, have a crucial function as the medium of conversing the price of goods and services between countries. Changes in exchange rates have effects on the change in prices of goods and services, demand for foreign currencies, and the value of international economic transactions [1]. Changes in exchange rates are unpredictable since there are so many factors that affect the fluctuation of the exchange rates such as economic fundamentals, monetary policy, fiscal policy, global economy, speculation, domestic and foreign political issues, market psychology, rumors, and technical factors, etc. The exchange rate fluctuation poses a risk to business sectors, in particular, the importers and exporters or those who associate with international businesses. One of the ways to manage their foreign exchange risk is to use Forward, Futures, and Options with commercial banks [2]. However, the researchers suggest that there is another way to predict the changes in foreign exchange rates by using various algorithms techniques.

As in foreign studies, Nagpure [3] suggested to apply Deep Learning models (Support Vector Regression, Artificial Neural Network, Long Short-Term Memory, and Neural Network with Hidden Layers) to predict daily multi-currency exchange rates. The results found that the average accuracy of the predicting model at high level. Similarly, Babu and Reddy [4] used ARIMA, Neural Network and Fuzzy neuron models to predict the daily exchange rates. The result found that the most accurate and stable model was ARIMA. Fuzzy neuron model performs better than neural network model. Whereas, Nayab, Khan, and Mahmud [5] predicted daily exchange rates by applying Cartesian Genetic Programming Evolved Artificial Neural Network (CGPANN). It showed that CGPANN was computationally cost effective and accurate at high level, as it is dependent on least amount of previous data for future data prediction. Meanwhile, Galeshchuk and Mukherjee [6] compared the precision of the technique to predict the direction of the daily exchange rates. It revealed that Deep Convolution Neural Networks was the best technique for predicting the direction of the daily exchange rates. Lastly, Shamar, Hota, and Handa [7] compared regression technique with ensemble regression techniques in view of two ensemble learning: Bagging and Boosting for predict daily exchange rates. The comparative results showed that Regression Ensemble with Least Square Boost performs better than others for one day ahead prediction at the testing stage. However, all 5 studies above agreed that using algorithms were able to predict daily exchange rates, but they did not consider other factors that may affect the exchange rate change, and some studies used not much data to forecast. As the study of Ling, Tsui, and Zhang [8] that employed a combined model of a parametric Markov Logistic model with a nonparametric multilayer Feedforward Neural Network to forecast weekly exchange rates by using the macroeconomic fundamental variables. The results confirmed that the combination models were significantly able to predict exchange rates but this model could not predict daily exchange rates. Most importantly, Rojas and Herman [9] used monthly economic factors to forecast exchange rate by using Linear Regression, Logistic Regression, Support Vector Machine, Support Vector Regression, Gradient Boosting Classifier, and Neural Network. They found that Support Vector Machine was best performance to forecast exchange rate.

In Thailand, Jansod [10] compared models for daily foreign exchange rate forecasting Thai Baht to US Dollar. The results found that the model having the most accurate predictive ability was ARIMA with GARCH-M, followed by ARIMA and Neural Networks models, respectively. Meanwhile Lekkla and Thongkam [11] examined the efficiency of the models to forecast daily foreign exchange rates. The results found that Sequential Minimal Optimization Regression had the smallest error in calculation, but Support Vector Machine Regression was most suitable for predicting exchange rate trend. According to all 2 studies above, there is a lack of other variables that may affect the change in exchange rate in Thaland.

From the aforementioned studies, algorithms were used to forecast the exchange rate in various patterns. Therefore, this study aimed to forecast daily changes in exchange rate between Thai Baht and US Dollar by using data mining technique. Algorithm of this study will be covered several methods and difference from aforementioned studies such as 1) Linear method: Generalized Linear Model, Logistic Regression, Fast Large Margin, and Support Vector Machine, 2) Tree structure method: Decision Tree, Random Forest, Gradient Boosted Trees, 3) Deep learning, and 4) Probability concept through Naïve Bayes. Like, study of Fischer, Krauss, and Treichel [12] that used Multi-Layer Perceptron (MLP), Logistic Regression, Naïve Bayes, K-Nearest Neighbors, Decision Trees, Random Forests, and Gradient-Boosting Trees to forecast exchange rate under simulation. Moreover, the researchers took factors in consider, following: gold prices, US Dollar price index, crude oil price, and price stock exchange index existing in Thailand, USA, Europe, Britain, Japan, and China. The result will be contributed to help support risk management strategies in foreign exchange rates for administrative and authorities, as well as being a benefit to forecast other foreign exchange rates.

