Trading Decision Models based on the Index Movement Analysis of East Asian Stock Exchange Markets

Review of Integrative Business & Economics Research

Kittisak Kerdprasop*
Knowledge Engineering Research Unit, School of Computer Engineering,
Suranaree University of Technology, Nakhon Ratchasima, Thailand

Nittaya Kerdprasop
Knowledge Engineering Research Unit, School of Computer Engineering,
Suranaree University of Technology, Nakhon Ratchasima, Thailand

ABSTRACT

We present in this work the application of data mining to analyze and predict the index movement of the Stock Exchange of Thailand (SET). The daily SET index movement prediction of our analysis is either stable, increase, or decrease. Our short-term prediction models are derived based on the co-movement analysis of two major stock markets of East Asia: Nikkei and Hong Kong Hang Seng (HSI). The opening and closing prices of the Nikkei and HSI are used to predict the direction of SET index. The models are created from different schemes: the same-day analysis, date-lagging analysis using the same train-test data, and date lagging analysis using separate train-test data. The models from these three strategies are demonstrated in this work. From our experimentation, the most accurate model is the one built from both lagging and current data of Nikkei and HSI indexes. Our best model can predict the SET index movement with 72.73% accuracy using a separate set of test data.

Keywords: Decision model, Co-movement analysis, Stock market, Data mining.

1. INTRODUCTION

In a current globalization world, domestic economy in most countries, including Thailand, has been inevitably impacted by the fluctuation in the international financial markets. The increasing significance of international movement to the domestic motion has been reported in many research works (Liao and Chou, 2013; Wang, 2014; Aloui and Hkiri, 2014; El Alaoui et al., 2015; Burzala, 2016). The co-occurrence movement had previously been studied with the time series analysis technique (Cavalcante et al., 2016). Since the widespread of machine learning and data mining technologies successfully applied to many application areas, financial and business analysts have gradually adopted this new technology as their beneficial tool. The advantage of machine learning and data mining to the data analysis task is their intelligent search for the optimal solution to a specified problem. Intelligence is an important factor in modern business analytics.

During the last decades, machine learning technique that has been applied extensively in business and other fields is artificial neural network, including its various

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

extensions (Oh and Kim, 2002; Lam, 2004; Chen and Du, 2009; Dai et al., 2012; Vanstone et al., 2012; Yoshihara et al., 2014; Gocken et al., 2016). Recently, an efficient and accurate learning technique using support vector machine has been introduced and rapidly adopted by researchers and practitioners to perform analytics and modeling tasks (Chen, 2010; Chao et al., 2012; Lin et al., 2013; Nayak et al., 2015). For data exploration and knowledge discovery tasks, data mining technology has alternatively employed. The various data mining tasks used in financial and other business applications include clustering (Aghabozorgi and The, 2014), decision tree induction (Wu et al., 2006), and other intelligent techniques (Chourmouziadis and Chatzoglou, 2016).

The superiority of data mining technique over conventional methods is the accurate modeling and at the same time the easy-to-apply of the obtained model. From the ease-of-use point of view to serve general users, the final product of modeling is thus presented as a small set of simple rules (Yu at al., 2013). To get such a simplified model, reducing number of predicting variables is therefore an important step to be applied prior to the modeling phase. Statistical methods such as principal component analysis (Sopipan et al., 2012; Guo et al., 2015; Wang and Wang, 2015) and independent component analysis (Lu et al., 2009) are normally used by data analysts. The major drawback of these statistical feature selection methods is the transformation process that makes data attributes changing their forms and values causing difficulty to interpretation and understanding. Recent research works are thus trying to discover new feature selection techniques that do not alter data format and content (Tsai and Hsiao, 2010; Lin et al., 2014; Kerdprasop and Kerdprasop; 2016).

