Songklanakarin J. Sci. Technol.
43 (5), 1428-1434, Sep. - Oct. 2021

http://www.sjst.psu.ac.th

Original Article

Ensemble classification method for daily return stock market

Meilany Nonsi Tentua[1*] and Dedi Rosadi[2]

[1] *Department of Informatics, Faculty of Science and Technology, Universitas PGRI Yogyakarta, Yogyakarta, 55182 Indonesia*

[2] *Department of Statistics, Gadjah Mada University, Yogyakarta, 55281 Indonesia*

Received: 27 June 2020; Revised: 26 October 2020; Accepted: 22 November 2020

Abstract

Stock price movements in Indonesia are measured using indices, one of which is the Composite Stock Price Index (CSPI). CSPI is a stock index that measures a combination of all shares from various sectors listed on the Indonesia Stock Exchange.

The ensemble method is to build predictive models by combining the strengths of the classical classification method. In this research, the purpose of ensemble based on Boosting for Regression appeared to enhance simple tree analysis and deals with some of the weaknesses found in uncomplicated techniques. The ensemble tree combines the prediction values of many simple trees into a single prediction value. Based on the experiments that have been carried out, the ensemble method proved to have a better accuracy rate, which amounted to 82%. It is assumed that the ensemble model can obtain the relationships between variables that are more precise than the previous model.

Keywords: ensemble method, stock market returns, classification

1. Introduction

Stock is one of the most popular financial market instruments. Stock can be interpreted as a sign of capital participation of a person or party (business entity) in a company or limited liability company ("PT Bursa Efek Indonesia," n.d.). Share price movements in Indonesia are measured using indices, one of which is the Composite Stock Price Index (CSPI). CSPI is a stock index that measures a combination of all shares from various sectors listed on the Indonesia Stock Exchange. Investors usually use parameters in the CSPI to read price developments and make a reference to the portfolio.

A stock market index is a measurement of the value of a part of the stock market. It is often used to describe market aggregate trends that will be predicted by a basic financial issue. Such stochastic value-share is very difficult to predict (Bruni, 2017). The difficulty in predicting the return of shares purchased by investors means that an investor is still guessing about the returns that investors will obtain in the future. There are four approaches used in predicting stock price movements such as technical analysis (Sedighi, Mohammadi, Fard, & Sedighi, 2019), fundamental analysis (Miralles-Quirós, Miralles-Quirós, & Oliveira, 2017), traditional time series forecasting (Park, Lee, & Lee, 2019), and statistical or machine learning methods (Kim & Han, 2016).

Statistical learning refers to a set of tools for modeling and understanding complex data sets. Statistical learning is a newly developed area in statistics and integrated with parallel developments in computer science and, in particular, machine learning (Gareth James, Daniela Witten, Trevor Hastie, 2013). Predictive modeling using data-sets is one of the tasks in statistical learning. An approach to predict the qualitative response is a process known as classification. Classification is the attribution of labels to notes according to criteria that are automatically learned from a series of labeled record training.

Some classical classification algorithms are widely used today (Trevor Hastie, Robert Tibshirani, 2017) such as

*Corresponding author
Email address: meilany@upy.ac.id

Linear Discriminant Analysis, Logistic Regression (Gündüz, Çataltepe, & Yaslan, 2017) (Shah, Ismail, & Bin Shahrin, 2019), K-Nearest Neighbors (Tanuwijaya & Hansun, 2019), Naïve Bayes (Marianto & Maruddani, 2020), Support Vector Machines (Selakov, Cvijetinović, Milović, Mellon, & Bekut, 2014), and Trees (Sitorus & Tarihoran, 2018). However, using a relatively small amount of data in stock market data, models built with Linear Discriminant Analysis, Logistic Regression, K-Nearest Neighbors, Naive Bayes, and Tress will easily over-fit. Their variance also tends to increase. Thus, it has an impact on the low accuracy achieved.

In making predictions for the daily stock return market, an appropriate method is needed to produce an accurate model. Classification using the Random Forests method in the study of Jyothirmayee *et al.* (2019), and the boosting process in the research of Nayak, Pai, & Pai (2016), have provided a slightly better level of accuracy compared to previous classical methods. These methods are also called ensemble methods. They can improve the accuracy of prediction models formed. However, in their research, they used analytical sentiment variables that would be difficult for new investors. The idea of ensemble learning is to build predictive models by combining the strengths of classical classification methods (Trevor Hastie, Robert Tibshirani, 2017).

