International Journal of Economics, Commerce and Management
United Kingdom Vol. VI, Issue 12, December 2018
http://ijecm.co.uk/ ISSN 2348 0386

MODELING POLITICAL OPINION ON THE JAKARTA COMPOSITE INDEX USING MODEL AVERAGING IN INDONESIA

Agung Eddy Suryo Saputro
Central Bureau of Statistics, Jakarta, Indonesia
agung.eddyss@gmail.com

Khairil Anwar Notodiputro
Bogor Agricultural University, Faculty of Mathematics and Natural Sciences,
Department of Statistic, Bogor, Indonesia
khairilnotodiputro@gmail.com

Indahwati
Bogor Agricultural University, Faculty of Mathematics and Natural Sciences,
Department of Statistic, Bogor, Indonesia
indah.stk@gmail.com

Abstract

The movement of the Jakarta Composite Index (JCI) can be used to describe the economic conditions in Indonesia. The JCI is directly or indirectly influenced by the Election of Regional Heads that are taking place in 171 regions of Indonesia. The purpose of this study to determine the best method can be used to predict the influence of political opinion on the JCI accurately. Data collected is in the form of high dimensional data. Least Absolute Shrinkage and Selection Operators (LASSO) and Model Averaging are statistical method can be used to predict high dimensional data. The results showed that Random Model Averaging with Ridge and weight AIC (RRA) is the best method to predict the influence of political opinion on the JCI.

Keywords: Jakarta Composite Index, LASSO, Model Averaging, Political Opinion, Weight AIC

INTRODUCTION

The economic condition of a country can be described through the stock market. When the stock price index increases, this shows that economic conditions in the country are good, and vice versa. The movement of a country's price index, both directly and indirectly, is influenced by political events taking place in the country. Several studies have shown that the movement of the stock price index is influenced by political events that are happening in a country, for example, research conducted by Chan and Wei (1996) and Zach (2003). Chan and Wei (1996) showed that political news in China had a significant effect on the 5% level of the Hang Seng Index stock. Whereas Zach (2003) shows that Tel Aviv Stock Exchange Index stock movements are directly affected by political events that are happening in Israel.

The development of technology and information, especially the internet, makes information about political events that have just taken place more quickly and easily spread throughout the world. Social media has an important role in disseminating this information. One of the most popular and growing social media in the dissemination of information since 2006 is Twitter. Twitter is a site that allows users to present information in the form of writing with a maximum number of characters of 140 characters or often called microblogging (Java *et al.*, 2007). Twitter is one source of big data providers, this is because the number of tweets or data provided by Twitter is very large, with an average of 600 tweets every second.

This study examine political events and stock price indices in Indonesia for several reasons. The first reason, on June 27, 2018, Indonesia hold an Election of Regional Heads covering 17 provinces, 39 cities, and 115 districts. The second reason, the stock index in Indonesia which is represented by the Jakarta Composite Index (JCI) scored the highest score on history on February 19, 2018, which amounted to 6689.29 points. Third reason, discussions on political topics often become the trending topic of Twitter Indonesia during January 2018 until May 2018. Based on these three reasons, this study focus on the Election of Governors in 17 provinces whose data comes from political opinions on Twitter and JCI.

Regression analysis is a popular statistical method for analyzing the effect of an explanatory variable on the response variable and predicting the response variable accurately. However, when the number of explanatory variables is greater than the number of observations, regression analysis cannot be used to predict the response variable, this condition is called high dimensional data (Rahardiantoro, 2016). Salaki (2018) shows that there are 3 models that can be used to solve high dimensional data, namely *Selection Model*, *Model Averaging*, and *Penalized Regression*. First, *Selection Model* works by forming the best model based on several criteria, such as *Akaike Information Criterion* (AIC), *Cp Mallows*, and *Cross Validation* (CV). Second, *Model Averaging* works by combining candidate models, estimation methods, and weights to deal with high dimensional data. Third, *Penalized Regression* works by shrinking several parameters to be simpler. Some examples of *Penalized Regression* models, such as *Least Absolute Shrinkage and Selection Operators* (LASSO), *Elastic Net*, and *Smoothly Clipped Absolute Deviation* (SCAD). This study will use 3 methods, namely *Random Model Averaging with Ridge and weight AIC* (RRA), *Marginal Correlation Model Averaging with Ridge and weight AIC* (MRA), and LASSO.

The objectives of this research are comparing the predictive results of the three methods based on the *Root Mean Square Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), and *Rsquared* (R^2).

LITERATURE REVIEW

The JCI is the company stocks that are combined and listed on the Indonesia Stock Exchange (IDX). The movement of JCI shows the condition of the capital market and measuring whether the stock price index has increased or decreased. The JCI calculation is carried out every day after the close of trading, August 10, 1982 is used as the base year of the JCI calculation. The JCI has three functions, namely the marker of market direction, profit level gauges, and portfolio performance benchmarks.

There are several formulas used to calculation stock index. IDX uses formula weighted to calculate JCI. Weighting the calculation of JCI for each stocks depends on the number of stocks registered. Weighting differently on each stocks, resulting in a JCI value that is strongly influenced by the movement of stocks that have a large weight. The JCI formula is as follows:

$$JCI_t = \frac{\sum_{i=1}^{n} H_{ti} W_{ti}}{N_{to}} \qquad (1)$$

Where:

JCI_t = Jakarta Composite Index day *t*; H_{ti} = day-to-day stock price based on current price; W_{ti} = weight of the first day's stock *t*; and N_{to} = total day-to-day stock value based on base year price.

Nisar and Yeung (2018) conducted a research on public opinion regarding regional head elections held in the United Kingdom against The Financial Time Stock Exchange (FTSE) 100 stocks index. Research Nisar and Yeung (2018) have not been able to prove that Twitter political opinion has an influence on the FTSE 100. However, the results of his research have proven that there is a significant relationship at the 5% level between the political sentiment for regional head elections in the United Kingdom against the FTSE 100. The method used in the Nisar and Yeung (2018) is the analysis of Multiple Regression.

