Stock Price Trend Prediction using Artificial Neural Network Techniques

Case Study: Thailand Stock Exchange

Weerachart Lertyingyod and Nunnapus Benjamas

Department of Computer Science, Faculty of Science, Khon Kaen University, Thailand 40002
weerachart.lertyingyod@kkumail.com, nunnapus@kku.ac.th

Abstract — **This paper presents a predictive model which to predict the trends of stock prices using Data Mining techniques. This research will allow the investor to make a more informed decision to buy and sell stocks, and in the most appropriate period. The predictive concept in this work implies learning historical price patterns, indicators, and behavior; and then predicting the future trends in one, five, and ten day periods.**

We compare the effectiveness of feature selection using Gain Ratio Attribute with the Ranker Search Method and Wrapper Selection with Greedy Step Wise Search Methods.

Interestingly, we can reduce the attributes from 14 to 6 (57.14%) using Wrapper Subset Evaluation with Greedy algorithm through forward selection. Accuracy improved over the models which were built from the original number of attributes.

The results of our experiment demonstrate that the predictive model for weekly (5 and 10 days) stock price direction is improved through the use of Artificial Neural Network (ANN) classification, in which the maximum accuracy of the model reached 93.89% at 10 days prediction, which were a vast improvement to the daily and 5 day predictions employing only six selected input attributes.

Keywords—Artificial Neural Network, Feature Selection, trend prediction, stock indicators, data mining , directional movement, data processing, and SET50.

I. Introduction

The successful stock investment is one of the most challenging to all investors, such as business owners, fund managers, and even profit minded individuals outside of the business sector. Previous research has included numerous forms of prediction of stock analysis, including stock price at the close of each business day [1], and their developing (future) trends. [2]

Typically, they separate the investor into two types: [3]

1. *Fundamental analysis*, which analyzes company fundamentals, such as fact sheets depicting growth, profit and loss, as well as ratios determining price per earning (P/E), price per book value (P/BV), and earnings per share (EPS) in order to predict future stock prices.
2. *Technical analysis*, in which stock prices are matched to signal indicators, in order to forecast future stock prices and influencing the decision to buy and sell.

However, troublesome, unsolved issues remain, such as:

- External factors which affect market fluctuation and cannot be predicted. [4]
- Stock prediction based upon company fundamentals, which remains incomprehensible for the new investor.

Through technical analysis, we can follow the trends of stock prices in any given period. Technical terms have been presented in graph form, in which data depicts stock prices as well as indicators; all of which reflect upon the needs of the investor.

A buy signal [5] indicates the appropriate time to in which to buy a stock, implying that the stock's price will predictably go up, or become *uptrend*. In contrast, a sell signal [5] suggests the proper time to sell a stock, indicating that the prices is predicted to go down or move into a *downtrend* position.

However, indicators may signal in error; for example, when a buy signal is given for a price direction which is only short-term. By issue of this type of error signal, our research focuses on solving the problem by creating a model capable of predicting the trend of stock price by weighing a variety of indicators, known as *combination indicators*. [6] The model was generated using an Artificial Neural Network (ANN) algorithm, and evaluated through accuracy rate and F-Measure of prediction, forecasting stock price direction in one, five, and ten day periods.

II. Related Works

Over the past few years, several research efforts have been made in the field of stock prediction. Having surveyed the numerous efforts which have led us to our current research, we

978-1-5090-4420-7/16/$31.00 ©2016 IEEE

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY BOMBAY. Downloaded on June 03,2021 at 05:40:07 UTC from IEEE Xplore. Restrictions apply.

have categorized stock prediction into two groups: stock trend prediction and stock price prediction.

