Inflation Rate Prediction in Indonesia using Optimized Support Vector Regression Model

Irvi Oktanisa\(^1\), Wayan Firdaus Mahmudy\(^2\), Ghozali Maski\(^3\),

\(^1,2\) Faculty of Computer Science, Brawijaya University
\(^3\) Faculty of Economics and Business, Brawijaya University

\{^1\text{irvioktanisa93@gmail.com},^2\text{wayanfm@ub.ac.id},^3\text{ghozalimaski@ub.ac.id}\}

Received 23 Desember 2019; accepted 10 February 2020

Abstract. Inflation is an indicator that illustrated the economic condition of a country. This monetary phenomenon is signed with the increase of price in the entire case. It can cause an effect on the political sector which impacts economic stability in a nation. The importance of inflation control is very important due to the high and unstable inflation that will harm economic and social in society. One of the solutions to control the inflation rate is determining an appropriate monetary policy based on future prediction of the inflation rate. This research using SVR as machine learning that is being optimized by GA as an evolutionary algorithm as a predicting method. SVR can solve nonlinear regression problems to linear regression using the Kernel function that easier to implement. But, in SVR there is no general rule to set the parameters of SVR. Therefore, this research proposed to use GA to optimize the parameters of SVR. GA can solve the optimization problems in various research on economics prediction problem. Based on the testing that has been conducted, GA-SVR generates the MSE value is 0.03767, lower than SVR basic method is 0.053158. It proves that the GA-SVR method can be utilized for predicting.

Keyword: Inflation, prediction, genetic algorithm, support vector regression

1 Introduction

Inflation is one of problems in various countries such as both developed and developing countries in the world, no exception in Indonesia. The term inflation is used by economics experts as the representation of all price rates in the economy was go up [1]. The inflation rate is a percentage change in the size of the price level from one period to the next [2]. Inflation is considered a serious problem by a country because inflation is an indicator that describes economic growth and market condition in a country. This is a monetary phenomenon characterized by a rise in overall prices.

The inflation occurs because of several factors, the main factor that common cause of almost all inflation cases is money growth [2]. The cycle of public consumption is characterized by the increasing circulation of money, the continued decline in the exchange rate of the Rupiah with foreign currencies which has an impact on the overall trade movement, as well as the increase in the import prices of goods from their home regions [3]. Indonesia will experience a prolonged monetary crisis when the problems caused by inflation are left unchecked and the inflation rate cannot be controlled. If that
happens, both political and social stability will be affected.

The inflation was begin in Indonesia in the 1960s that has reached 635% with average inflation of 196.08% [4]. It was a bad experience for the government and the people of Indonesia as a whole. Dornbusch and Fischer [5] classify the average inflation rate in Indonesia in the period 1970 to 1980 in the moderate range (15% to 30%). After the monetary crisis hit Indonesia, the economy began to experience a better development to control the inflation rate with various policy instruments carried out by the government.

Based on Bank Indonesia [6], various factors have depicted as reason for the importance of controlling the inflation rate. A high inflation will get an impact on the decline in the real income and the standard people living. The unstable inflation will get an impact on uncertainty in making decisions to consume, produce and invest. Furthermore, the high level of domestic inflation compared to the inflation rate of neighboring countries will make the real domestic interest rate uncompetitive, which can put pressure on the value of the rupiah.

One way that can be done to control inflation is to predict the inflation rate. It is able to provide the information for government policies in anticipating the case of the rate of increased inflation for the next period and also it may help to identify the factors for sudden inflation increase. It requires a method that can be used to solve the prediction of inflation problems. Predicting is made by time-series analysis technique based on historical data. This study will also use some external factors to determine the level of inflation. Some external factors used in this study include the Consumer Price Index (CPI), Money Supply, the BI rate, and Exchange Rate [7].