Text Mining

Political opinion data from *crawling* Twitter is still textual data that is not structured. The textual data must be converted into structured data, so that the textual data that has been structured can be further analyzed using statistical methods. The method used in unstructured textual data processing is called text mining. There are several stages of text mining analysis carried out in this study, namely as follows:

a. *Information Retrieval* (IR), is defined as a process of rediscovering the information needed from textual data so that information can be stored, represented, compiled, and processed (Manning *et al.*, 2008).

b. *Pre-processing*, is one of the important stages in text mining analysis, because the better the results of *pre-processing* will provide better predictive results. Several stages in *pre-processing*, namely: tokenizing, parsing, cleaning, filtering, case folding, word normalization, stop words removal, and stemming.

c. *Word volume*, is the process of counting the number of words as a result of *pre-processing*.

d. *TFIDF*, is a matrix-making process based on the number of occurrences of terms in a document.

e. *Sentiment mining* is extraction and processing of textual data so that positive, negative, and neutral sentiment classes are obtained (Pang and Lee, 2008).

METHODOLOGY

Regional head election in Indonesia, conducted on June 27, 2018. Therefore, the period of data collection in this study, began a month before regional head elections until a month after regional head elections, namely May 27, 2018 –July 27, 2018.

There are two types data used in this study, namely administrative data and textual data. JCI is administrative data sourced from https://www.idx.co.id. While political data is textual data that comes from *crawling* Twitter. Table 1 shows there are one response variables (JCI) and four main explanatory variables (*VWord*, *Pos, Neg*, and *TFIDF*). The provinces that were used as the sample for data collection were the provinces that conducted the gubernatorial election, which were 17 provinces. Each province has 4 explanatory variables, so the number of explanatory variables to be used in this study is 4 x 17 = 68 explanatory variables.

Table 1 List of research variables

Variables	Description	Unit
JCI	Jakarta Composite Index	Point
$VWord_i$	Word volume	*Word*
Pos_i	Positive sentiment	%
Neg_i	Negative sentiment	%
$TFIDF_i$	Average TFIDF every day	*Word/Tweet*

Source: https://www.idx.co.id and processed results R 3.5.1

The index number *i* in the explanatory variable show the provincial code, is as shown in table 2.

Table 2 Provincial code

Code	Province	Code	Province	Code	Province
1	North Sumatra	7	East Java	13	South Sulawesi
2	Riau	8	Bali	14	Southeast Sulawesi
3	South Sumatra	9	West Nusa Tenggara (NTB)	15	Maluku
4	Lampung	10	East Nusa Tenggara (NTT)	16	North Maluku
5	West Java	11	West Kalimantan	17	Papua
6	Central Java	12	East Kalimantan		

The software used is R version 3.5.1 with packages *twitteR, ROAuth, tm, sentimentr, e1071, wordcould, stringr, Hmics, caret,* and *glmnet*.

Least Absolute Shrinkage and Selection Operator (LASSO)

High dimensional data has a singularity problem which results in explanatory variables in the regression model correlating with each other or often called multicollinearity. Tibshirani (1996) develops the Ordinary Least Square (OLS) method by adding constraints$\sum_{j=1}^{p}|\beta_j| \leq t$so that the multicollinearity that occurs in the model can be overcome, the method developed is known as the *Least Absolute Shrinkage and Selection Operator* (LASSO). The coefficient in LASSO can be estimated using the following equation:

$$\hat{\beta}^{LASSO} = argmin\left\{\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j\right)^2 + \lambda\sum_{j=1}^{p}|\beta_j|\right\} \qquad (2)$$

Closed equations cannot be used to estimate coefficient LASSO, but LASSO solutions can be solved by quadratic programming (Tibshirani, 1996) and modifications to the Least Angle Regression (LAR) algorithm for LASSO (Efron *et al.*, 2004). The LASSO coefficient can be reduced to zero and some coefficients are exactly zero because LASSO uses constraints$\sum_{j=1}^{p}|\beta_j| \leq t$. The equation used to predict the response variable using the LASSO method is as follows:

$\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}^{\boldsymbol{LASSO}}$ (3)

Where**:** $\hat{\boldsymbol{y}}$= vector response variable (JCI) size 62 x 1; $\boldsymbol{X}$= matrix explanatory variable (political opinion) size 62 x 68; and$\hat{\boldsymbol{\beta}}^{\boldsymbol{LASSO}}$= vector estimator LASSO size 68 x 1.

Model Averaging

The *Model Averaging*(MA) can be used to predict the variable response to high dimension data (Rahardiantoro, 2016). The development of the MA method by Perrone (1993) and Claeskens and Hjort (2008) was used to improve the accuracy of the regression model predictions and overcome uncertainty models. The MA works by combining part or all of the regression predictions $\hat{f}_i(\boldsymbol{X})$to predict $\hat{f}(\boldsymbol{X})$ with $\hat{y}_i = \hat{f}_i(\boldsymbol{X})$ which is called the candidate model (Rahardiantoro, 2016). The final model prediction is formed by all candidate models combined with weighted averages. The final model prediction is used as the final prediction on the MA.

The performance of the MA is determined by three main stages, namely the construction of candidate models, estimates, and weighting criteria for candidate models. The candidate model construction is done by dividing the explanatory variables into several classes, where each class functions as a design matrix for each candidate model. Ando and Li (2014) proposed a *Marginal Correlation Model Averaging* (MCMA) in constructing candidate models based on *Marginal Correlation* values between the response variables with each explanatory variable. For example,ρ_j is the *Marginal Correlation* between the response variables $\boldsymbol{y} = (y_1, \ldots, y_n)'$with the explanatory variable $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})'$ so that $\rho_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}\, y_i$ (Salaki, 2018). Sort the explanatory variables $x_1^*, x_2^*, \ldots, x_p^*$ based on the value ρ from highest to lowest. Whereas Perrone (1993) proposes *Random Model Averaging* (RMA) as a process of forming candidate models because it allows all explanatory variables to be chosen randomly in each candidate model.

This study adopt the estimation method used Salaki (2018), namely using *Ridge Regression*. *Ridge Regression* is a development of OLS that adds a constraint $\sum_{j=1}^{p}\beta_j^2 \leq t$ (Hoerl and Kennard, 1970). *Ridge Regression* coefficient can be estimated by the following equation:

$$\hat{\beta}^{RIDGE} = argmin\left\{\frac{1}{2}\sum_{i=1}^{N}\left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\,\beta_j\right)^2 + \lambda\sum_{j=1}^{p}\beta_j^2\right\} \quad (4)$$

Estimator $\hat{\beta}$ in *Ridge Regression* is depreciated towards zero as the value of λ increases. *Ridge Regression* cannot select explanatory variables because coefficients that are assumed to be simultaneous may not be zero (Hastie *et al.,* 2008).