1. Stock Trend Prediction

Stock trends are generally predicted as *uptrend* or *downtrend*. The basis of these predictions is either through models employing various algorithms, or through data research. Because algorithms are selected based upon the data input into the model, data processing become an important tool in each method. Both methods attempt to reach the same outcome, which is, the predicted direction of the stock price. Kuo-Ping Wu, Yung-Piao Wu, and Hahn-Ming Lee [7] employed K-Means clustering to determine the relationship between indicators, through Association Rules, referred to as the AprioriAll algorithm. This unique trading strategy yielded a net profit of 49.73% over four years. Proponents of the data research technique focused on the data preparation process, analyzed with data mining techniques, resulting in a highly accurate model with less processing time. Patel, Shah, Thakkar, and Kotecha [6] applied the Rain Forest technique in order to predict the stock's trend, which yielded 90.79% accuracy. Nearly by complex algorithms, such as the Artificial Neural Network, this demonstrate the importance of data preprocessing in terms of both performance and processing time.

Some researchers have suggested that traders who frequently buy and sell through such high performance strategies prove more successful than those who buy and hold for the long term. Maragoudakis and Serpanos [2] practiced the Random Forest technique with financial news from various on-line media. The price trend predictions, evaluated through F-Measure (73.44%), further yielded margins of 16-48% from investments in the Greek Stock Market over a two week period.

2. Stock Price Prediction

Stock prediction may also be based upon the stock's daily closing price, using historical price patterns. Shaverdi, Fallahi, and Bashiri [7] predicted the stock price of the Iranian Petrochemical Industry using a genetic algorithm with the Group Method of Data Handling (GMDH) neural network. They determined the model accuracy of the statistic values as the absolute fraction of variance (R^2), the Root-Mean Squared Error (RMSE), and the Mean Absolute Deviation (MAD). The resulting R^2 reached 99%, but was limited to each individual model; and could not be applied to other stocks, as each maintain a difference price pattern and require new training and test data. Kannan, Sekar, Sathik, and Arumugam [4] combined an indicator signal with the previous day's closing price in order to predict the price for the following day. The annual return of the algorithm was 32.17%, with trading wins of 58% in all buying signals.

Sample research on groups 1 and 2 were designed under the same concept, yet for different purposes. Group 1 focused on the class of trend, rather than the estimation value. The target of Group 2 was the estimation of the stock price, which fluctuated daily.

III. Indicator Backgroud

There are three types of forecast indicators examined in our research: data indicators, feature selection, and classification.

A. Data Indicators

General pricing trends follow the path of *uptrend*, *downtrend*, or *sideways*, as shown in Figure 1.

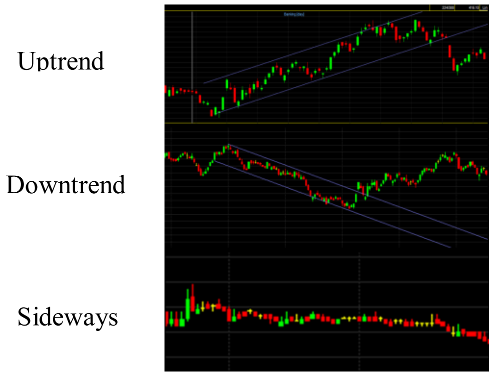

Figure 1. Price Uptrend, Downtrend and Sideways.

The objective of each indicator is to determine the right time to buy and sell a stock. Several data indicators are explained below.

Moving Average Convergence/Divergence (MACD)

The MACD [5] is a moving average oscillator that shows potential overbought/oversold phases of market fluctuation, and is a calculation of two moving averages of the underlying price/indicator. This indicator consists of the following attributes:

- MACD line, by default, calculated from the difference of the exponential moving average price (26 days and 12 days) as shown in Equation 1.

$$\text{MACD} = \text{EMA}(12) - \text{EMA}(26) \quad (1)$$

- Signal line, which shows the exponential moving average of the closing price over the last 9 days, or EMA 9.

MACD Buy Signal: MACD > Signal value.

MACD Sell Signal: MACD < Signal value.

Relative Strength Index (RSI)

RSI [5] is one of most popular indicators used in stock trend analysis. It shows the strength of the price trend within a single company. The value is continuous between 0-100, and is calculated through Equation 2.

$$\text{RSI} = 100-[100/1+\text{RS}] \quad (2)$$
$$\text{RS} = \text{Ua} / \text{Da}$$

- Ua is the average price change when the closing price has moved up in the last 14 days.
- Da is the average price change when the closing price has moved down in last 14 days.