We, therefore, present in this paper the new feature construction method instead of the transformation of the existing data attribute values. The constructed new features are lagged data of stock market opening and closing values. We consider the two major stock markets of Asia, that are, Nikkei and Hong Kong Hang Seng. The detail of data used in our research is described in Section 2. The model building process, in which we pay particular attention to the decision tree model because of its comprehensibility, is explained in Section 3. The experimentation to confirm model efficiency is illustrated in Section 4. We conclude our research work in Section 5.

2. DATA SET CHARACTERISTIC

The data used in this work are opening and closing indexes from the three stock exchange markets: Stock Exchange of Thailand (SET), Hong Kong Hang Seng (HSI), and Japan Nikkei 225. Stock indexes used in this study are during September 28, 2016 to November 14, 2016. Data are taken from the site https://www.google.com/finance. The 30 opening-day records that are the same date among the three markets are selected. That means for some specific date that was the holiday in some country, that date would not be used in our analysis for building a predictive model. The SET index movement (as either stable, increase, or decrease) is our analysis target, while the HSI and Nikkei are predictors. Data attributes and their meanings are summarized in Table 1.

The original data taken from the source contain only 6 attributes (or fields, features, variables), that are, the opening and closing prices of SET, HSI, and Nikkei. We

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

firstly construct other 3 attributes to represent movement of each stock market. These fields are named SIGN_CAL_SET, SIGN_CAL_HANG, and SIGN_CAL_NI. The movement of each stock market is computed by subtracting the closing price of current date from the previous day closing price, and then making it be percentage comparative to the current date closing price. If the price difference is less than 0.5%, the SIGN_CAL_ will be set to 0 (to signal stable or small change). If the index price decreases to 0.5% or lower, the SIGN_CAL_ will be set as -1 (to signal negative change). The last possibility is that the index price increases to 0.5% or higher; for this case, the SIGN_CAL_ will be set as +1 (to signal positive change).

We then also compute closing-price difference of the current date comparative to the previous two-day closing price. There are thus 3 more attributes, named, SIGN_CAL_2_SET, SIGN_CAL_2_HANG, and SIGN_CAL_2_NI. The opening price of each market are also lagged 1 and 2 days. These 6 lagged attributes are SET_open_1, SET_open_2, HANG_open_1, HANG_open_2, NI_open_1, and NI_open_2. Totally, there are 17 predictive attributes and 1 target attribute, which is SIGN_CAL_SET. Figure 1 shows example of 10 data instances. Data distribution for opening and closing prices of the three stock markets is also provided in Figure 2.

Table 1. Data attributes and their meanings.

Attribute name	Meaning
OPEN_HANG	Opening price of Hong Kong Hang Seng (HSI) stock market
CLOSE_HANG	Closing price of HSI stock market
SIGN_CAL_HANG	Signal calculation of HSI closing price compared to the previous day closing price. Possible values: 0 (small to no change), -1 (negative change), +1 (positive change).
SIGN_CAL_2_HANG	Signal calculation of HSI closing price compared to the previous two-day closing price. Possible values: 0 (small to no change), -1 (negative change), +1 (positive change).
HANG_open_1	Opening price on the previous day of HSI stock market
HANG_open_2	Opening price on the previous two-day of HSI stock market
OPEN_NI	Opening price of Nikkei stock market
CLOSE_NI	Closing price of Nikkei stock market
SIGN_CAL_NI	Signal calculation of Nikkei closing price compared to the previous day closing price. Possible values: 0 (small to no change), -1 (negative change), +1 (positive change).
SIGN_CAL_2_NI	Signal calculation of Nikkei closing price compared to the previous two-day closing price. Possible values: 0 (small to no change), -1 (negative change), +1 (positive change).
NI_open_1	Opening price on the previous day of Nikkei stock market
NI_open_2	Opening price on the previous 2-day of Nikkei stock market
OPEN_SET	Opening price of Thailand SET stock market
CLOSE_SET	Closing price of SET stock market
SIGN_CAL_SET	Signal calculation of SET closing price compared to the previous day closing price. Possible values: 0 (small to no change), -1 (negative change), +1 (positive change).