2. Materials and methods

The research process shown in Fig. 1 is a research processing method called "Cross-Industry Standard Process for Data Mining" or "CRISP-DM" was used to analyze data. There are 6 steps as follows [13].

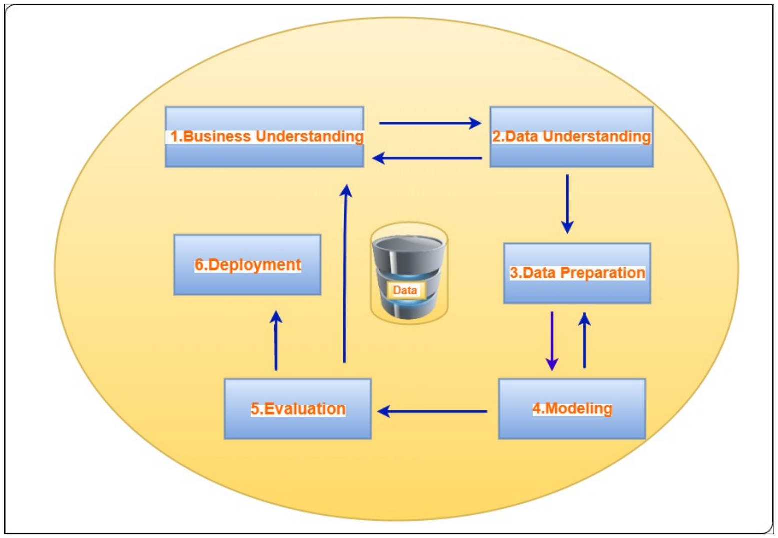

Fig. 1 The Research Process

Business Understanding

Daily dataset related to the foreign exchange market in Thailand was gained. Also, the information on factors influencing foreign exchange rates was also cultivated. In this study, the researchers obtained daily factors that may affect change in exchange rate for example, price of crude oil, gold price, USD index, and price index.

Data Understanding

In this step, daily factors were categorized into 4 parts as follows:

- Foreign exchange rate data of Thai Baht to US dollar were gained from the Bank of Thailand from 3 January 2002 to 18 April 2019 [14].
- West Texas Intermediate (WTI) and Brent crude oil (Brent) were gained from the Federal Reserve Bank of Saint Louis from 3 January 2002 to 18 April 2019 [15].
- US dollar price index, Gold price (XAU/USD), Stock Exchange of Thailand Index (SET index), and Nasdaq Price Index were obtained from th.investing.com from 3 January 2002 to 18 April 2019 [16].
- Dow Jones Price Index, S & P Price Index, New York price index, Hangseng Price Index, Shanghai Price Index, Nikkei Price Index, UK100 price Index, EUROSTOXX 50 price index, and EUROSTOXX 600 price index were obtained from finance.yahoo from 3 January 2002 to 18 April 2019 [17].

Data Preparation

Firstly, the dataset was arranged in a time-series basis within the amount of 3,256 days. Then, all factors were converted according to this numbering suggestion. 1 refers to increasing of dataset whereas 0 refers to the decreasing of the dataset from previous days. An example of how to convert values for forecasting is shown in Fig. 2.

Actual	Difference	Dummy
44.11	-	-
44.06	-0.05	0
44.03	-0.03	0
43.92	-0.11	0
43.95	0.03	1
43.87	-0.08	0
43.89	0.02	1

Fig. 2 Example of how to convert values for forecasting

Modeling (Forecasting)

Firstly, relationship between the exchange rate of Thai Baht to US dollar and factors influencing on exchange rate were investigated. Correlation was weighed for considering which factors have significant effect on changes in exchange rate. Calculation of weighed correlation was performed in equation (1) [18].

$$R = \frac{(X(i) - \overline{X})(Y(i) - \overline{Y})}{(n-1)S_X S_Y} \tag{1}$$

Where:

- R refers to correlation coefficient.
- X, Y refer to attributes of factors/variables.
- N refers to the total number of example.
- S_X, S_Y refer to standard deviation of X and Y.
- i refers to the increment variable of summation.