In this study, we aim to predict daily stock market returns in the movement of the stock Jakarta Composite Index (JCI). We also use the most regulated daily share price, such as PT Bank Negara Indonesia (Persero) Tbk, PT Perusahaan Perkebunan London Sumatra Indonesia Tbk, PT Unilever Indonesia Tbk, PT Astra International Tbk, PT Bank Mandiri (Persero) Tbk, and PT Hanjaya Mandala Sampoerna Tbk. To predict the daily stock market, we use the Boosted Regression Tree (Boost RT) method. Besides, we also made comparisons with other classical methods such as LDA, Logistic regression, KNN, and Naïve Bayes.

The contributions of our study is the use of simple features, only by calculating the stock closing price's return value from the five previous days. The classification method that we use is the Boosted Regression Tree (BoostRT) method, which is an ensemble method that combines the Boosting method and the Decision Tree method. This method is rarely used for predicting stock prices in Indonesia. We use accuracy techniques in evaluating the classification model.

2. Materials and Methods

Several studies have been conducted in the application of statistical learning in the stock market domain. The use of the bootstrap method is modified with Random Forest to predict the direction of the movement of stock index prices (Kim & Han, 2016). Training sets are generated by a modified bootstrap considering the impact of response variables simultaneously and applied in a random forest. The data used are the Korea composite stock price index with the variable close price, open price, high price, low price, trading volume, and training amount.

A hierarchical Deep Neural Network (DNN) is a method that can be applied to predict stock market returns. A study applying this method had been conducted to predict the stock market returns within the next five minutes (Lachiheb & Gouider, 2018). The data used were the main index of the Tunisian stock market (TUNINDEX) with the variables of TUNINDEX current value, Number of up direction stocks, Number of down direction stocks, Number of unchanged stocks, and Total stock values in Tunisian dinars. Other research that uses DNN is a large comprehensive data analysis process to predict the daily return direction of the S&P 500 SPDR ETF (ticker symbol: SPY) based on 60 financial and economic features (Zhong & Enke, 2019).

The approach of a cross-sectional statistical estimation model for selecting stocks on the Chinese stock market allows investors to identify stocks that are likely to perform well and to build a suitable portfolio (Wu, Chen, Xu, He, & Tindall, 2019). The data used are sourced from the Shanghai Composite Index.

Predicting stock market prices using Fuzzy Metagraph (FM) based on stock market decision making, classification, and prediction are proposed for short-term investors from the Indian stock market (Anbalagan & Maheswari, 2015). Simple Moving Average (SMA), Exponential Moving Average (EMA), Moving Average Convergence Divergence (MACD), and Relative Strength Index (RSI) are some technical indicators that are used as input to train an integrated system with Fuzzy Meta graph. The data used in the Indian stock market data are from the Bombay Stock Exchange (BSE).

Artificial Neural Network (ANN) is a method in machine learning that can be applied for prediction and classification. Implementation of ANN back-propagation is used to predict the closing price of the S & P 500 stock exchange. Historical data consist of five variables, namely open, high, low, close, and volume.

The Adaptive Neuro-Fuzzy Inference System (ANFIS) method is used to determine the Jakarta Composite Index (JCI) value movement. This study provides two outputs, namely prediction results and classification results based on several relative error values. The classification method used is Fuzzy Kernel C-Means (Fanita & Rustam, 2018).

This research is done to classify and predict the daily return of the stock market in Indonesia. The method undertaken in this research can be seen in Figure 1.

Step 1: Collecting CSPI data from the Indonesia Composite Index through the Yahoo Finance web.

Step 2: Calculating the market stock daily return value from the data collected in step 1 using the formula:

$$\text{Return Value} = \frac{\text{Currentvalue - Originalvalue}}{\text{Originalvalue}} \times 100 \qquad (1)$$

Step 3: Pre-processing data by handling the missing valuesand excessive data from the data-set.

Step 4: Making sure that the training data-set is larger compared to the test set. The training data-set is taken from 2015 to 2018 data, while the test data-set is taken from 2019 data.

Step 5: Applying the ensemble Boosting Regression Tree (BoosRT).

Step 6: Analyzing the results of the ensemble model performance in terms of its accuracy, memory, accuracy, and specificity.