The weighting used in this study uses the *Akaike Information Criterion* (AIC) weights. AIC weighting is based on the AIC value generated by each candidate model (AIC_k). Claeskens and Hjort (2008) formulated the AIC weighting (w_k) as follows:

$$w_k = \frac{\exp(-\frac{1}{2}\Delta_k)}{\sum_{k=1}^{M} \exp(-\frac{1}{2}\Delta_k)} \; ; \; \Delta_k = AIC_k - AIC_{min} \qquad (5)$$

Each candidate model has a different w_k weighting. If the w_k weighting has a higher weight indicates that the candidate model *k* has a good prediction.

The equation for estimating the response variable in the candidate model *k* can be formulated as follows:

$$\hat{\boldsymbol{y}}_{\boldsymbol{k}} = \boldsymbol{X}_{\boldsymbol{k}} \hat{\beta}_k^{RIDGE} \quad (6)$$

The equation to predict the final response variable in *Model Averaging* can be formulated as follows:

$$\hat{\boldsymbol{y}} = \sum_{k=1}^{K} \frac{1}{w_k} \boldsymbol{X}_{\boldsymbol{k}} \hat{\beta}_k^{RIDGE} \quad (7)$$

Where**:** $\hat{\boldsymbol{y}}$= vector final response variable (JCI) size 62 x 1; $\boldsymbol{X}_{\boldsymbol{k}}$= matrix explanatory variable (political opinion) of candidate model *k* size 62 x *m*; $\hat{\beta}_k^{RIDGE}$= vector predictive coefficient parameter *Ridge Regression* of candidate model *k* size *m* x 1; and w_k= weighted AIC candidate model *k.*

RESULTS AND DISCUSSION

The pattern of the JCI movement for the period of May 27, 2018 –July 27, 2018 fluctuated and there was a tendency to decline from 6022.04 points to 5989.14 points (Figure 1). In the period May 27, 2018 –June 6, 2018, the JCI movement tends to increase and reached its highest value of 6106.70 points on June 6, 2018. On June 6, 2018 –July 3, 2018, the JCI movement tended to decline and reached its lowest point of 5633.94 on July 3, 2018. The period of decline in the JCI was in conjunction with the campaign period and voting for regional head elections. This indicates there is a possibility of the influence of the regional head elections, both directly and indirectly, on the movement of the JCI. Therefore, it need to be studied more deeply so that the influence of the regional head elections on the JCI movement can be proven statistically. For the period July 3, 2018 - July 27, 2018, the JCI movement tend to increase, this indicates the economic conditions in Indonesia are starting to improve.

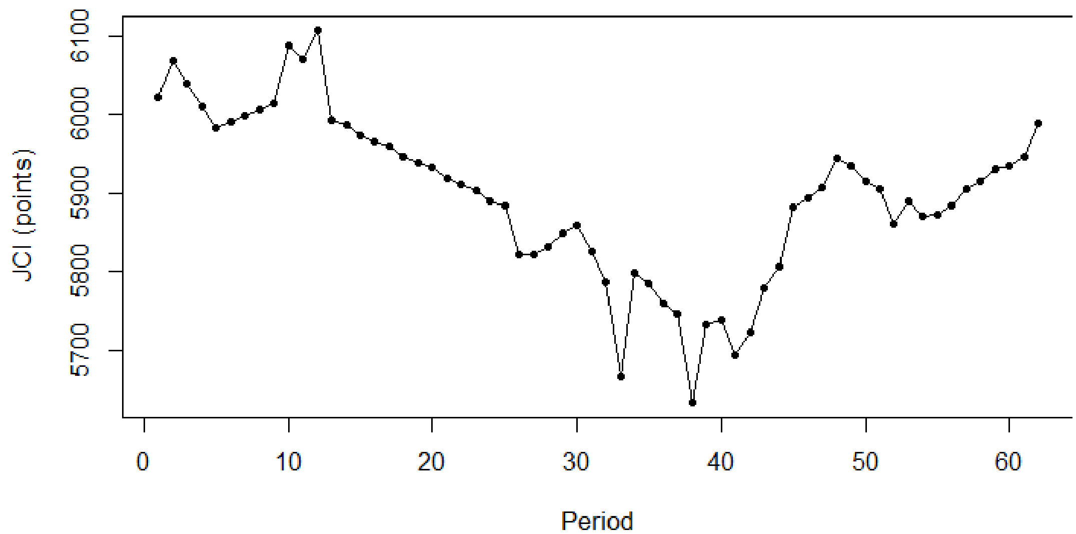


Figure 1 The JCI movement for the period May 27, 2018 - July 27, 2018

Source: processed results R 3.5.1

Application of the LASSO Method to JCI and Political Opinion

One important step in LASSO is the selection of the value λ which acts as a controller in the selection process of explanatory variables. Figure 2 shows the $\log(\lambda)$value that gives the best shrinkage results at the interval 2.73 – 3.48 or if transformed in the form of λ at the interval 15.40 – 32.41. The selection of the value λ in (Figure 2) uses the *Cross Validation* (CV) method by making the training data as a model design and testing data as a test of the goodness of the model.

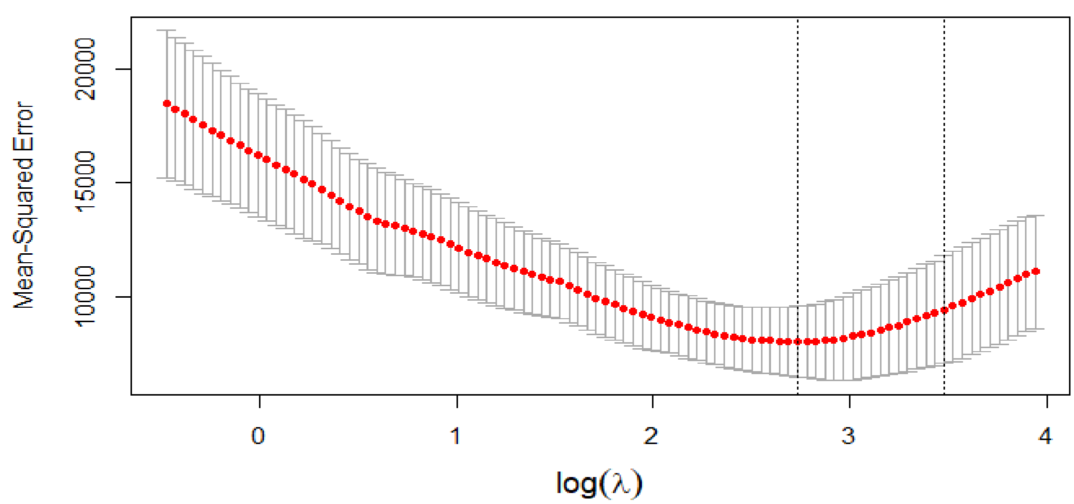


Figure 2 *Cross Validation* for log*(λ)* using the LASSO method

Source: processed results R 3.5.1

The modeling results using the LASSO method show that the JCI is influenced by 8 political factors, namely Lampung's positive sentiment, West Kalimantan's positive sentiment, NTB's negative sentiment, Maluku's negative sentiment, Riau TFIDF, South Sumatra TFIDF, Bali TFIDF, and NTB TFIDF. The equation produced by the LASSO method is as follows:

$$\widehat{JCI} = 6172.69 + 0.243Pos_4 + 0.068Pos_{11} + 1.621Neg_9 + 0.392Neg_{15} - 7.42TFIDF_2 - 57.234TFIDF_3 - 48.661TFIDF_8 - 55.629TFIDF_9$$