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

	This is the target attribute for modeling.
SIGN_CAL_2_SET	Signal calculation of SET closing price compared to the previous two-day closing price. Possible values: 0 (small to no change), -1 (negative change), +1 (positive change)
SET_open_1	Opening price on the previous day of SET stock market
SET_open_2	Opening price on the previous two-day of SET stock market

Table | Annotations

	OPEN_HANG	CLOSE_HANG	SIGN_CAL_HANG	SIGN_CAL_2_HANG	HANG_open_1	HANG_open_2	OPEN_NI	CLOSE_NI	SIGN_CAL_NI
1	22622.460	22531.090	1.000	0.000	23048.320	22981.910	17526.6...	17374.790	1.000
2	23048.320	22415.190	-1.000	1.000	22981.910	22682.710	17281.9...	16251.540	-1.000
3	22981.910	22909.470	0.000	0.000	22682.710	22611.540	17242.7...	17171.380	0.000
4	22682.710	22801.400	1.000	-1.000	22611.540	22708.550	17126.0...	17177.210	1.000
5	22611.540	22642.620	0.000	-1.000	22708.550	22946.060	16964.5...	16905.360	-1.000
6	22708.550	22683.510	-1.000	1.000	22946.060	23015.060	$null$	$null$	$null$
7	22946.060	22810.500	-1.000	0.000	23015.060	22845.820	17238.0...	17134.680	-1.000
8	23015.060	23147.070	1.000	-1.000	22845.820	23088.830	17380.5...	17442.400	0.000
9	22845.820	22934.540	0.000	-1.000	23088.830	23347.190	17360.8...	17425.020	0.000
10	23088.830	22954.810	-1.000	-1.000	23347.190	23396.140	17448.2...	17446.410	1.000

SIGN_CAL_2_NI	NI_open_1	NI_open_2	OPEN_SET	CLOSE_SET	SIGN_CAL_SET	SIGN_CAL_2_SET	SET_open_1	SET_open_2
0.000	17281.950	17242.700	1508.620	1494.530	-1.000	1.000	1498.810	1504.490
1.000	17242.700	17126.030	1498.810	1509.430	0.000	1.000	1504.490	1496.060
-1.000	17126.030	16964.500	1504.490	1509.840	1.000	0.000	1496.060	1489.400
-1.000	16964.500	17238.000	1496.060	1502.270	1.000	0.000	1489.400	1499.140
0.000	17238.000	17380.540	1489.400	1485.700	0.000	0.000	1499.140	1502.190
$null$	$null$	$null$	1499.140	1493.080	0.000	1.000	1502.190	1500.150
0.000	17380.540	17360.890	1502.190	1498.650	0.000	0.000	1500.150	1494.040
1.000	17360.890	17448.250	1500.150	1504.520	1.000	0.000	1494.040	1500.720
0.000	17448.250	17346.390	1494.040	1495.720	0.000	0.000	1500.720	1492.200
0.000	17346.390	17329.030	1500.720	1494.440	0.000	-1.000	1492.200	1505.990

Figure 1. Example of 10 data instances containing 18 attributes.

Field	Graph	Measurement	Min	Max	Mean	Std. Dev	Skewness
OPEN_SET		Continuous	1387.030	1513.180	1487.840	27.256	-2.359
CLOSE_SET		Continuous	1406.180	1513.860	1487.774	26.952	-2.078
Field	**Graph**	**Measurement**	**Min**	**Max**	**Mean**	**Std. Dev**	**Skewness**
OPEN_HANG		Continuous	22611.540	24013.840	23289.877	390.641	-0.035
CLOSE_HANG		Continuous	22415.190	23952.500	23243.747	412.312	-0.231
Field	**Graph**	**Measurement**	**Min**	**Max**	**Mean**	**Std. Dev**	**Skewness**
OPEN_NI		Continuous	16474.450	17526.610	17037.632	284.467	-0.070
CLOSE_NI		Continuous	16251.540	17446.410	17014.554	317.453	-0.448