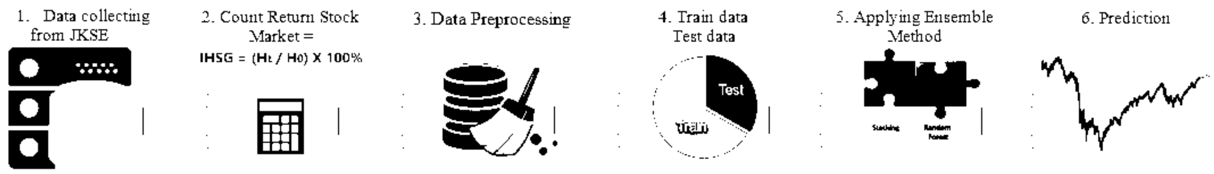

Figure 1. Steps under taken in the research

2.1 Boosted Regression Tree (BoosRT)

Boosted Regression Tree (BoosRT) models are a combination of two techniques: decision tree algorithms and boosting methods (Shin, 2015). In these algorithms, the BoosRT approximates the function (x) as preservative growth of the base learner:

$$f(x) = f_0(x) + \beta_1 f_1(x) + \beta_2 f_2(x) + \cdots + \beta_m f_m(x) \quad (2)$$

A single study cannot predict enough using training data, so we must use the lowest residuals to improve predictive performance. BoosRT uses an iterative algorithm, so every iteration m, and a new regression tree. The formula to predict a separate constant value is:

$$h(x; \{R_{lm}\}_l^L = \textstyle\sum_{l-1}^{L} \overline{y_{lm}}\,(x \epsilon R_{lm}) \quad (3)$$

where y_{lm} = mean $x_i \in R_{lm}$ $(\tilde{y}im)$ is the mean of pseudo-residuals the m^{th} iteration:

$$\tilde{y}_{im} = -\left[\frac{\partial\Psi(y_i, F(x_i))}{\partial f(x_i)}\right]_{f(x)=F_{m-1}(x)} \quad (4)$$

After that the approximation obtained will be updated with the following equation

$$f_m(x) = f_{m-1}(x) + v.Y_{lm} \,|\, (x \in R_{lm}), \quad (5)$$

Where

$$Y_{lm} = \arg min_Y \textstyle\sum_{x_i \epsilon R_{Im}} \Psi\,(y_i,\ F_{m-1}\,(x_i) + Y) \quad (6)$$

BoosRT algorithm for the generalized boosting of regression trees are:

a) Initialize (x), $F0(x)$ = arg minY $\sum Ni{-}1$ $\Psi(yi, \gamma)$.
b) For m=1 to M do
c) From the full training dataset, select a subset randomly,

$$\{\pi(i)\}_l^N = rand\_perm\{i\}_l^N \quad (7)$$

d) Fit the base learner

$$\tilde{y}_{\pi(i)} = -\left[\frac{\partial\Psi\left(y_{\pi(i)m}, F(x_i)\right)}{\partial F(x(i))}\right]_{F(x)=f_{m-1}(x)}, i = 1, \tilde{N} \quad (8)$$

e) Update model for the current iteration,

$$\{R_{lm}\}_l^L L - terminalmodetree\left(\{\widetilde{y_{\pi(i)}}, x_{\pi(i)}\}_l^{\tilde{N}}\right) \quad (9)$$

f) Choose a gradient descent step size as

$$Y_{lm} = \arg min_y \textstyle\sum_{x(i)\epsilon R_{lm}} \Psi(y_{\pi(i)}, F_{m-1}(x_{\pi(i)}) + Y) \quad (10)$$

g) Update the estimate of (x) as

$$F_m(x) = F_{m-1}(x) + v.Y_{lm} \,|\, (x \epsilon R_{lm}) \quad (11)$$

2.2. Performance evaluation

The performance of the classification model can be measured by counting the number of classes that were predicted correctly (true positive), the predicted numbers that were not included in the class and were true (true negative), and those that were wrongly predicted (false positive or false negative). The formulation of accuracy, precision, and recall is used in this study using the following formula:

$$Accuracy = \frac{tp + tn}{tp + fn + fp + tn} \quad (12)$$

$$Precision = \frac{tp + tn}{tp + fp} \quad (13)$$

$$Recall = \frac{tp}{tp + fn} \quad (14)$$

where, tp = true positive, tn = true negative, fp = false positive, and fn = false negative.