Figure 3 shows the pattern of the JCI movement which is predicted to have approached the actual JCI movement pattern, although there are still a number of points in the opposite direction. Prediction using LASSO produces a value of Mean Square Error (MSE) of 7079.95 and Root Mean Square Error (RMSE) of 84.14. While the value of Rsquare (R2) produced by LASSO is 29.74%, this shows the contribution of political opinion variables in explaining the JCI by 29.74%.

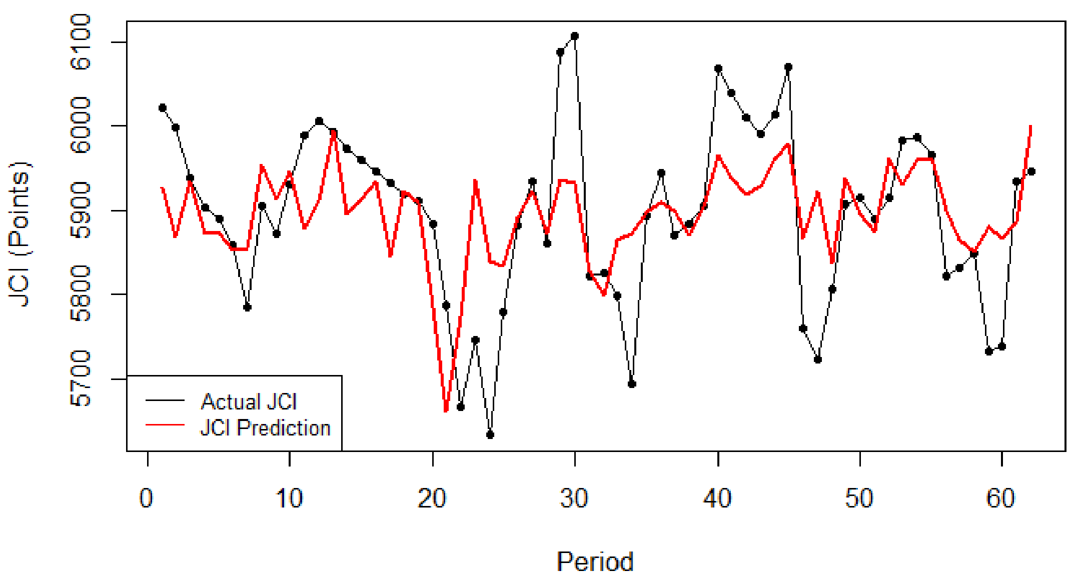


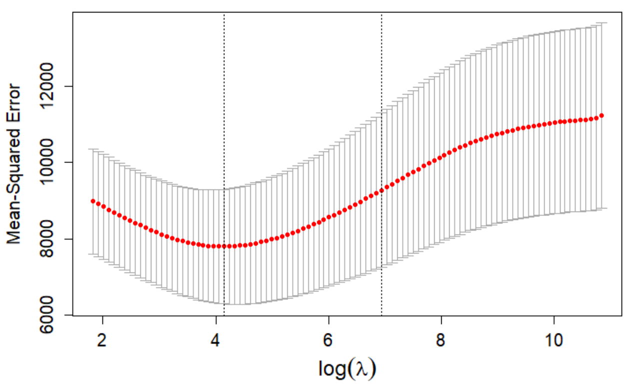
Figure 3 The movement of actual JCI and JCI prediction using LASSO method

Source: processed results R 3.5.1

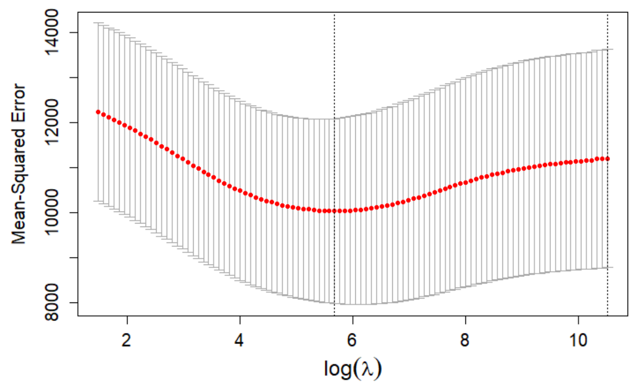
Application of the MRA Method to JCI and Political Opinion

The estimation method used in the MRA is *Ridge Regression*. As with LASSO, *Ridge Regression* also requires the stage of selecting the value λ so that the prediction results obtained are the best. Figure 4 shows each candidate model that forms a *Model Averaging* has a different interval of $\log(\lambda)$values. Candidate model 1 has $\log(\lambda)$ interval between 4.15 - 6.94, candidate model 2 has $\log(\lambda)$ interval between 5.68 – 10.52, and candidate model 3 has

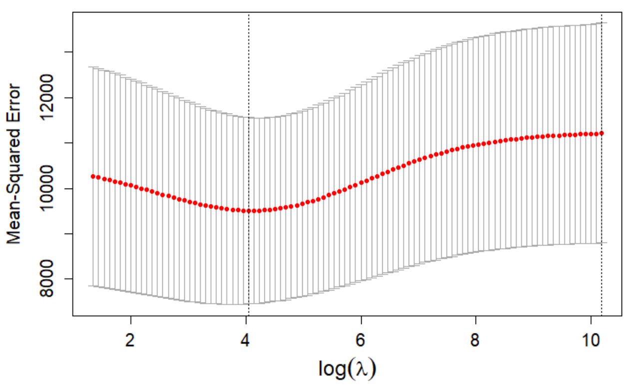
$\log(\lambda)$ interval between 4.04 - 10.18. While candidate model 4 and 5 only have a lower limit for $\log(\lambda)$, ie candidate model 4 has a lower limit of 9.66 and candidate model 5 has a lower limit of 8.89.