Figure 2. Distribution for 30-day opening and closing prices of SET, HSI, and Nikkei

3. CO-MOVEMENT ANALYSIS METHOD

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

The prepared stock data as explained in the previous section are then used for building the predictive model. The attribute SIGN_CAL_SET has been set its role as a target, and the other 17 attributes are eligible for being used as predictors. Some data instances containing null value are removed. The final data set is comprised of 30 instances and 18 attributes. We build predictive models using three different schemes of training and testing data. The final 3 models and then compared on their predictive accuracy. The most accurate model is then reported as the best trading model for general traders to make decision regarding to the movement of SET index on that day. The analysis process is demonstrated as a flow chart in Figure 3. The analysis tool to implement our modeling scheme is IBM SPSS Modeler version 14 as shown in Figure 4.

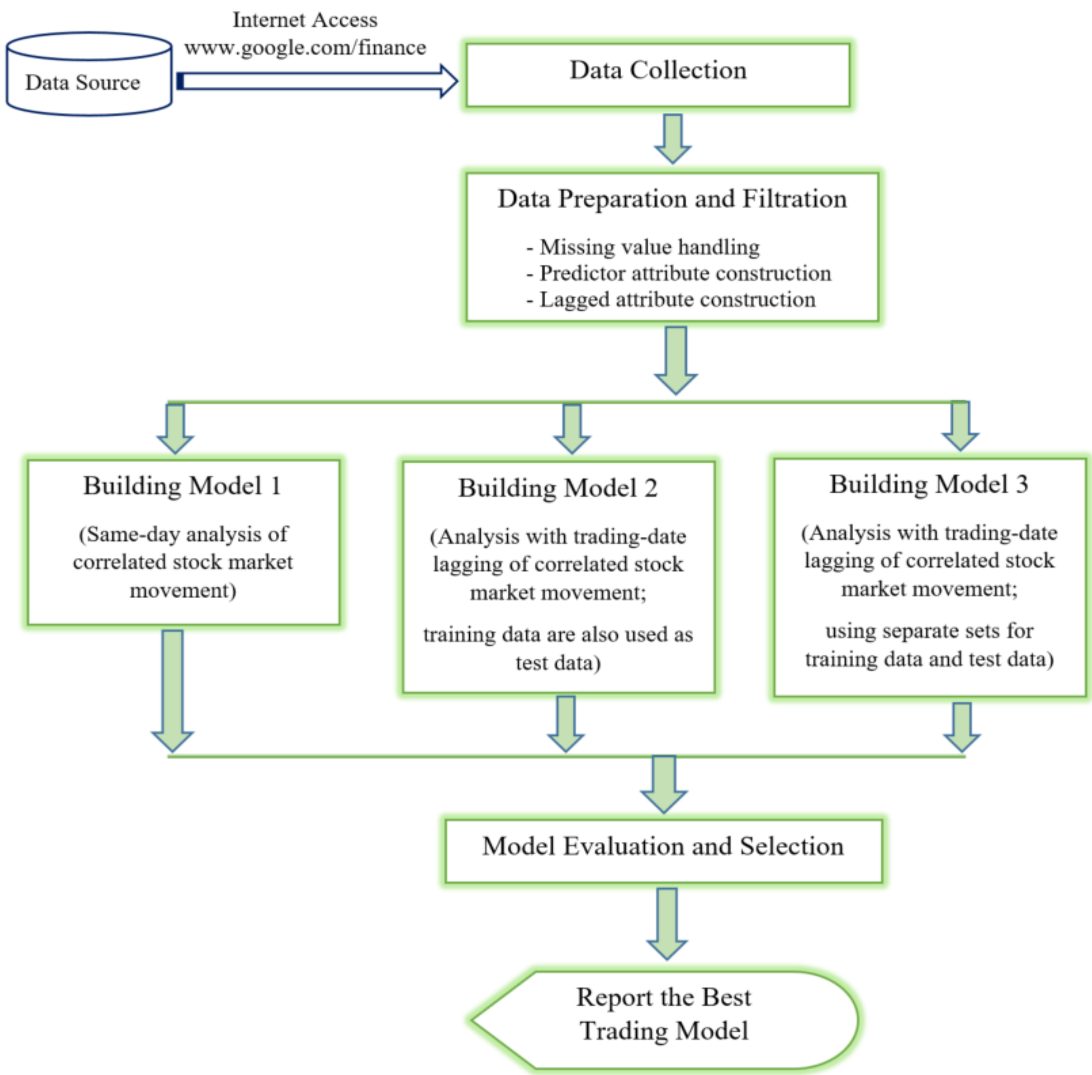


Figure 3. The steps to build trading model.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

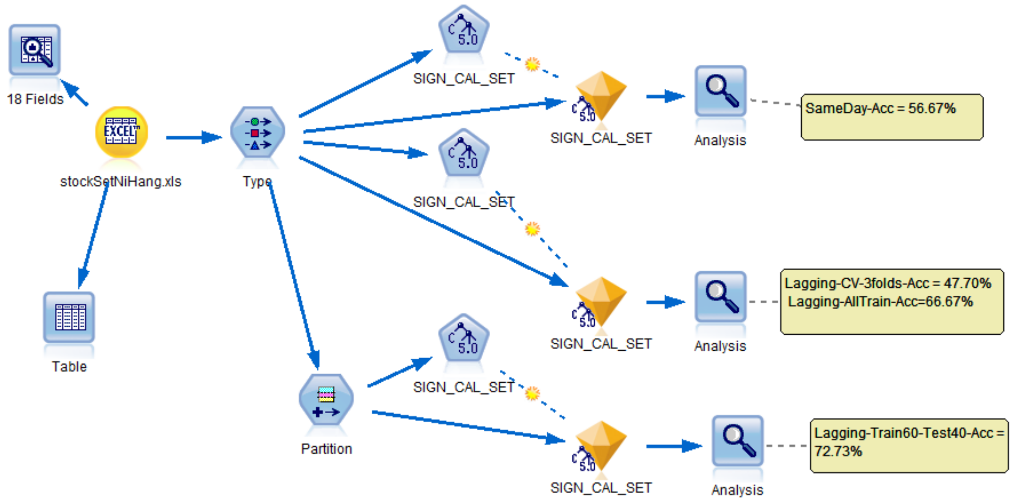
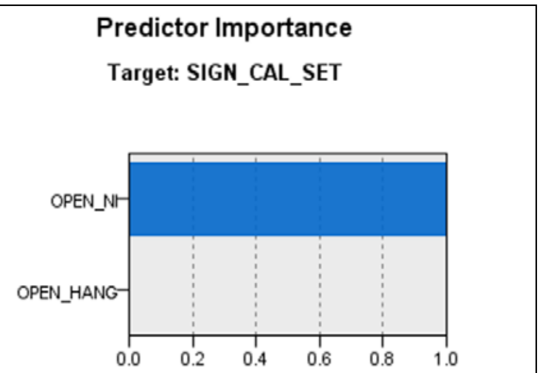


Figure 4. Implementation of the designed process with IBM SPSS Modeler.

4. ANALYSIS RESULTS

Upon implementing the designed modeling process, the three predictive models for predicting SET index movement can be shown as in Figure 5. Predictive model and importance of predictors are also given in the figure. We then evaluate the obtained three trading models by comparing the actual SET index movement (as either not change, positive change, or negative change) against the index movement outcomes predicted by the models. It turns out that the SET index movement prediction model using trading-date lagging strategy (model 3) shows the best performance of 72.73% predictive accuracy when tested with a separate test set. Model 2, which also uses trading-date lagging scheme but tested with training data, shows the predictive performance as accurate as 66.67%. The model 2 when tested with 3-fold cross validation, its accuracy drops to 47.70%. The low performance may due to the small set of training data. Model 1 that has been built from the same-day trading data of HSI and Nikkei markets without considering information from the previous date shows predictive performance at 56.67% accurate. The best trading model is thus the one with lagging scheme, that is, model 3. The model and its meaning are given in Figure 6.