3. Results and Discussion

3.1. Results

The daily return stock market classification using the ensemble method is implemented using R. Some of the libraries contained in R are tools for statistical analysis and can also be used to create models with classification techniques (James, Witten, & Hastie, 2013), (Rosadi, 2017). The data-sets used in this study are JCI, PT Bank Negara Indonesia (Persero) Tbk, PT Perusahaan Perkebunan London Sumatra Indonesia Tbk, PT Unilever Indonesia Tbk, PT Astra International Tbk, PT Bank Mandiri (Persero) Tbk, and PT Hanjaya Mandala Sampoerna Tbk. The daily data were collected from the Jakarta Composite Index via the Yahoo Finance website. IHSG data were collected from December

25, 2014 to December 25, 2019. A data frame with 1193 observations comprises the following nine variables, namely:

Year: The year that the data was recorded
X1: return value at t-1
X2: return value at t-2
X3: return value at t-3
X4: return value at t-4
X5: return value at t-5
Volume: Volume of shares traded (number of daily shares traded in billions)
Today: return value at today (t)
Direction: the label Down and Up, indicating that the market had a positive or negative return on a given day.

The observation data taken from Yahoo Finance consists of 7 (seven) variables: Date, Open, High, Low, Close, Adj Close, and Volume. In pre-processing data, we will process the values in the observation data into a return value. The return value is calculated by using the close price value in the observed data. The results of return value based on observation data is shown in Table 1.

$$Today = \frac{6.236{,}69 - 6.219{,}44}{6.219{,}44} * 100 = 0{,}277$$

$$X1 = \frac{6.219{,}44 - 6.334{,}84}{6.334{,}84} * 100 = -1{,}82$$

$$X2 = \frac{6.334{,}84 - 6.342{,}17}{6.342{,}17} * 100 = -0{,}12$$

$$X3 = \frac{6.342{,}17 - 6.381{,}95}{6.381{,}95} * 100 = -0{,}63$$

$$X4 = \frac{6.381{,}95 - 6.336{,}67}{6.336{,}67} * 100 = 0{,}715$$

$$X5 = \frac{6.336{,}67 - 6.326{,}21}{6.336{,}67} * 100 = 0{,}165$$

The return value of a day becomes the label Direction, which indicating the market had a positive (Up) or negative (Down) return on a given day. The distribution of JKSE return value for 2015-2019 can be seen in Figure 2.

We compared the performance of the BoosRT method with the Logistic Regression, Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), k-Nearest Neighbors(kNN), and Naive Bayes methods.

The performance of the proposed method is validated by comparing it in terms of its accuracy, recall, precision, and specificity. The analysis is done by taking different proportions of tests and training. In conducting the training, data from 2015 to 2018 were used, as many as 951 data. Data testing used data from 2019, as many as 242 data.

Table 1 shows the results of the accuracy, recall, precision, and specificity of the classification method, where the BoosRT methods as ensemble methods. While the other method is the classical method.

Figure 2 shows the level of accuracy, recall, precision, and specificity of each method used in the JCI daily data between 2015 and 2019.

Table 1. Data observation

Date	Open	High	Low	Close
Sep 05, 2019	6,294.28	6,307.35	6,281.95	6,306.80
Sep 06, 2019	6,329.41	6,336.91	6,305.02	6,308.95
Sep 09, 2019	6,328.28	6,333.90	6,306.74	6,326.21
Sep 10, 2019	6,331.73	6,342.01	6,311.34	6,336.67
Sep 11, 2019	6,334.59	6,381.95	6,328.69	6,381.95
Sep 12, 2019	6,399.00	6,414.48	6,337.52	6,342.17
Sep 13, 2019	6,369.42	6,375.80	6,318.92	6,334.84
Sep 16, 2019	6,262.29	6,266.14	6,193.51	6,219.44
Sep 17, 2019	6,215.24	6,240.35	6,205.30	6,236.69

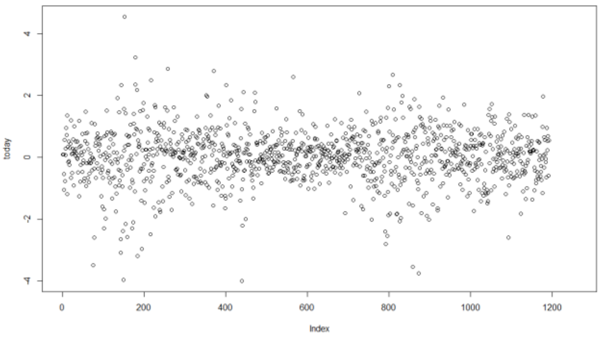

Figure 2. Return value today of JKSE 2015-2019

This ensemble method is also tested on other data sets, such as:

a) PT Kimia Farma (Persero) Tbk, produces and sells medicines and their derivative products.
b) PT Bank Negara Indonesia (Persero) Tbk, provides various banking products and services in Indonesia.
c) PT Perusahaan Perkebunan London Sumatra Indonesia Tbk, a plantation business in Indonesia.
d) PT Unilever Indonesia Tbk, engages in manufacturing, marketing, and distribution of consumer goods.
e) PT Astra International Tbk, together with its subsidiaries, engages in the automotive and financial services in Indonesia.
f) PT Indofarma (Persero) Tbk, manufactures and markets generic, over the counter, and branded generic drugs in Indonesia.
g) PT Hanjaya Mandala Sampoerna Tbk, manufactures and trades cigarettes in Indonesia.
h) PT Telekomunikasi Indonesia Tbk, provides telecommunications, information, media, and edutainment services worldwide.

We used the stock market data from these companies because their shares greatly affect JCI. This data set is taken from a model that has been formed. Table 3 shows the results of the accuracy, recall, precision, and specificity of the BoosRT classification method.

3.2 Discussion

In the first experiment we used IHSG data; it appeared that the logistic regression method had an accuracy of 54.32% or had 100-54, 3 2 = 47.8% training error rate. By removing variables that were not very influential in the

Table 2. Evaluation of classification methods

Methods	Accuracy	Recall	Precision
Logistic Regression	0.54	0.14	0.61
Linear Discriminant Analysis (LDA)	0.54	0.16	0.52
Quadratic Discriminant Analysis (QDA)	0.55	0.52	0.51
kNN	0.65	0.63	0.62
Naïve Bayes	0.77	0.76	0.77
BoosRT	0.82	0.83	0.71

Table 3. Ensemble classification with the other data set

Data set	Accuracy	Recall	Precision
PT Kimia Farma (Persero) Tbk	0,76	0,77	0,73
PT Bank Negara Indonesia (Persero) Tbk	0,81	0,97	0,73
PT Perusahaan Perkebunan London Sumatra Indonesia Tbk	0,79	0,81	0,70
PT Unilever Indonesia Tbk	0,82	0.83	0,81
PT Astra International Tbk	0,82	0,59	0,96
PT Indofarma (Persero) Tbk	0,8	0,71	0,82
PT Hanjaya Mandala Sampoerna Tbk	0,82	0,84	0,71
PT Telkom	0,78	0,74	0,76

formation of logistic regression models, an effective model was obtained. The accuracy level of 54.32% was the highest level of accuracy obtained by using very influential variables in the direction prediction.

The use of the Linear Discriminant Analysis (LDA) method did not significantly differ in accuracy compared with the Logistic Regression method. The LDA method had an accuracy rate of 54.48%. Likewise, the QDA method had an accuracy rate only 1% better than the Logistic Regression and LDA methods.

For kNN method, the models created were used in the data testing. The level of accuracy achieved was 65 % with k = 2. This model was better than LDA and QDA. In using this method, we tried k values from 1 to 15. When k = 2 the experiment achieved the highest accuracy values.

Training data using the Naïve Bayes method turned out to have a better level of accuracy when applied to test data, compared to the previous method. The accuracy of this method was 77 %, or 12 % better for training error.

Moreover, by using the ensemble method called BoosRT, it had an accuracy rate of 82 %. This is significant for stock market data, which is known to be very difficult to be predicted accurately. It shows that the model assumed by BoosRT can capture the relationship between variables more accurately than the previously used methods.

In the second experiment we used the BoosRT method on the return value of eight companies, namely PT Bank Negara Indonesia (Persero) Tbk, PT Perusahaan Perkebunan London Sumatra Indonesia Tbk, PT Unilever Indonesia Tbk, PT Astra International Tbk, PT Bank Mandiri (Persero) Tbk, and T Hanjaya Mandala Sampoerna Tbk. The results of the classification using ensemble models that have been obtained on 8 (eight) data sets showed an accuracy level of 76% to 82%. The average level of accuracy of the data set tested using the BoosRT model was 80.4%. The average recall from the model was 79.2%, which means that the success rate of the model in finding information had a better performance. The average precision model of 77.6% means that the level of prediction matches the data. It also means that the model has a good performance. From the measurement of accuracy, recall, and precision, the BoosRT model used had a better average than the logistic Regression, LDA, QDA, kNN, or Naïve Bayes methods. This shows that the ensemble model using the BossRT model is stable to be used in different data sets. The accuracy of the BoosRT model in the test data set can be seen in Figure 4.