Secondly, Data mining was used to analyze and forecast daily changes in foreign exchange rate according to 9 algorithms drawn from RapidMiner: 1) Naive Bayes, 2) Generalized Linear Model, 3) Logistic Regression, 4) Fast Large Margin, 5) Deep Learning, 6) Decision Tree, 7) Random Forest, 8) Gradient Boosted Trees and 9) Support Vector Machine. Thirdly, each model was tested accuracy by using 10-fold cross validation with train/test ratio following: 90:10, 80:20 to 10:90 respectively. 3 algorithms were set parameter for analysis, such as 1) Decision Tree used maximal depth = 20, 2) Random forest used number of trees = 20 and Maximal depth = 20, Random forest used number of trees = 20 and maximal depth = 20 and 3) Gradient Boosted Trees used number of trees = 20, maximal depth = 20 and learning rate = 0.01. In the meantime, other algorithms used as a default condition.

Evaluation

Percentage of Accuracy was used to evaluate each algorithm. Calculation of accuracy is shown in equation (2);

$$\text{Accuracy} = \frac{TN + TP}{TN + FP + TP + FN} \tag{2}$$

Where:

- TN refers to True Negative (an outcome where the model correctly predicts the negative class).
- TP refers to True Positive (an outcome where the model correctly predicts the positive class).
- FN refers to False Negative (an outcome where the model incorrectly predicts the negative class).

- FP refers to False Positive (an outcome where the model incorrectly predicts the positive class).
- Accuracy refers to a ratio of correctly predicted observations to the total observations.

Deployment

The research results will be report for administration and authorities to support or develop risk management strategies about exchange rate.

3. Results and Discussion

Results were categorized into 2 sections. Firstly, the data shows relationship between 17 predictors and the exchange rate of Thai Baht to US Dollar. Secondly, the data exhibits percentage of accuracy in forecasting change in exchange rate of Thai Baht and US Dollar of 9 algorithms.

Section 1

The relationship between changes in the 17 factors and change in the exchange rate of Thai Baht to US Dollar was found that US dollar price index was the most significant correlation with the exchange rate at 0.28, Gold price reached 0.24 and Nasdaq Price Index was reported at 0.11 respectively. While the least predictor was the Nikkei price index at 0.02. (shown in table 1)

Table 1 The correlation value between change in predictors and change in the exchange rate of Thai Baht to US Dollar

Predictors	Exchange rate (Thai Baht to US Dollar)
US dollar price index	0.28
Gold price (XAU/USD)	0.24
Nasdaq Price Index	0.11
S & P Price Index	0.11
Dow Jones Price Index	0.10
Stock Exchange of Thailand Index (SET index)	0.09
West Texas Intermediate (WTI)	0.09
US dollar for selling of commercial banks	0.08
US dollar for buying of commercial banks	0.07
UK100 price Index	0.05
Shanghai Price Index	0.05
Hangseng Price Index	0.05
Brent crude oil (Brent)	0.05
New York price index	0.03
EUROSTOXX 600 price index	0.03
EUROSTOXX 50 price index	0.03
Nikkei Price Index	0.01

Table 1 showed that 5 key factors forecasting change in the exchange rate were US Dollar price index, Gold price, Nasdaq Price Index, S&P Price Index and Dow Jones Price Index. These factors put the largest effects on the exchange rate of Thai Baht to US Dollar. This results were similar to; 1) Concept of Lobel [19] arguing that the US Dollar Index was important for traders both as a market in its own right and as an indicator of the relative strength of the US Dollar around the world. 2) Studies of Ibrahim, Kamaruddin, & Hasan [20] together with Beckmann, Czudaj, & Pilbeam [21] found significant relationship between exchange rates and gold prices. And 3) studies of Tsai [22], together with Tepdang et al. [23] found the relationship between stock price index and exchange rate.

Section 2

Percentage of accuracy from 9 algorithms to forecast changes in the exchange rate between Thai Baht and US Dollar was shown in Fig. 3.