By default, an RSI under 30 will alert a buy signal, just as an RSI above 70 will alert a sell signal.

Average Directional Movement Index (ADX)

ADX [8] is an indication of the severity of a stock in either an uptrend or downtrend, consisting of three indicators:

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY BOMBAY. Downloaded on June 03,2021 at 05:40:07 UTC from IEEE Xplore. Restrictions apply.

- ADX confirms the trend of stock prices. The higher the value, the more likely a trend will occur, however the direction is supported by the PDI (DI+) and MDI (DI-).
- DI + indicates a stock uptrend, if it increases and crosses over DI-.
- DI – indicates a stock downtrend, if it increases and crosses down DI+.

All measurements are calculated by the following equations:

$$PDI(n) = \sum_{i=1}^{n}[(PDM)_i] / \sum_{i=1}^{n}[(TR)_i] \qquad (3)$$

$$MDI(n) = \sum_{i=1}^{n}[(MDM)_i] / \sum_{i=1}^{n}[(TR)_i] \qquad (4)$$

$$DX = Abs(\frac{[PDI(14)-MDI(14)]}{[PDI(14)-MDI(14)]}) \qquad (5)$$

$$ADX(n) = \sum_{i=1}^{n}[(DX)_i]/n \qquad (6)$$

Parabolic Stop and Reverse (SAR)

The Parabolic SAR [8] was developed by Welles Wilder. This indicator is most often used to set trailing price stops. The SAR alerts when the price penetrates a Parabolic SAR level.

Figure 2. SAR Buy and Sell Signal

The SAR buy signal occurs at the first point when the SAR falls below the closing price. The SAR sell signal occurs at the first point when the SAR goes over the closing price.

Slow Stochastic Oscillator (SSTO) %K and %D

The Stochastic Oscillator [6], developed by George C. Lane in 1950, considers the trend of a price within a specific period. Values are represented as %K and %D, and range from 0-100, and are calculated as:

$$\%K = \frac{\text{Recent Close (t)} - \text{Lowest Low (n)}}{\text{Highest High (n)} - \text{Lowest Low (n)}}; \qquad (7)$$

n = 9 is default.

%D = MA n days of %K ; n=3 is default (8)

STO buy signal: when %K > %D value
STO bell signal: when %K < %D value

Exponential Moving Average (EMA)

An exponential moving average [8] is similar to a simple moving average, and is calculated by applying a small percentage of the current value to the previous value. An EMA applies more weight to recent values, and is calculated as:

$$EMA_{(n,t)} = [EMA_{(n,\,t-1)} + SF\,(P_{(t)} - EMA_{(n,\,t-1)})] \qquad (9)$$

By $EMA_{(t-1)}$: EMA at time t-1
SF: Smooth Factor = 2 / (n+1)
$P_{(t)}$: Current Close price
n: period (n days)

In order to analyze a trend using EMA, you must compare EMA short-term values with EMA long-term values. Unlike some indicators, the buy/sell signal is not self- generated, and the best period (n) cannot be specified.

EMA Buy Signal: when EMA-short > EMA-long term
EMA Sell Signal: when EMA-short < EMA-long term

This research applied the Feature Selection [9] and classification techniques (B and C) to filter the necessary attributes before creating a predictive model using the classification algorithm.

B. Feature Selection

GainRatioAttributeEval [9]

This algorithm determines the relationship between the attribute and class that determines a rank in terms of its weight. This ensures that the necessary attributes are selected for input before building a prediction model. We applied the GainRatioAttributeEval as the attribute evaluator with Ranker search methodology to sort the necessary attributes.

Wrapper Subset Evaluation

The Wrapper Selection [9] is used evaluate each attribute and learning expectation through Cross-Validation (k-Fold). Our research uses the Greedy Algorithm [10] search method with Wrapper Subset Evaluation [11] for attribute evaluation via a Forward Selection algorithm.

The process of forward selection is as follows:

1. Select only a single attribute from all candidates.
2. Add the selected attribute to the model.
3. Evaluate the classification model using an added set of attributes.