Model 1 (Same-day trading analysis)
 	
OPEN_HANG <= 23859.400
    OPEN_NI <= 16,751 => **1.0**
    OPEN_NI > 16,751 => **0.0**
OPEN_HANG > 23859.40 => **-1.0**
 	
(Accuracy = 56.67%)
	

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

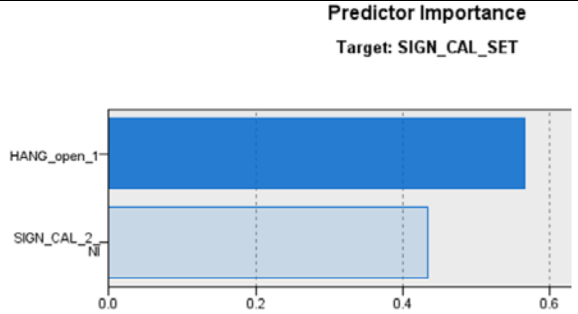
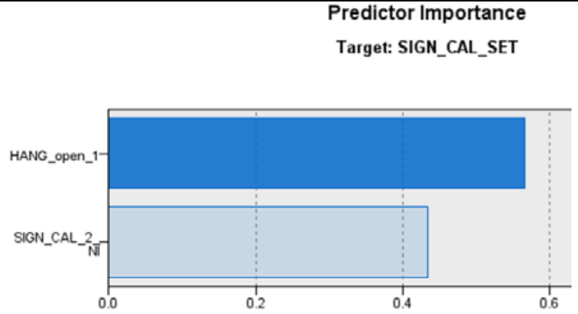
Model 2 (Analysis with lagging scheme, use all train data) HANG_open_1 <= 23665.30 SIGN_CAL_2_NI <= -1 => **1.0** SIGN_CAL_2_NI > -1 => **0.0** HANG_open_1 > 23665.30 => **-1.0** (Accuracy = 66.67%)	
Model 3 (Analysis with lagging scheme, use separate train-test data) HANG_open_1 <= 23665.30 SIGN_CAL_2_NI <= -1 => **1.0** SIGN_CAL_2_NI > -1 => **0.0** HANG_open_1 > 23665.30 => **-1.0** (Accuracy = 72.73%)	

Figure 5. Trading models built from three different feature selection schemes.

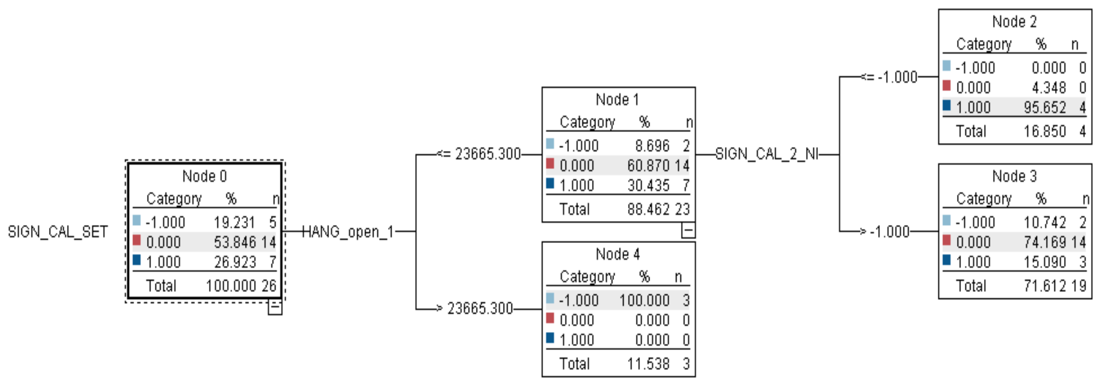


Figure 6. The most accurate trading model built from trading-date lagging scheme.