Train/ Test	Naïve Bayes		Generalized Linear Model		Gradient Boosted Trees		Fast Large Margin		Logistic Regression		Deep Learning		Decision Tree		Random Forest		Support Vector Machine	
	AC	S.D.	AC	S.D.	AC	S.D.	AC	S.D.	AC	S.D.	AC	S.D.	AC	S.D.	AC	S.D.	AC	S.D.
10:90	61.39	1.03	62.25	1.94	61.95	2.02	61.94	1.98	60.93	1.34	59.11	1.39	59.80	5.66	62.37	2.04	61.65	1.79
20:80	62.26	1.72	63.41	1.65	62.87	3.86	63.55	1.02	63.38	1.62	58.95	2.60	59.34	5.88	63.46	1.98	62.78	3.69
30:70	62.65	0.94	63.87	0.76	63.19	2.29	64.00	0.90	63.81	0.54	60.76	1.42	57.74	5.71	64.22	1.18	63.95	1.20
40:60	62.55	1.28	63.71	1.64	63.92	1.24	63.83	1.38	63.48	1.69	61.49	1.44	61.87	4.85	63.75	1.58	63.94	1.33
50:50	63.34	1.59	64.40	1.07	64.48	0.96	64.37	1.14	64.39	1.01	60.98	1.28	62.64	3.45	64.57	1.22	63.70	1.38
60:40	63.28	0.80	64.28	1.00	63.78	1.10	64.37	1.10	64.14	0.99	61.30	1.54	63.64	0.83	64.31	0.96	63.64	0.93
70:30	63.52	1.46	64.62	1.15	64.36	1.25	64.39	1.31	64.61	1.22	61.86	1.55	64.07	1.24	64.38	1.40	63.96	1.11
80:20	63.18	2.04	64.48	1.71	64.24	1.67	64.66	1.55	64.86	1.89	61.43	2.66	64.11	2.09	64.59	1.46	64.06	2.05
90:10	64.48	2.22	64.31	2.42	64.03	2.86	64.66	2.85	64.43	2.42	60.40	2.26	61.92	4.75	64.19	2.73	63.81	2.45

*note: 1) AC refer to percentage of accuracy
2) S refers to Standard Deviation

Fig. 3 Summarized average accuracies of 9 algorithms

Fig. 3 found that Logistic Regression was reached the highest accuracy at 64.86% (Standard Deviation = 1.89%) in train/test portion of 80:20, Fast Large Margin was reported at 64.66% in train/test portion of 80:20 and 90:10, whereas Logistic Regression was exhibited at 64.61% in train/test portion 70:30. Decision Tree was shown the lowest accuracy at 57.74% in train/test portion of 20:80. An example of result from prediction is shown in Fig. 4. let 1) predict refers to actual value, 2) prediction refers to forecast results, 3) confidence refer to confidential ratio from factors, 4) range1 refers to increasing of exchange rate and 5) range2 refers to decreasing of exchange rate.

predic	prediction(predic)	confidence(range2)	confidence(range1)
range1	range2	0.562	0.438
range2	range2	0.640	0.360
range1	range2	0.654	0.346
range1	range1	0.433	0.567
range1	range1	0.441	0.559
range1	range1	0.410	0.590
range2	range2	0.544	0.456
range1	range1	0.435	0.565
range2	range2	0.611	0.389
range1	range1	0.480	0.520

Fig. 4 Example of Result from prediction

Fig. 4 showed that the confidence ratio between range and range2 performed not much different. It affected the overall forecasting error was reached exceed 35%, and the forecasting efficiency of the exchange rate was decreased accordingly. This may be caused by 1) various in pattern of changes in factors, 2) low of correlation ratio between changes in exchange rate and factors and 3) a daily dataset was very dynamic and complication.

The result from the aforementioned indicates that Logistic Regression, Generalized Linear Model, Support Vector Machine, and Fast Large Margin were the most accurate models. It might be because relationships between the exchange rate and each factor are linear, so the dataset may be suitable for these 4 algorithms [8 – 9, 24 – 26]. Especially, these Regression techniques estimated the effect between dependent variable and independent variables [7].

In the meantime, Decision Tree, Random Forest, and Gradient Boosted Trees were not the methods using linear classifier. Similarly, Naïve Bayes and Deep Learning were also the least accurate for predicting (example of algorithms shows in Fig. 5(b)). This might because non-linear algorithm method reported lower accuracy than linear methods. Due to the fact that the change of dataset is fluctuated and complicated. So that the result cannot be clearly identified and put lower accuracy compare to linear methods. Therefore, non-linear algorithm method become the modeling techniques of choice [12].