If accuracy increases, continue to add from the selected attributes to the set of features, and repeat steps 1 to 3, until the accuracy of model no longer improves.

C. Classification

Artificial Neural Network (ANN) [12]

The ANN is one of the most popular classification algorithms, which provides a high-efficiency predictive model within complex data. This proposed ANN model consists of three main layers, show in Figure 3.

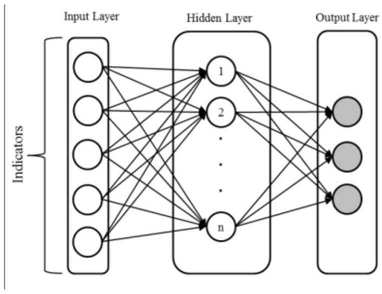

Figure 3. Artificial Neural Network

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY BOMBAY. Downloaded on June 03,2021 at 05:40:07 UTC from IEEE Xplore. Restrictions apply.

1. The Input Layer consists of nodes or all input attributes in the Training set, which delivers all input data into the Hidden Layer.
2. The Hidden Layer consists of the node responsible for the learning processing, input from the Input Layer, referred to as *perceptron*. Comparable to human brain cells, the hidden layer calculates the summary of input (x) and weight (W), in the following equation:

$$S = \sum_{i=1}^{n}[(w_i x_i)_i] + w_0 \tag{10}$$

3. The Output Layer consists of a class node, calculated by an Activate Function, usually a Sigmoid Function; unless the indicator is a classification model, in which case the Activate Function will be a Threshold Function:

$$f(x) = \begin{cases} 1 & \text{if } x \geq T \\ 0 & \text{if } x < T \end{cases} \tag{11}$$

IV. Research Process

The research process consists of four primary steps: data preparation, filter attribute selection, classification model creation, and model evaluation model, shown in Figure 4.

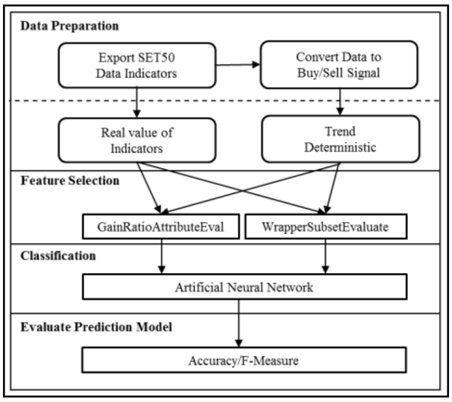

Figure 4. The Research Process

A. Data Collection

Data indicators were selected daily from SET50, and the training data set was collected from 2012-2015, and from January to March 2016 for the test set. All data were divided into real data indicators and data to be converted into signals.

B. Data Processing

The real data indicators consisted of 14 attributes, which were used without transformation to any another format, illustrated in Figure 5.

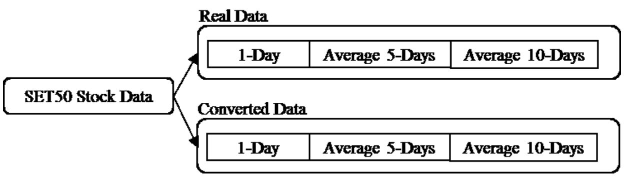

Figure 5. Group of data set.

RSI	EMA (5)	EMA (10)	EMA (20)	EMA (30)	SAR	MACD	Signal
Hist	ADX	DI+	DI-	SSTO %K	SSTO %D	Class	

Figure 6. Attribute of Real Data Indicator

On the other hand, the real dataset was transformed into Buy/Sell signals through the conditions in Table 1. As a result of this transformation, the attributes were reduced from 14 to 9.