In this study, we used the return value in the past five days as a feature. This feature considers that the stock price may be affected by stock movements from a few preceding days. This feature makes it easy for a new investor to predict the value of shares. Our research's use of features is different from research conducted by Jyothirmayee *et al.* (2019) using the random forest method; they used the data feature taken from Yahoo Finance. It is a weakness of the model because the prediction only considers the stock's movement on the previous day to lower accuracy.

Besides, a study by Nayak *et al.* (2016), shows that the use of features and stock value should also consider the sentiment analysis of the company. It will not be easy for a new investor in deciding to buy shares if he has to know the company's sentiment analysis.

The limitation of this study is that it uses only five return value variables and does not pay attention to the socio-political and security conditions that occur in Indonesia which also affect the stock market performance.

5. Conclusions

According to the experiments, the ensemble method has a high accuracy (82%). It is significant for stock market data, which is known to be difficult to model accurately. It shows that the model assumed by BoosRT can capture the relationship between variables more accurately than the methods used before. The classification using ensemble models had been obtained on the other data sets. The model has shown an accuracy level of 76% to 82%. It shows that the ensemble model using the BossRT model is stable to be used and generalized in different data sets.

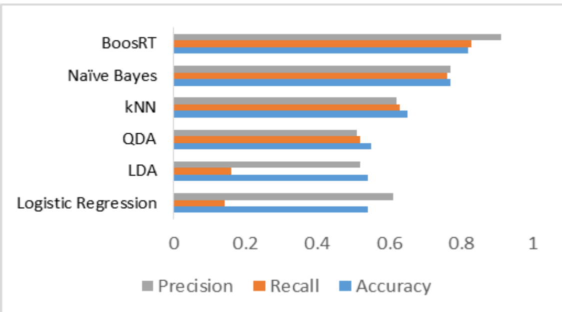

Figure 3. Analysis of classification methods with JKSE data set

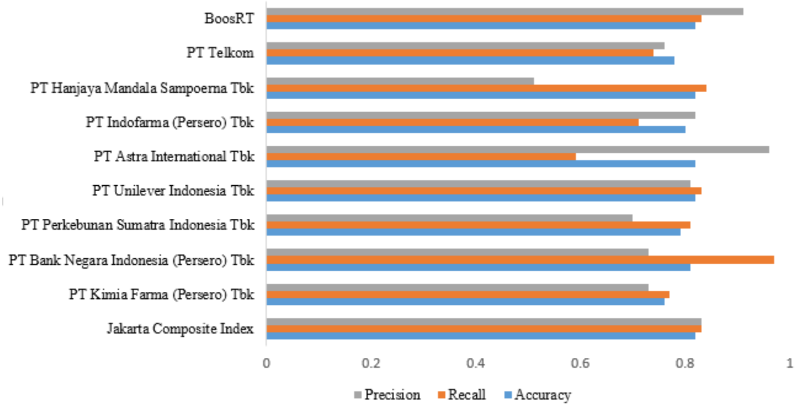

Figure 4. Model accuracy in other data sets

The policy implications in this study can use from two sides, investors and companies. For investors, predictions of the stock price can determine their investment strategy. It will relate to the rate of return they expect in investing. For companies, this prediction will be considered for the company to determine policies related to the company's stock price.

References

Anbalagan, T., & Maheswari, S. U. (2015). Classification and prediction of stock market index based on Fuzzy Metagraph. *Procedia Computer Science, 47*(C), 214–221. doi:10.1016/j.procs.2015.03.200

Bruni, R. (2017). Stock market index data and indicators for day trading as a binary classification problem. *Data in Brief, 10*, 569–575. doi:10.1016/j.dib.2016.12.044

Fanita, F., & Rustam, Z. (2018). Predicting the Jakarta composite index price using ANFIS and classifying prediction result based on relative error by fuzzy Kernel C-Means. *AIP Conference Proceedings, 2023*. doi:10.1063/1.5064203

Gareth James, Daniela Witten, Trevor Hastie, R. T. (2013). An introduction to statistical learning with applications in R. In G. Casella S. Fienberg I. Olkin (Ed.), *Springer Texts in Statistic*. doi:10.2174/0929867003374372