This research proposed Support Vector Regression (SVR) to solve the prediction problems. It is because the SVR can solve the high dimensional feature space of linear function using Radial Basis Function (RBF) as a kernel that needed to solve the prediction case. Furthermore, SVR can produce the high prediction accuracy with the excellent generalization capability and it has proven to become an effective implement in the real-value function estimation [8].

The previous study was conducted which applied SVR [9] to predict the inflation using money gap data based on time series techniques. Based on a previous study [9], SVR used to predict the inflation with an accuracy error of an average of 0.1 with a smaller average error rate than other machine learning methods in research that had done before. It is because SVR applies the principle of structural risk minimization and leads to generalizations that are better than conventional techniques. However, on the SVR model, there are no general rules for determining the parameters. Therefore, we proposed to use the optimization method to find the optimal parameters of SVR by using Genetic Algorithm (GA). To optimize the problems, SVR makes a modify of the function into dual convex quadratic programs [10].

A recent study has used the Genetic Algorithm (GA) as an optimization method to find the optimal parameters of SVR for USD/CNY prediction. The accuracy obtained using GA as an optimization method and used the RBF kernel is 0.00901 [11]. The previous study was conducted which applied Genetic Algorithm (GA) method [11] to find the optimal parameters for model selection to improve the performance of the SVR model. GA in various research is a popular algorithm in finance and economics, which can optimize the accuracy of prediction.

The optimization of the SVR's parameters model using the GA method with RBF kernel function [11] shows a lower average error value than another kernel function such as linear kernel function and polynomial kernel function. Therefore, this research proposed the GA-SVR method and RBF as a kernel function that used to solve the inflation problem with external factors as experimental data.
2 Previous Study

Several methods can be applied to solve the prediction of inflation problems, one of them is using Support Vector Regression (SVR). The previous study has done to solve the core of inflation series using SVR model. The study used 6 periods and 12 periods. The study using RMSE and MAE to evaluate the prediction result. Based on the result, the SVR has a lowest RMSE and MAE than backpropagation neural network and Maximum Likelihood Estimation (MLE) [12].

The previous study was applied SVR to predict rainfall in Bangladesh. The rainfall data is a time-series data and it changes as a season time and climate change. The accuracy of the proposed approach prediction is almost 99.92% [13].

Another case that applied SVR to solve the prediction problem was conducted which applied SVR [14] to predict the demand and supply of pulpwood in India. This study using time series data month per-month. This study using a Libsvm as a library for SVM by integrating with MATLAB. The result shows the SVR can predict demand and supply for pulpwood with produced a good prediction accuracy. This study claimed that the appropriately tuned of SVR based prediction model can outperform other complex models.

Another study using an optimization method to optimize the parameter of SVR was conducted which applied Genetic Algorithm (GA) [15] to optimize the SVR parameters to solve their problems. The results show that the RMSE of the GA-SVR method has a lowest RMSE than random forest (RF) and the combined of GA and partial least square regression (GA-PLSR).

The next study used the GA-SVR to investigate the accuracy of the method to solve the daily reference evapotranspiration ($ET_0$) evaluation. The study using two categorical accuracy of the prediction there are the epsilon loss and quadratic function. Based on the result, GA-SVR has a lower RMSE than SVR model are 0.0016 and 0.022 for epsilon loss function. The next result GA-SVR has a lower RMSE than SVR of 0.0018 and 0.026 for quadratic function [16].

The previous study used a hybrid Genetic Algorithm-Neural Network to predict the Peso exchange rate towards the US Dollar. The accuracy obtained using GA-NN is 0.1 [17]. Some study about using GA as an optimization method has been successfully to do by previous researchers to accomplish different problems [18].

3 Methodology

3.1 The Data Set

This research uses the data set based on historical data from Bank Indonesia, and Badan Pusat Statistik (BPS) as a statistical data resources. The used data is 126 records from January 2009 until June 2019. The data used is time series data with the dependent variable is inflation and the independent variables are the multivariable that consist of CPI, money supply, BI rate, and exchange rate. These factors are the external factors that affects the inflation rate in Indonesia. They were done selected using Vector Autoregression (VAR). Its purpose is to determine the significance of variables used in prediction [19]. The data set is shown in Table 1.