RSI	ADX	EMA (5>10)	EMA (5>20)	EMA (5>30)	SAR Signal
MACD Signal	DI Signal	SSTO Signal	Class		

Figure 7. Attribute of converted data indicator

Table I .Transform Data Condition

Indicator	Formula Condition	Result
MACD	MACD Buy = [MACD(26,12,9) – Signal >0]	1
	MACD Sell = [MACD(26,12,9) – Signal <0]	0
DI	DI Buy = (DI+) – (DI-) >0	1
	DI Sell = (DI+) – (DI-) <0	0
SSTO	SSTO Buy = SSTO(%K) – SSTO(%D) > 0	1
	SSTO Sell = SSTO(%K) – SSTO(%D) < 0	0
SAR	SAR Buy = Close price - SAR(Value) > 0	1
	SAR Sell = Close price - SAR(Value) < 0	0
EMA	EMA Buy = EMA(Short Term) - EMA(Long Term) > 0	1
	EMA Sell = EMA(Short Term) - EMA(Long Term) < 0	0

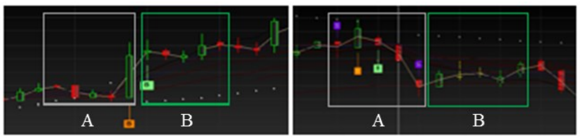

Figure 8. Comparing Average Prices over five days

The label or class of each instance was determined by the closing price over a five day period, Figure 8. “A” represents the previous five day period, and “B” is the class of instance in the training set, labeled by the following conditions:

- *Uptrend* when the average closing price B > A
- *Downtrend* when the average closing price B < A
- *Steady* when the average closing price B = A

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY BOMBAY. Downloaded on June 03,2021 at 05:40:07 UTC from IEEE Xplore. Restrictions apply.

V. Experimental Results

Applying the filter algorithm using GainRatioAttributeEval with Ranker search method, compared with the Wrapper algorithm using WrapperSubsetEvaluate, operated with a Greedy Forward search method, results in the selected attribute and their respective accuracy and performance.

The accuracy compared with the selected attributes of 18 models using K-Fold Validation (K=10) is shown in Table II and Figure 9.

Table II. Selected Attributes

No.	Dataset	Feature Selection	Ori.	Selected	Reduce
1	Real 1 Day	-	14	14	0.00%
2	Real 5 Day	-	14	14	0.00%
3	**Real 10 Day**	**-**	**14**	**14**	**0.00%**
4	Real 1 Day	GainRatioAttributeEval	14	14	**0.00%**
5	Real 5 Day	GainRatioAttributeEval	14	9	**-35.71%**
6	Real 10 Day	GainRatioAttributeEval	14	9	**-35.71%**
7	Real 1 Day	WrapperSubsetEval	14	6	**-57.14%**
8	Real 5 Day	WrapperSubsetEval	14	6	**-57.14%**
9	Real 10 Day	WrapperSubsetEval	14	6	**-57.14%**
10	Con. 1 Day	-	9	9	0.00%
11	Con. 5 Day	-	9	9	0.00%
12	**Con.10Day**	**-**	**9**	**9**	**0.00%**
13	Con. 1 Day	GainRatioAttributeEval	9	9	**0.00%**
14	Con. 5 Day	GainRatioAttributeEval	9	9	**0.00%**
15	Con.10Day	GainRatioAttributeEval	9	9	**0.00%**
16	Con. 1 Day	WrapperSubsetEval	9	3	**-66.67%**
17	Con. 5 Day	WrapperSubsetEval	9	4	**-55.56%**
18	Con.10Day	WrapperSubsetEval	9	6	**-33.33%**