Meaning of the best SET trading model is as follows:

(1) If (opening price of HSI on the previous day was less than or equal to 23665.30)
AND
(calculated signal of Nikkei on the previous two days was negative change)
Then
SET index movement today shall be <u>positive change</u>.

(2) If (opening price of HSI on the previous day was less than or equal to 23665.30)

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

AND
(calculated signal of Nikkei on the previous two days was positive or unchanged)
Then
SET index movement today shall be <u>slightly or not change</u>.

(3) If (opening price of HSI on the previous day was higher than 23665.30)
Then
SET index movement today shall be <u>negative change</u>.

5. CONCLUSION

We present in this work the Thailand stock market movement prediction based on the analysis of previous Thailand stock market movement and the other two major markets of East Asia: Nikkei and Hong Kong Hang Seng. The main target for our prediction is the forecasting of SET closing price direction as either unchanged, increase, or decrease, comparative to the SET closing price on the previous day. To build the predictive model, we use opening and closing prices of SET, Nikkei, and HSI stock markets to train the tree-based induction learning algorithm. During the data preparation phase, which is the steps prior to the core learning phase, we also construct additional features using opening and closing prices that are the lagging of one to two days. Closing price difference computed from the previous day comparison has also been used as movement signal. This signal reflects three different directions of the market: unchanged, positive change, and negative change. The final tree-based model induced from our feature construction method has been evaluated on a hold out test set. Its predictive performance is 72.73% accurate. This level of predictive accuracy is satisfied. We, nevertheless, plan to improve predictive performance with additional background information and supplement the model with some advanced techniques.

REFERENCES

[1] S. Aghabozorgi and Y.W. The, "Stock market co-movement assessment using a three-phase clustering method," Expert Systems with Applications, vol. 41, no. 4, pp. 1301-1314, 2014.
[2] C. Aloui and B. Hkiri, "Co-movements of GCC emerging stock markets: new evidence from wavelet coherence analysis," Economic Modelling, vol. 36, pp. 421-431, 2014.
[3] M.M. Burzala, "Contagion effects in selected European capital markets during the financial crisis of 2007-2009," Research in International Business and Finance, vol. 37, pp. 556-571, 2016.
[4] R.C. Cavalcante, R.C. Brasileiro, V.L.F. Souza, J.P. Nobrega, and A.L.I. Oliveira, "Computational intelligence and financial markets: a survey and future directions," Expert Systems with Applications, vol. 55, pp. 194-211, 2016.
[5] H. Chao, H. Li-li, and H. Ting-ting, "Financial time series forecasting based on wavelet kernel support vector machine," Proceedings of the 8th IEEE International Conference on Natural Computation, pp. 79-83, 2012.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