3.2 Cross-Validation

This research use Cross-Validation (CV) to divide the dataset into training data and testing data. The CV is one of the statistical analysis methods to validate the feature of the classifier [11]. Based on our exploration, we use 10 CV in our preprocessing on the data set.
3.3 Support Vector Regression

The main purpose of regression is to find a function which related to each variable. Support Vector Regression (SVR) is based on the theory of Support Vector Machine (SVM). Support Vector Machine or SVM is an update algorithm of neural network algorithms based on the statistical learning theory [20]. One of the basic characteristics of SVR is minimizing training errors to achieve the main goal of controlling the complexity of the hypothesis space [21]. SVR is not only able to solve problems with linear data but to solve overcome real data that has non-linear. In non-linear data, SVR converts input vectors to higher dimensions using kernel functions.

Based on the training data \( \{(x_1, y_1), \ldots, (x_n, y_n)\} \), the target of learning from SVR is to find a function that represents the relation between \( x \) and \( y \), and when a new \( x \) is given, the function can produce the appropriate approximate value. The SVR function is expressed in Eq. (1).

\[
f(x) = \sum_{i=1}^{n} w_i \phi_i(x) + b
\]  

\( \phi(x) \) is a feature nonlinear mapping from the input space \( x \); \( b \) is constant, and \( w \) is denoted as the weight vector estimated by minimizing the risk function can be shown as \( w = \sum_{i=1}^{n}(\alpha_i - \alpha_{i}^*)\phi(x_i) \). The study strategy of SVR is to minimizing the expected risk. SVR performs linear regression in the feature space to lower the expected risk using \( \epsilon \)-insensitive loss and, at the same time try to reduce the complexity of the model by minimizing \( |w|^2 \) [21]. In Eq. (2), it can be realized where \( \xi^+, \xi^- (i=1, \ldots, n) \). They are the non-negative slack variables, representing the deviation between the function value and the actual value.

\[
\begin{align*}
\text{Min}: & \frac{1}{2}\|w\|^2 + C \sum_{i=1}^{n}(\xi_{i}^+ + \xi_{i}^-) \\
\text{st.} & \quad y_i - w_i \phi(x_i) + b \leq \varepsilon + \xi_{i}^+ \\
& \quad w_i \phi(x_i) + b - y_i \leq \varepsilon + \xi_{i}^- \\
& \quad \xi^+ + \xi^- \geq 0
\end{align*}
\]  

\( C \) in equation (2) is a regular constant and is considered to determine the trade-off between empirical errors and functional evenness. Therefore, the general form of SVR is based on a regression function, written as:

\[
f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\phi(x_i)\phi(x) + b = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) K(x_i, x) + b
\]  

where \( K(x_i, x) \) in equation (3) is called the kernel function. Although several options for Kernel functions are available, many are used, namely the Radial Basis

<table>
<thead>
<tr>
<th>Date</th>
<th>Input Variable</th>
<th>CPI</th>
<th>Money supply</th>
<th>BI rate</th>
<th>Exchange rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jan-09</td>
<td></td>
<td>9.17</td>
<td>113.78</td>
<td>1.859.891,00</td>
<td>8.75</td>
</tr>
<tr>
<td>Feb-09</td>
<td></td>
<td>8.6</td>
<td>114.02</td>
<td>1.890.430,00</td>
<td>8.25</td>
</tr>
<tr>
<td>Mar-09</td>
<td></td>
<td>7.92</td>
<td>114.27</td>
<td>1.909.681,00</td>
<td>7.75</td>
</tr>
<tr>
<td>Apr-09</td>
<td></td>
<td>7.31</td>
<td>113.92</td>
<td>1.905.475,00</td>
<td>7.5</td>
</tr>
<tr>
<td>May-19</td>
<td></td>
<td>3.32</td>
<td>137.40</td>
<td>5.860.641,40</td>
<td>6</td>
</tr>
<tr>
<td>Jun-19</td>
<td></td>
<td>3.28</td>
<td>138.16</td>
<td>5.918.515,27</td>
<td>6</td>
</tr>
</tbody>
</table>
Function (RBF) function [20][21]. Thus, the RBF function is applied as a Kernel function. The coefficient $\alpha_i$ and $\alpha_i^*$ can be obtained by solving the problem of convex quadratic programming that shows in Eq. (4).