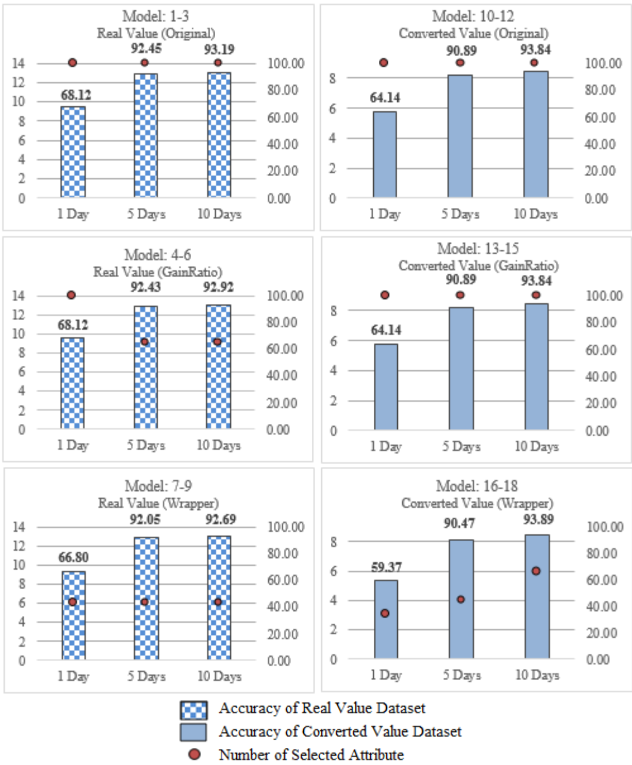

Figure 9. Result Comparison.

Details of the F-Measure for each model are given in Table III.

Table III. Compare Accuracy and F-measure

No.	Dataset	Accuracy	F-Measure		
			Uptrend	Downtrend	Steady
1	Real 1 Day	68.12	72.42	73.64	0
2	Real 5 Day	92.45	92.96	92.78	0
3	**Real 10 Day**	**93.19**	**93.29**	**93.5**	**0**
4	Real 1 Day	68.12	72.42	73.64	0
5	Real 5 Day	92.43	93.2	93.13	0
6	Real 10 Day	**92.92**	**93.34**	**93.33**	**0**
7	Real 1 Day	66.80	71.16	73.34	0.04
8	Real 5 Day	92.05	92.79	92.78	0
9	Real 10 Day	**92.69**	**92.69**	**93.69**	**0**
10	Con. 1 Day	64.14	68.07	68.87	0
11	Con. 5 Day	90.89	91.44	91.50	0
12	**Con.10Day**	**93.84**	**93.90**	**94.20**	**0**
13	Con. 1 Day	64.14	68.07	68.87	0
14	Con. 5 Day	90.89	91.44	91.50	0
15	Con.10Day	**93.84**	**93.90**	**94.20**	**0**
16	Con. 1 Day	59.37	63.04	64.50	0
17	Con. 5 Day	90.47	91.13	91.25	0
18	Con.10Day	**93.89**	**93.91**	**94.29**	**0**

Experiment results demonstrated the ability of the prediction model to predict the direction of stock prices over a period of 10 days. The maximum accuracy of the prediction model from 5 to 10 days differed slightly; however, the lesser number of attributes significantly reduced the processing time. By filtering the attribute via the Wrapper subset evaluation with the use of a Greedy Stepwise search method reduced the attributes to 57.14% in the real data indicator, and 33.33% in the converted dataset. The maximum accuracy of the prediction model reached 93.89% under the converted dataset, utilizing only six attributes within an average period of 10 days.

VI. Discussion

We can reduce the volatility of the data basis on a daily with the average value over the time such as by weekly. Result in the predictive model provide higher accuracy.

Filter attribute using Wrapper subset evaluation with Greedy Stepwise can reduce the number of attribute more than 50% because in case of the attribute which added to evaluation each round become the accuracy reduced, it will be cut off and won't be used to build the predictive model. On the other hands, gain ratio algorithm evaluates the worth an attribute by measuring the gain ratio with respect to the class, if the gain ratio of attribute is not equal to zero, it will be selected to the prediction model.

Notice that Gain Ratio Attribute Evaluation not work on the set of converted data shown in 13, 14 and 15 in Table II. Because the value of data is binary (0 and 1), there is a high probability that each attribute can be associated with the class, becomes the gain ration of each attribute is not equal to zero. Even the priority of attribute will be different but all of them will be used to build the predictive model.

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY BOMBAY. Downloaded on June 03,2021 at 05:40:07 UTC from IEEE Xplore. Restrictions apply.

The ANN algorithm is a highly effective technique which can be applied to stock data with high accuracy prediction in data input format, in which there are real value indicators and conversion values.