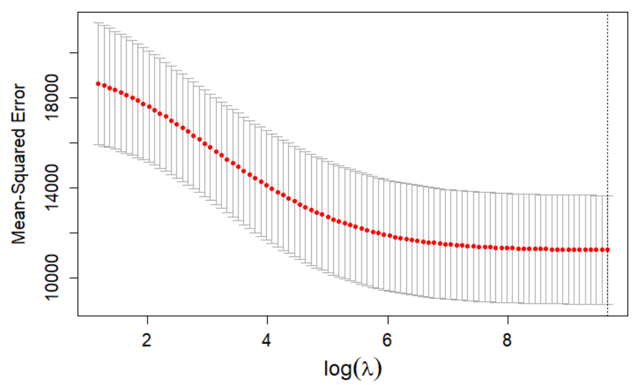
Candidate Model 1



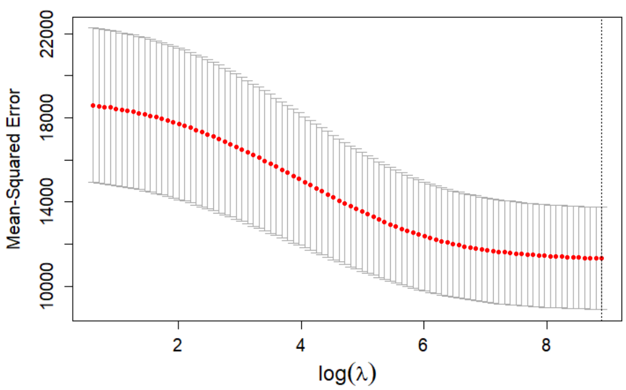
Candidate Model 2



Candidate Model 3



Candidate Model 4



Candidate Model 5

Figure 4 *Cross Validation* for log*(λ)* using the MRA method for each candidate model

Source: processed results R 3.5.1

Weighting is also one of the important elements in the *Model Averaging*. Each candidate model of MRA, has 1 AIC weight value. Weighting for candidate model 1 is 0.99723, candidate model 2 is 0.00004, candidate model 3 is 0.00273, candidate model 4 is 0.0000002, and candidate model 5 is 0.000000002. The resulting weighting value shows candidate model 1 has the biggest contribution in forming *Model Averaging*, while candidate model 5 has the smallest contribution in forming *Model Averaging*.

The *Model Averaging* equation used to predict the JCI value is a combination of all the equations produced by the candidate model. The equation of the *Model Averaging* produced by the *Marginal Correlation with Ridge and Weight AIC* (MRA) is as follows:

$$\widehat{JCI} = 6313.01 - 0.000VWord_1 - 0.066VWord_2 - 0.023VWord_3 + 0.000VWord_4 - 0.000VWord_5$$
$$-0.000VWord_6 - 0.000VWord_7 + 0.034VWord_8 - 0.000VWord_9 + 0.047VWord_{10}$$
$$-0.000VWord_{11} - 0.095VWord_{12} - 0.000VWord_{13} - 0.106VWord_{14} - 0.150VWord_{15}$$
$$-0.000VWord_{16} - 0.000VWord_{17} - 0.000Pos_1 + 0.002Pos_2 - 0.000Pos_3 + 0.000Pos_4$$
$$+0.000Pos_5 + 0.000Pos_6 + 0.000Pos_7 - 0.003Pos_8 - 0.000Pos_9 - 0.004Pos_{10} + 0.004Pos_{11}$$
$$+0.000Pos_{12} - 0.004Pos_{13} + 0.000Pos_{14} - 0.000Pos_{15} + 0.000Pos_{16} - 0.000Pos_{17} - 0.000Neg_1$$
$$-0.000Neg_2 + 0.000Neg_3 - 0.000Neg_4 - 0.003Neg_5 - 0.000Neg_6 - 0.000Neg_7 + 0.000Neg_8$$
$$+0.006Neg_9 + 0.000Neg_{10} - 0.000Neg_{11} - 0.000Neg_{12} - 0.000Neg_{13} + 0.000Neg_{14}$$
$$+0.000Neg_{15} - 0.000Neg_{16} - 0.000Neg_{17} - 0.000TFIDF_1 - 17.934TFIDF_2 - 48.598TFIDF_3$$
$$-0.000TFIDF_4 - 0.068TFIDF_5 - 0.002TFIDF_6 - 0.085TFIDF_7 - 51.408TFIDF_8$$
$$-36.576TFIDF_9 - 0.000TFIDF_{10} - 12.904TFIDF_{11} - 0.000TFIDF_{12} - 0.065TFIDF_{13}$$
$$-0.037TFIDF_{14} - 14.929TFIDF_{15} - 0.000TFIDF_{16} - 12.327TFIDF_{17}$$

The coefficient of MRA which is 0.000 show the coefficient value is very small and close to 0. Figure 5 shows the predicted pattern of JCI movement has approached the actual JCI movement pattern, although there are still a number of points in the opposite direction. Prediction using MRA produces the value of *Mean Square Error* (MSE) of 8600.11 and *Root Mean Square Error* (RMSE) of 92.74. While the value of *Rsquare* (R^2) produced by the MRA is 32.12%, this shows the contribution of political opinion variables in explaining the JCI by 32.12%.