$$\text{Max: } -\frac{1}{2}\sum_{i,j=1}^{n}(a_i - a_j)(a_j - a_j^*)K(x_i, x_j) - \epsilon \sum_{i=1}^{n}y_i(a_i - a_i^*) + \sum_{i=1}^{n}a_i\gamma_i(a_i - a_i^*)$$

(4)

with the limits: $\sum_{i=1}^{n}a_i - a_i^* = 0$ dan $0 \leq a_i, a_i^* \leq \epsilon$. The most frequently used kernel functions are polynomial kernel function, sigmoid kernel function and the Radial Basis Function (RBF). This research uses a radial basis function, which is defined as Eq. (5):

$$K(x, z) = \exp \left( \frac{|x-z|^2}{2\gamma^2} \right)$$

(5)

Based on Eq. (5), the parameter $\gamma$ is needed to be manually set, as the same as $\epsilon$ and $C$ in Eq. (2), all of which have much influence on the prediction accuracy of SVR.

### 2.4 Genetic Algorithm

The application of the Genetic Algorithm is inspired by the research of the biological systems from the computer simulation. Evolutionary biology is GA origins which used to find approximate solutions for optimization problems [22]. A population in GA consist of some chromosome that represents the possible solutions [23]. There are three main processes in GA to create a new generation in each iteration, there are selection, crossover, and mutation [24]. The three processes in GA’s basic principles make an exploration and exploitation in the feasible solutions [25]. GA consist of three process [26]:

a. A population generated that consists of some random individual chromosome that has contexture of specific genes.

b. The fitness value calculated by the value of each individual.

c. The reproduction process conducted to produce offspring by operating crossover and mutation.

The individuals selected from a population that consists of parents and offspring to stopover for the next generations to recover the old population.

### 2.5 Hybrid GA-SVR

We proposed a classical GA to solve an optimization problem in SVR parameters. A genetic algorithm can produce the best solution for complex and complicated problems [22]. This approach is purposed to give a solution to generate the optimal value of SVR parameters as the main model to produce the model and value prediction. Parameter of SVR that we used in this study is $C$ (Complexity), $\epsilon$ (epsilon), and $\gamma$ (gamma). The procedure for establishing the GA-SVR model shown in Figure 1.

The hybrid GA-SVR process begins from the data configuration. The data set is divide into two types which are training data and testing data using cross-validation. Next, determine the parameters used in prediction using SVR. The parameters that used are $C$ (Complexity), $\epsilon$ (epsilon), and $\gamma$ (gamma). The initialization of the population consists of chromosome representation. The population consists of a set of individuals delineates by some chromosome [27]. The chromosome is random as many as population size [28]. The genes in the chromosome contain SVR parameters that consist of $C$, $\epsilon$, and $\gamma$. The representation of the chromosome in this research using real coded representation. The illustration of chromosome representation shown in Figure 2.

There are three operations in GA, selection, crossover, and mutation which crossover and mutation are genetic operators that called the reproduction process. Crossover is a section of the reproductive manner in GA that requires some strategy to select two parents from the previous generation. The part of GA that also important is mutation. It has a few random modify executed in each chosen chromosome [29] to obtain new varieties of a chromosome. The crossover and mutation processes will
generate a combination of various features to reach different directions in the search space [30]. The parameter used in the crossover operator is crossover-rate (cr) and the mutation operator used is mutation-rate (mr).