Focus on the F-measure found that the prediction model proved appropriate in determining stock uptrends and downtrends. An imbalance in the data set class [13] (48.71%: 50.84%: 0.44%, uptrend: downtrend: steady) caused model biasing on the majority data set, in both uptrends and downtrends, sending the F-measure of the steady class near zero.

VII. CONCLUSION

From our research we may conclude that stock price direction can be predicted more accurately on a weekly basis rather than daily. This is due to daily external factors which cause stock prices to fluctuate. By using an average price calculation, the stock trend stabilized and enhanced the prediction model accuracy. The future work, we will focus on the class definition using another condition such as comparing the buy and sell signal of each indicator in the weekly period then build the predictive model using another algorithm.

REFERENCES

[1] M. M. Aldin, H. D. Dehnavi and S. Entezari, "Evaluating the Employment of Technical Indicators in Predicting Stock Price Index Variations Using Artificial Neural Networks (Case Study: Tehran Stock Exchange)", *IJBM*, vol. 7, no. 15, 2012.

[2] M. Maragoudakis and D. Serpanos, "Towards Stock Market Data Mining Using Enriched Random Forests from Textual Resources and Technical Indicators", *IFIP Advances in Information and Communication Technology*, vol. 339, pp. 278-286, 2010.

[3] University of Cambridge, "Technical Analysis", *University of Cambridge*, Cambridge, England, 2011. [Online]. Available: http://www.mrao.cam.ac.uk/~mph/Technical_Analysis.pdf.

[4] K. Kannan, P. Sekar, M. Sathik and P. Arumugam, "Financial Stock Market Forecast using Data Mining Techniques", in *the International MultiConference of Engineers and Computer Scientists 2010*, Hong Kong, 2010, pp. 555-559.

[5] K. Wu, Y. Wu and H. Lee, "Stock Trend Prediction by Using K-Means and AprioriAll Algorithm for Sequential Chart Pattern Mining", *Journal of information science and engineering*, vol. 30, pp. 653-667, 2014.

[6] J. Patel, S. Shah, P. Thakkar and K. Kotecha, "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques", *Expert Systems with Applications*, vol. 42, no. 1, pp. 259-268, 2015.

[7] M. Shaverdi, S. Fallahi and V. Bashiri, "Prediction of Stock Price of Iranian Petrochemical Industry Using GMDH-Type Neural Network and Genetic algorithm", *Applied mathematical sciences*, vol. 6, no. 7, pp. 319-332, 2012.

[8] F. X. S. D. Nugroho, T. B. Adji and S. Fauziati, "Decision support system for stock trading using multiple indicators decision tree," *Information Technology, Computer and Electrical Engineering (ICITACEE), 2014 1st International Conference on*, Semarang, 2014, pp. 291-296.

[9] N. Rachburee and W. Punlumjeak, "A comparison of feature selection approach between greedy, IG-ratio, Chi-square, and mRMR in educational mining," *2015 7th International Conference on Information Technology and Electrical Engineering (ICITEE)*, Chiang Mai, 2015, pp. 420-424.

[10] A. K. Farahat, A. Ghodsi and M. S. Kamel, "An Efficient Greedy Method for Unsupervised Feature Selection," *2011 IEEE 11th International Conference on Data Mining*, Vancouver,BC, 2011, pp. 161-170.

[11] R. Kohavi and G. John, "Wrappers for feature subset selection", *Artificial Intelligence*, vol. 97, no. 1-2, pp. 273-324, 1997.

[12] A. Devadoss and T. Ligori, "Stock prediction using artificial neural networks", *International Journal of Data Mining Techniques and Applications*, vol. 2, pp. 283-291, 2013.

[13] N. V. Chawla , K. W. Bowyer , L. O. Hall and W. P. Kegelmeyer, "SMOTE: synthetic minority over-sampling technique", *Journal of artificial intelligence research*, vol. 16, pp. 321-357, 2002

Authorized licensed use limited to: INDIAN INSTITUTE OF TECHNOLOGY BOMBAY. Downloaded on June 03,2021 at 05:40:07 UTC from IEEE Xplore. Restrictions apply.