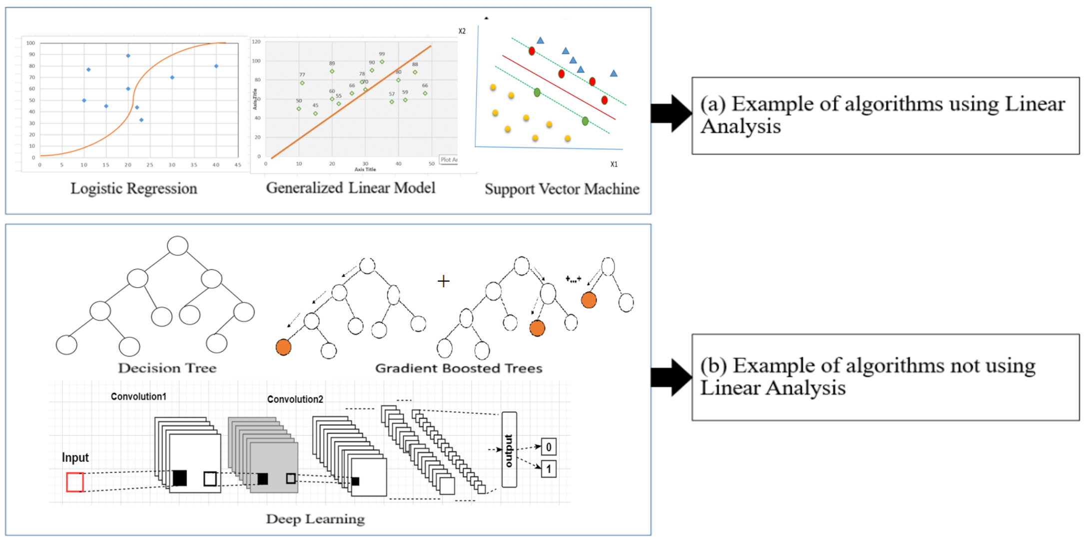

Fig. 5 Concept of Algorithms between Using Linear Analysis and not Using Linear Analysis

4. Conclusion

The research has discovered 2 parts. First, there are 5 factors affecting change in exchange rate (US Dollar price index, Gold price, Nasdaq Price Index, S & P Price Index, and Dow Jones Price Index).

Second, the most accurate algorithms to forecast changes in the exchange rate between Thai Baht and US Dollar was Logistic Regression, and algorithms that use linear methods. They reported better performance than other methods. However, the error of prediction was reached exceed 35%. This may be caused by 1) various pattern of changes in the 17 factors, 2) low correlation ratio between changes in exchange rate and the factors, 3) a daily dataset was very dynamic and complicated.

5. Suggestions

Executives and authorities should use results for supporting of risk management in exchange rate. First, they might develop application used for predicting exchange rate. Secondly, they should enhance dataset for speculating trend of exchange rate. More strongly, they should monitor the changes in 5 factors: US Dollar price index, Gold price, Nasdaq Price Index, S & P Price Index, and Dow Jones Price Index. Due to the data suggested that these factors help risk management in how much foreign currencies should be hold.

Finally, the accuracy of the daily exchange rate change forecast is not very high, it can be used as initial information. The 17 factors should be taken into account for forecasting change in exchange rate. However, future studies should find investigate other possible factors that can be further forecast the change in exchange rate. There should be a new form linear algorithm that may produce better results than existing algorithms used in this research.

6. Acknowledgement

The researchers put a special thank to the Bank of Thailand, the Federal Reserve Bank of Saint Louis, th.investing.com, and finance.yahoo, in supporting secondary data used to analyze in this study.

7. References

[1] K. Muenthaisong, P. Sathapornmongkonlert, T. Leemanonwarachai, Factors Affecting the Fluctuation of the Thai Baht/Malaysian Ringgit Exchange Rate, RMUTL Journal. 6(1) (2018), 100 – 109.

[2] Foreign Exchange Risk Management, https://www.bot.or.th/English/FinancialMarkets/ForeignExchangeMarket/ForeignExchangeHedgingManual/Pages/default.aspx, 21 June 2019.

[3] A.R. Nagpure, Prediction of Multi-Currency Exchange Rates Using Deep Learning, IJITEE. 8(6) (2019), 2278 – 3075.

[4] Y. Zhang, H. Wu, A Comparison of the Prediction Performance by the Linear Model and ARIMA, Model, Master degree, Umea University, Umeå, 2007.

[5] D. Nayab, G.M. Khan, S.A. Mahmud, Prediction of Foreign Currency Exchange Rates Using CGPANN, International Conference on Engineering Applications of Neural Networks, Halkidiki, Greece, 13 – 16 September 2013, 91 – 101.