Gündüz, H., Çataltepe, Z., & Yaslan, Y. (2017). Stock daily return prediction using expanded features and feature selection. *Turkish Journal of Electrical Engineering and Computer Sciences, 25*(6), 4829–4840. doi:10.3906/elk-1704-256

Jyothirmayee, S., Dilip Kumar, V., Someswara Rao, C., & Shiva Shankar, R. (2019). Predicting stock exchange using supervised learning algorithms. *International Journal of Innovative Technology and Exploring Engineering, 9*(1), 4081–4090. doi:10.35940/ijitee.A4144.119119

Kim, H., & Han, S. T. (2016). The enhanced classification for the stock index prediction. *Procedia Computer Science, 91*(Itqm), 284–286. doi:10.1016/j.procs.2016.07.077

Lachiheb, O., & Gouider, M. S. (2018). A hierarchical Deep neural network design for stock returns prediction. *Procedia Computer Science, 126*, 264–272. doi:10.1016/j.procS.2018.07.260

Marianto, F. Y., & Maruddani, D. A. I. (2020). *Perbandingan Metode Naive Bayes dan Bayesian Regularization Neural Network (BRNN) untuk Klasifikasi Sinyal Palsu pada Indikator Stochastic Oscillator, 9*, 16–25.

Miralles-Quirós, M. del M., Miralles-Quirós, J. L., & Oliveira, C. (2017). The role of liquidity in asset pricing: the special case of the Portuguese Stock Market. *Journal of Economics, Finance and Administrative Science, 22*(43), 191–206. doi:10.1108/JEFAS-12-2016-0001

Nayak, A., Pai, M. M. M., & Pai, R. M. (2016). Prediction models for indian stock market. *Procedia Computer Science, 89*, 441–449. doi:10.1016/j.procs.2016.06.096

Park, M., Lee, M. L., & Lee, J. (2019). Predicting stock market indices using classification tools. *Asian Economic and Financial Review, 9*(2), 243–256. doi:10.18488/journal.aefr.2019.92.243.256

Rosadi, D. (2017). *Analisis statistika dengan R*. Retrieved from http://opac.perpusnas.go.id/DetailOpac.aspx?id=985555#

Sedighi, M., Mohammadi, M., Fard, S. F., & Sedighi, M. (2019). The Nexus between stock returns of oil companies and oil price fluctuations after heavy oil upgrading: Toward theoretical progress. *Economies, 7*(3), 1–17. doi:10.3390/economies7030071

Selakov, A., Cvijetinović, D., Milović, L., Mellon, S., & Bekut, D. (2014). Hybrid PSO-SVM method for short-term load forecasting during periods with significant temperature variations in city of Burbank. *Applied Soft Computing Journal, 16*, 80–88. doi:10.1016/j.asoc.2013.12.001

Shah, S. Q., Ismail, I., & Bin Shahrin, A. R. (2019). Stock market manipulation: A comparative analysis of East Asian emerging and developed financial markets. *Management Science Letters, 9*(1), 183–192. doi:10.5267/j.msl.2018.10.006

Shin, Y. (2015). *Application of boosting regression trees to preliminary cost, 2015*.

Sitorus, H., & Tarihoran, Y. (2018). Analisis Harga Saham PT Astra Internasional Tbk Menggunakan Data Dari Bursa Efek Indonesia dalam Jangka Waktu Pendek Menggunakan Metode Naïve Bayes dan Decision Tree-J48. *TeIKa*, *8*(1), 21–33. doi:10.36342/teika.v8i1.2239

Tanuwijaya, J., & Hansun, S. (2019). LQ45 stock index prediction using k-nearest neighbors regression. *International Journal of Recent Technology and Engineering*, *8*(3), 2388–2391. doi:10.35940/ijrte.C4663.098319

Trevor Hastie, Robert Tibshirani, J. F. (2017). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). *SpringerPlus*. doi:10.1111/j.1751-5823.2009.00095_18.x

Wu, W., Chen, J., Xu, L., He, Q., & Tindall, M. L. (2019). A statistical learning approach for stock selection in the Chinese stock market. *Financial Innovation*, *5*(1), 1–18. doi:10.1186/s40854-019-0137-1

Zhong, X., & Enke, D. (2019). Predicting the daily return direction of the stock market using hybrid machine learning algorithms. *Financial Innovation*, *5*(1). doi:10.1186/s40854-019-0138-0