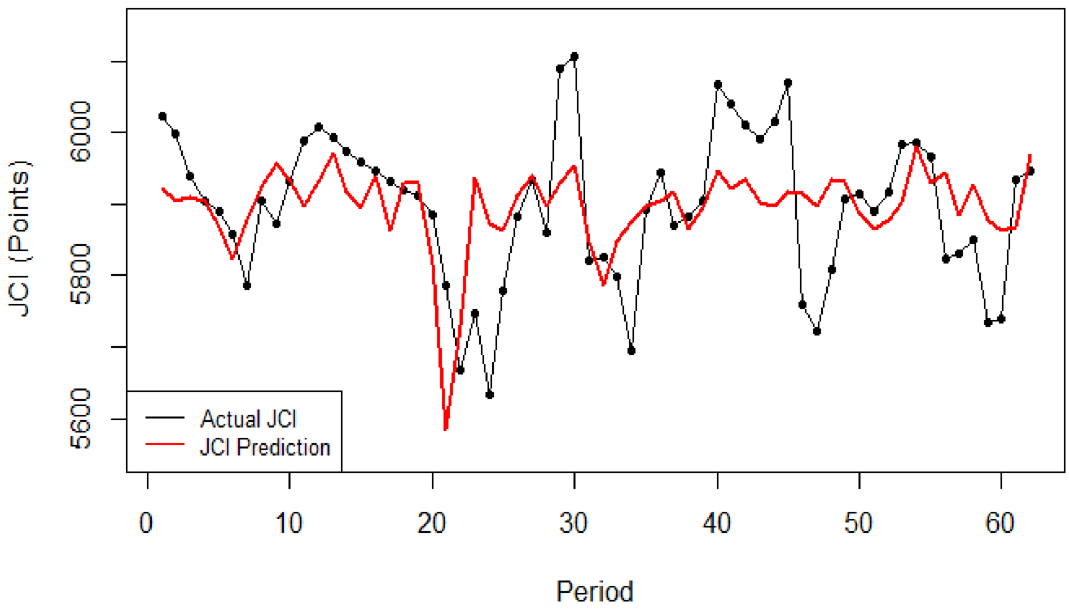


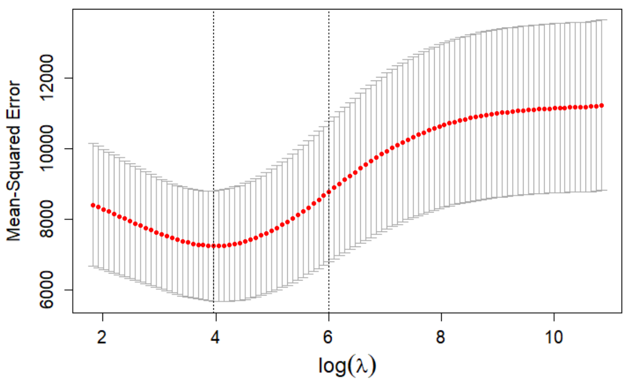
Figure 5 The movement of actual JCI and JCI prediction using MRA method

Source: processed results R 3.5.1

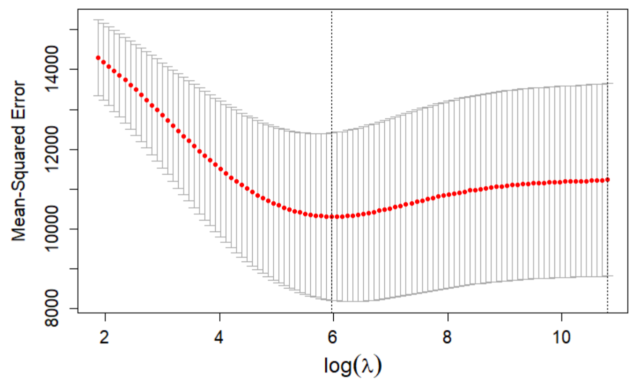
Application of the RRA Method to JCI and Political Opinion

The estimation method used in RRA is *Ridge Regression*. Figure 6 shows each candidate model that forms a *Model Averaging* has a different interval of $\log(\lambda)$values. Candidate model 1 has $\log(\lambda)$ interval between 3.97 – 6.01, candidate model 2 has $\log(\lambda)$ interval between 5.97 – 10.81, candidate model 3 has $\log(\lambda)$ interval between 6.37 – 10.65, candidate model 4 has $\log(\lambda)$ interval between 5.89 - 10.63, and candidate model 2 has $\log(\lambda)$ interval between 5.60 – 10.53.

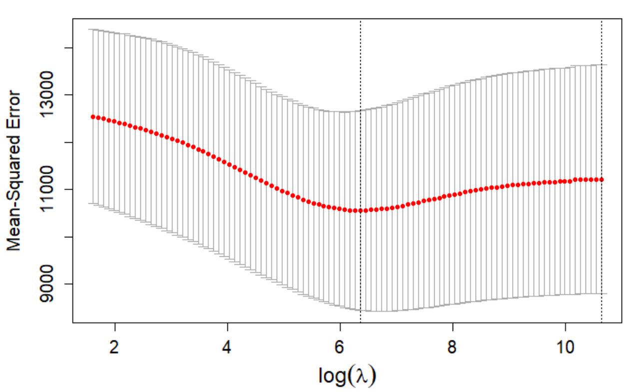
Each candidate model of RRA has 1 AIC weight value. Weighting for candidate model 1 is 0.99997, candidate model 2 is 0.000003, candidate model 3 is 0.00000005, candidate model 4 is 0.000000005, and candidate model 5 is 0.00000003. The resulting weighting value shows candidate model 1 has the biggest contribution in forming *Model Averaging*, while candidate model 4 has the smallest contribution in forming *Model Averaging*.



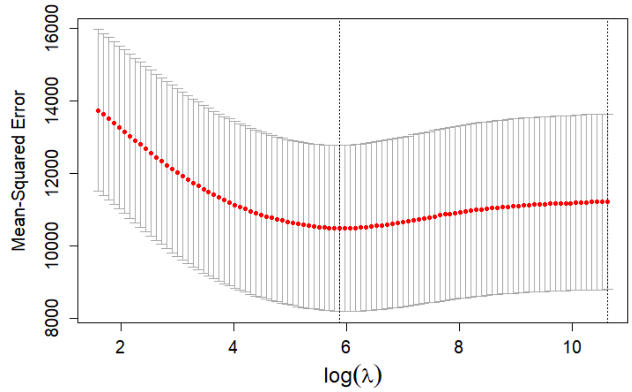
Candidate Model 1



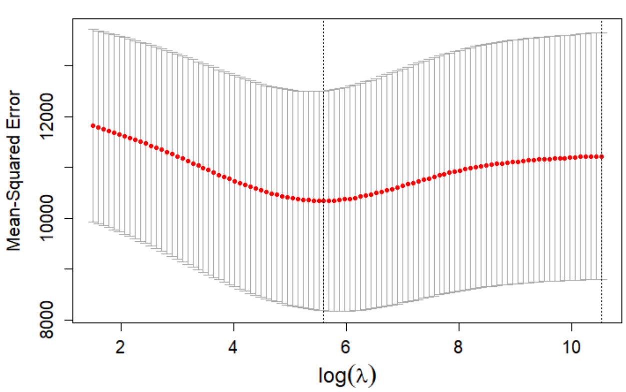
Candidate Model 2



Candidate Model 3



Candidate Model 4



Candidate Model 5

Figure 6 *Cross Validation* for log*(λ)* using the RRA method for each candidate model