[6] J. Chen, "SVM application of financial time series forecasting using empirical technical indicators," Proceedings of the IEEE International Conference on Information Networking and Automation, pp. V1-77 – V1-81, 2010.
[7] W.-S. Chen and Y.-K. Du, "Using neural networks and data mining techniques for the financial distress prediction model," Expert Systems with Applications, vol. 36, no. 2, pp. 4075-4086, 2009.
[8] K. Chourmouziadis and P.D. Chatzoglou, "An intelligent short term stock trading fuzzy system for assisting investors in portfolio management," Expert Systems with Applications, vol. 43, pp. 298-311, 2016.
[9] W. Dai, J.-Y. Wu, and C.-J. Lu, "Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes," Expert Systems with Applications, vol. 39, no. 4, pp. 4444-4452, 2012.
[10] A.O. El Alaoui, G. Dewandaru, and S.A. Rosly, "Linkages and co-movement between international stock market returns: case of Dow Jones Islamic Dubai Financial Market index," Journal of International Financial Markets, Institutions & Money, vol. 36, pp. 53-70, 2015.
[11] M. Gocken, M. Ozcalici, A. Boru, and A.T. Dosdogru, "Integrating metaheuristics and artificial neural networks for improved stock price prediction," Expert Systems with Applications, vol. 44, pp. 320-331, 2016.
[12] Z. Guo, H. Wang, J. Yang, and D.J. Miller, "A stock market forecasting model combining two-directional two-dimensional principal component analysis and radial basis function neural network," PLoS ONE, vol. 10, no. 4, article e0122385, 19 pp., 2015.
[13] K. Kerdprasop, and N. Kerdprasop, "A short term co-movement detection of financial data," Proceedings of the 2nd IEEE International Conference on Computer and Communications, pp. 42-45, 2016.
[14] M. Lam, "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis," Decision Support System, vol. 37, no. 4, pp. 567-581, 2004.
[15] S.-H. Liao and S.-Y. Chou, "Data mining investigation of co-movements on the Taiwan and China stock markets for future investment portfolio," Expert System with Applications, vol. 40, no. 5, pp. 1542-1554, 2013.
[16] Y. Lin, H. Guo, and J. Hu, "An SVM-based approach for stock market trend prediction," Proceedings of the 2013 International Joint Conference on Neural Networks, pp. 1-7, 2013.
[17] F. Lin, D. Liang, C.-C. Yeh, and J.-C. Huang, "Novel feature selection methods to financial distress prediction," Expert Systems with Applications, vol. 41, no. 5, pp. 2472-2483, 2014.
[18] C.-J. Lu, T.-S. Lee, and C.-C. Chiu, "Financial time series forecasting using independent component analysis and support vector regression," Decision Support Systems, vol. 47, no. 2, pp. 115-125, 2009.
[19] R.K. Nayak, D. Mishra, and A.K. Rath, "A naïve SVM-KNN based stock market trend reversal analysis for Indian benchmarks indices," Applied Soft Computing, vol. 35, pp. 670-680, 2015.
[20] K.J. Oh and K. Kim, "Analyzing stock market tick data using piecewise nonlinear model," Expert Systems with Applications, vol. 22, pp. 249-255, 2002.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)

[21] N. Sopipan, W. Kanjanavajee, and P. Sattayatham, “Forecasting SET50 index with multiple regression based on principal component analysis,” Journal of Applied Finance & Banking, vol. 2, no. 3, pp. 271-294, 2012.
[22] C.-F. Tsai and Y.-C. Hsiao, “Combining multiple feature selection methods for stock prediction: union, intersection, and multi-intersection approaches,” Decision Support Systems, vol. 50, no. 1, pp. 258-269, 2010.
[23] B. Vanstone, G. Finnie, and T. Hahn, “Creating trading systems with fundamental variables and neural networks: the Aby case study,” Mathematics and Computers in Simulation, vol. 86, pp. 78-91, 2012.
[24] Y. Wang, “Stock price direction prediction by directly using prices data: an empirical study on the KOSPI and HSI,” International Journal of Business Intelligence and Data Mining, vol. 9, no. 2, pp. 145-160, 2014.
[25] J. Wang and J. Wang, “Forecasting stock market indexes using principle component analysis and stochastic time effective neural networks,” Neurocomputing, vol. 156, pp. 68-78, 2015.
[26] M.-C. Wu, S.-Y. Lin, and C.-H. Lin, “An effective application of decision tree to stock trading,” Expert Systems with Applications, vol. 31, pp. 270-274, 2006.
[27] Yoshihara, K. Fujikawa, K. Seki, and K. Uehara, “Predicting stock market trends by recurrent deep neural networks,” Proceedings of the PRICAI 2014 Trends in Artificial Intelligence, pp. 759-769, 2014.
[28] H. Yu, G.V. Nartea, C. Gan, L.J. Yao, “Predictive ability and profitability of simple technical trading rules: recent evidence from Southeast Asian stock markets,” International Review of Economics and Finance, vol. 25, pp. 356-371, 2013.

ISSN: 2304-1013 (Online); 2304-1269 (CDROM); 2414-6722 (Print)