[6] S. Galeshchuk, S. Mukherjee, Deep Networks for Predicting Direction of Change in Foreign Exchange Rates, Intell Sys Acc Fin Mgmt. 24(4) (2017) 100 – 110.

[7] D.K. Shamar, S.H. Hota, R. Handa, Prediction of Foreign Exchange Rate using Regression Techniques, RBTR. 14(1) (2017) 29 – 33.

[8] J.Z. Ling, A.K. Tsui, Z.Y. Zhang, Forecast Foreign Exchange with both Linear and Nonlinear Models Coupled with Trading Rules for Selected Currencies, The 21st International Congress on Modelling and Simulation (MODSIM2015), Gold Coast Convention and Exhibition Centre, Broadbeach, Queensland, Australia, 29 November – 4 December 2015, 1112 – 1118.

[9] Foreign Exchange Forecasting via Machine Learning, https://pdfs.semanticscholar.org/de5f/6396861fd941f1197f5850d84b6c018bcb6c.pdf?_ga=2.40279394.1952362639.1593010381-28814222.1586343858, 15 June 2020.

[10] A. Jansod, Accuracy Comparison in Foreign Exchange Rate Forecasting Between Neural Networks and ARIMA GARCH-M Models, M.Econ. Economics, Chiang Mai University, Chiang Mai, 2007.

[11] S. Lekkla, J. Thongkam, Forecasting the Trend of Foreign Exchange Rates Using Time Series Analysis Techniques, JOURNAL ITRMU. 5(2) (2018) 94 – 103.

[12] T. Fischer, C. Krauss, A. Treichel, Machine Learning for Time Series Forecasting - a Simulation Study, FAU Discussion Papers in Economics 02/2018, Friedrich-Alexander University Erlangen-Nuremberg, Institute for Economics, 2018.

[13] E. Pacharawongsakda, An Introduction to Data Mining Techniques, first ed., Asia Digital Press Co., Ltd., Bangkok, 2014.

[14] Rates of Exchange of Commercial Banks in Bangkok Metropolis, https://www.bot.or.th/App/BTWS_STAT/statistics/BOTWEBSTAT.aspx?reportID=123&language=ENG, 21 April 2019.

[15] Oil-Economic Data Series, https://fred. stlouisfed.org/tags/series?t=oil, 22 April 2019.

[16] Indices, https:// th.investing. com/indices/, 22 April 2019.

[17] Yahoo Finance, https:// finance.yahoo.com/, 22 April 2019.

[18] Weight by Correlation (RapidMiner Studio Core), https://docs.rapidminer.com/latest/studio/operators/modeling/feature_weights/weight_by_correlation.html, 20 June 2020.

[19] What is the US Dollar Index & Why is it Important?, https://www.dailyfx.com/us-dollar-index/what-is-us-dollar-index.html, 9 April 2020.

[20] S.N. Ibrahim, N.I. Kamaruddin, R. Hasan, The Determinants of Gold Prices in Malaysia, JOAMS. 2(1) (2014), 38 – 41.

[21] J. Beckmann, R. Czudaj, K. Pilbeam, Causality and Volatility Patterns between Gold Prices and Exchange Rates, N Am J Econ Finance. 34 (2015) 292 – 300.

[22] I.C. Tsai, The Relationship between Stock Price Index and Exchange Rate in Asian Markets: A Quantile Regression Approach, J. Int. Financial Mark. Inst. Money. 22(3) (2012) 609 – 621.

[23] S. Tepdang, R. Pongprasert, K. Keawsathuan, S. Wangnum, An Investigation of Association of SET Index between Exchange Rate and Traders by Using Data Mining Technique, Sri-Ayutthaya Cluster 9th Rajabhat University National Conference, Valaya Alongkorn Rajabhat University under the Royal Patronage. 18 – 19 October 2018, 1017 – 1026.

[24] Generative and Discriminative Classifier: Naïve Bayes and Logistic Regression, https://www.cs.cmu.edu/~tom/mlbook/NBayesLogReg.pdf, 8 April 2020.

[25] Support Vector Machines(SVM) — An Overview, https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989, 8 April 2020.

[26] G. Yuan, C. Ho, C. Lin, Recent Advances of Large-Scale Linear Classification, Proc. IEEE. 100(9) (2012) 2584 – 2603.