Source: processed results R 3.5.1

The *Model Averaging* equation used to predict the JCI value is a combination of all the equations produced by the candidate model. The *Random Model Averaging with Ridge and weight AIC* (RRA) equation produced is as follows:

$$\widehat{JCI} = 6076.98 - 0.000VWord_1 - 0.000VWord_2 - 0.000VWord_3 - 0.000VWord_4 + 0.000VWord_5$$
$$-0.000VWord_6 - 0.000VWord_7 - 0.000VWord_8 - 0.000VWord_9 - 0.000VWord_{10}$$
$$-0.000VWord_{11} - 0.000VWord_{12} - 0.000VWord_{13} - 0.000VWord_{14} - 0.000VWord_{15}$$
$$-0.457VWord_{16} - 0.000VWord_{17} - 0.000Pos_1 + 0.000Pos_2 + 0.000Pos_3 + 0.997Pos_4$$
$$+0.000Pos_5 + 0.000Pos_6 + 0.076Pos_7 - 0.000Pos_8 - 0.000Pos_9 - 0.000Pos_{10} + 0.000Pos_{11}$$
$$-0.000Pos_{12} - 0.000Pos_{13} + 0.000Pos_{14} - 0.000Pos_{15} + 0.000Pos_{16} - 0.000Pos_{17} - 0.000Neg_1$$
$$-0.000Neg_2 + 0.000Neg_3 - 0.262Neg_4 - 0.820Neg_5 - 0.000Neg_6 - 0.000Neg_7 + 0.129Neg_8$$
$$+2.275Neg_9 + 0.000Neg_{10} - 0.234Neg_{11} - 0.000Neg_{12} - 0.000Neg_{13} + 0.000Neg_{14}$$
$$+1.813Neg_{15} - 0.000Neg_{16} - 0.000Neg_{17} - 5.513TFIDF_1 - 0.000TFIDF_2 - 63.249TFIDF_3$$
$$+16.203TFIDF_4 - 0.000TFIDF_5 - 0.000TFIDF_6 - 0.000TFIDF_7 - 48.468TFIDF_8$$
$$-0.000TFIDF_9 - 0.000TFIDF_{10} - 36.801TFIDF_{11} - 0.000TFIDF_{12} - 0.000TFIDF_{13}$$
$$-0.000TFIDF_{14} - 0.000TFIDF_{15} - 0.000TFIDF_{16} - 0.000TFIDF_{17}$$

The coefficient of RRA which is 0.000 show the coefficient value is very small and close to 0. Figure 7 shows the predicted pattern of JCI movement has approached the actual JCI movement pattern, although there are still a number of points in the opposite direction. Prediction using RRA results in a value of *Mean Square Error* (MSE) of 6730.25 and *Root Mean Square Error* (RMSE) of 82.04. While the value of *Rsquare* (R^2) produced by RRA is 44.59%, this shows the contribution of political opinion variables in explaining the JCI by 44.59%.

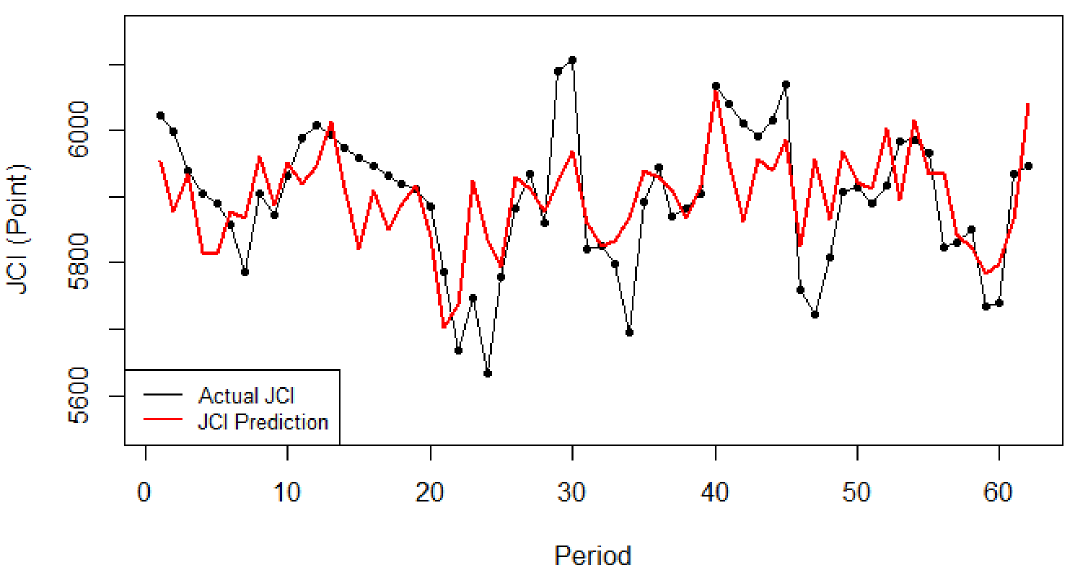


Figure 7 The movement of actual JCI and JCI prediction using RRA method

Source: processed results R 3.5.1

Comparison of Prediction Results Produced by LASSO, MRA, and RRA Methods

The JCI prediction movement pattern produced by the LASSO, MRA, and RRA methods shows a pattern similar to the actual JCI, although there is little movement that is in contrast to the actual JCI (Figure 8). The movement of the JCI prediction produced by the three methods is still not fluctuating sharply and tends to move in the middle compared to the actual JCI.

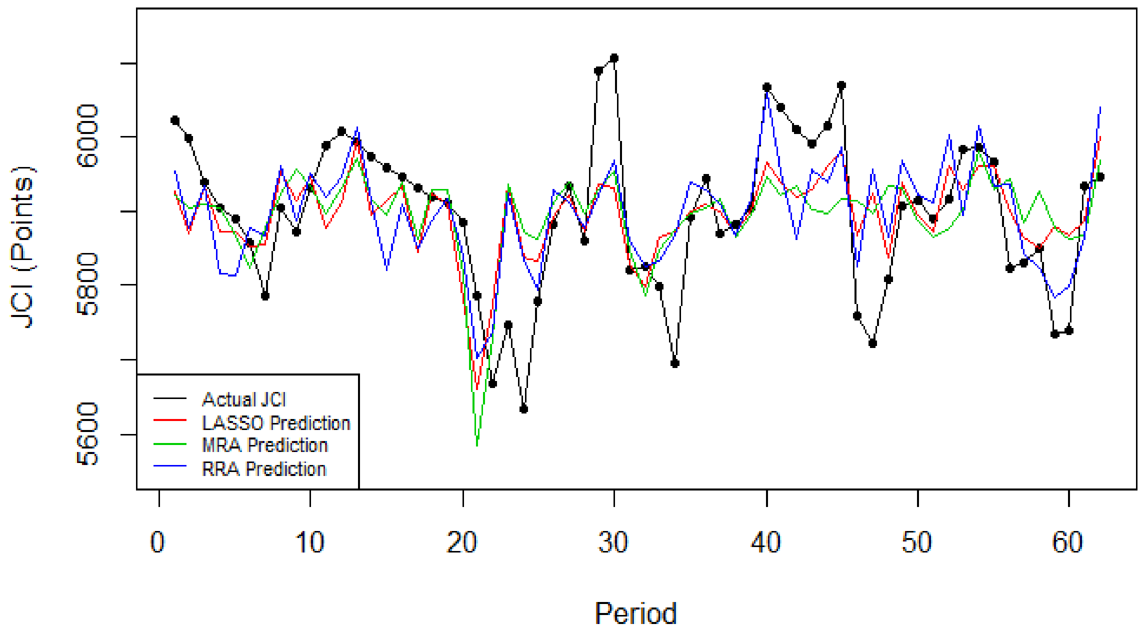


Figure 8 The movement of actual JCI and JCI prediction according to method

Source: processed results R 3.5.1

Table 3 shows average value produced by the prediction of the JCI of the three methods yields a value that approaches average the actual JCI (5898 points), which is 5895.75 points (LASSO), 5895.86 points (MRA), and 5897.82 points (RRA). The smallest MSE and RMSE values were obtained using the RRA method, which amounted to 6730.25 (MSE) and 82.04 (RMSE). The smallest value of the *Mean Absolute Percentage Error* (MAPE) is produced by LASSO and RRA, which is equal to 1.08. The RRA method produces the largest R^2 compared to the other 2 methods, which is 44.59%. Based on the value of RMSE, MAPE, and R^2, it can be concluded that the *Random Model Averaging with Ridge and Weight AIC* (RRA) method is the best method for predicting JCI values compared to the LASSO and MRA methods.

Table 3 Comparison of evaluation values according to method

Evaluation	Prediction LASSO	Prediction MRA	Prediction RRA
MSE	7079.95	8600.11	6730.25
RMSE	84.14	92.74	82.04
MAPE	1.08	1.24	1.08
R^2 (%)	29.74	32.11	44.59
Mean (Points)	5895.75	5895.86	5897.82

Source: processed results R 3.5.1

CONCLUSIONS

The movement of the Jakarta Composite Index (JCI) for the period May 27, 2018 – July 27, 2018 fluctuated and there was a downward trend. On June 6, 2018 – July 3, 2018, the JCI movement fluctuated and there was a tendency to decline sharply. The period of the decline in the JCI, along with the campaign period for the regional head elections and voting. This allows that the movement of the JCI, both directly and indirectly, is influenced by the regional head elections, which is happening in Indonesia.

The *Random Model Averaging with Ridge and weight AIC* (RRA) method is the best method in modeling the JCI with political opinion compared to the *Least Absolute Shrinkage and Selection Operator* (LASSO) and the *Marginal Correlation Model Averaging with Ridge and weight AIC* (MRA). This is indicated by the value of RMSE, and MAPE produced by RRA smaller than the LASSO and MRA methods, which are equal to 82.04 (RMSE) and 1.08 (MAPE). In addition, RRA as the best method is also shown by higher R^2 value than LASSO and MRA, which is 44.59%.

Subsequent research is recommended to extend the research period so that *crawling* data on political opinion on Twitter produces very large amounts of data. Collection of political data can be done through other social media, such as: Facebook. The development of a *model averaging* method is suggested in constructing the candidate model, estimation method, and weight selection.

REFERENCES

Ando, T., Li, K.C. (2014). "A Model Averaging Approach for High Dimensional Regression". Journal of the American Statistical Association. Vol. 109, No. 505: 254-265.

Chan, Y.C., Wei, K.C.J. (1996). "Political risk and the stock price volatility: the case of Hong Kong". Pacific Basin Finance Journal. Vol. 4, No. 2:259-275.

Claeskens, G., Hjort, N.L. (2008). "Model Selection and Model Averaging". New York: Cambridge University Press.

Efron, B., Hastie, T., Johnstone, I., Tibshirani, R. (2004). “Rejoinder to “Least Angle Regression” by Efron et al.”. The Annals of Statistics. Vol.32, No. 2: 494-499.

Hastie, T., Tibshirani, R., Friedman, J. (2008). “The Elements of Statistical Learning Second Edition”. New York: Springer.

Hoerl, A.E., Kennard, R.W. (1970). “Ridge Regression: Biased Estimation For Non-orthogonal Problems”. Technometrics. Vol.12 No. 1:55-67.

Java, A., Song, X., Finin, T., Tseng, B. (2007). “Why we Twitter: understanding microblogging usage and communities”. Proceedings of the Joint 9th WEBKDD and 1st SNA-KDD Workshop; 2007 August 12; California, United State. New York: ACM.

Manning, C.D., Raghavan, P., Schutze, H. (2008). “Introduction to Information Retrieval”. New York: Cambrige University Press.

Nisar, T.M., Yeung, M. (2018). “Twitter as a tool for forecasting stock market movements: a short-window event study”. The Journal of Finance and Data Science, 4: 101-119.

Pang, B., Lee, L. (2008). “Opinion mining and sentiment analysis”. Foundations and Trends in Information Retrieval. Vol. 2, No. 1-2: 1-135.

Perrone, M.P. (1993). “Improving Regression Estimation: Averaging Methods for Variance Reduction with Extensions to General Convex Measure Optimization [Dissertation]”. Providence: Brown University.

Rahardiantoro, S. (2016). “A New Approach for Constructing Model Averaging Candidates in High Dimensional Regression[Thesis]”. Bogor: Bogor Agricultural University.

Salaki, D.T. (2018). “A Study of Model averaging in High Dimensional Data with Various Correlation [Dissertation]”. Bogor: Bogor Agricultural University.

Tibshirani, R. (1996). “Regression shrinkage and selection via lasso”. Journal of the Royal Statistical Society 58: 267-288.

Zach, T. (2003). “Political events and the stock market: evidence from Israel”. International Journal of Business. Vol. 8, No. 3:243-266.