![GA-SVR prediction model](image)

**Figure 1 GA-SVR prediction model**

<table>
<thead>
<tr>
<th>C</th>
<th>ε</th>
<th>γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.01</td>
<td>0.001</td>
</tr>
<tr>
<td>1</td>
<td>0.02</td>
<td>0.9</td>
</tr>
<tr>
<td>10</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>20</td>
<td>0.03</td>
<td>0.5</td>
</tr>
<tr>
<td>100</td>
<td>0.0001</td>
<td>0.1</td>
</tr>
</tbody>
</table>

**Figure 2 Representation of chromosome**

These are used to determine the number of a new chromosome with \( cr \times popsize \) to generate offspring from the crossover process, \( mr \times popsize \) used to generate the mutation process [31]. The processes in GA operations will produce a new generation population. This research used a one-cut-point crossover and insertion mutation shown in Figure 3 and Figure 4.

![Crossover Illustration](image)

**Figure 3 Crossover Illustration**
The crossover and mutation calculated to produce new generation population. The fitness value calculated using the Mean Square Error (MSE) in SVR model result. The fitness function for this case shows in (6)

\[ fitness = \frac{1}{MSE} \]  

In the selection process, the following number of chromosome is chosen based on the fitness value as many as the number of population size. The selection process is done to pick individuals from a set population that consists of parents chromosomes and the offspring from the reproduction process to generate a new population. If the stopping conditions of fulfilled, SVR will get the solution as a value as the optimal parameters for each SVR parameters. It will be used for the final prediction process using SVR model and generate a value of prediction using GA-SVR.

3 Result and Discussion

This section shows the experimental result that consist of genetic parameters testing of GA as a optimization method. Testing is conducted in GA which consists of three testing types, (i) testing of population size, (ii) testing of the combination \( cr \) and \( mr \), and (iii) testing the generation number. This testing is done to obtain the optimal parameters of GA to optimize the SVR parameters value. The optimal GA’s parameters will be calculated to get the predicting value and the accuracy of prediction from SVR model that closes to actual data. Each scenario performed 10 times to obtain an average fitness value.

In testing population size use numbers with multiples of 10 starting from number 10 until 100. The \( cr \) value used is 0.5 and \( mr \) value is 0.5. The generation number used is 500. The parameters of SVR used is C of 0.1 to 1000, epsilon of 0.0001 to 1, and gamma of 0.0001 to 1. The result of population size testing shown in Figure 5.

\[
\begin{array}{c|c|c}
\text{p1} & 0.1 & 0.01 & 0.001 \\
\text{C1} & 0.1 & 0.001 & 0.02 \\
\end{array}
\]

Figure 4 Mutation representation

Figure 5 Test results of the population size

Figure 5 shows a optimal average fitness value produced by population size of 10 is 0.03838098 as a big population size that it can be obtain the optimum solutions to
produce high fitness value. In the testing of combination \(cr\) and \(mr\) values used to determine the most optimal combination of \(cr\) and \(mr\) to produce the best solution in this case. The generation number used is 500 and the population size used is 70 after testing. The SVR parameters used is \(C\) of 0.1 to 1000, epsilon of 0.0001 to 1, and gamma of 0.0001 to 1. The test results of the combination of \(cr\) and \(mr\) shows in Figure 6.

![Figure 6 Test result of the combination of \(cr\) and \(mr\)](image)

Figure 6 shows the optimum average of fitness value generate by the combination of \(cr\) is and \(mr\) is 0.5. If the combination of \(cr\) and \(mr\) are not well defined, hence the possibility of convergence may occur so it can not generate the optimal solution [31]. In testing generation number is done to get an optimal generation number to generate an optimal solution. The generation number used with multiples of 500 starting from number 100 to 4000. The other parameters used are the population size is 10, the \(cr\) is 0.5 and \(mr\) is 0.5. The SVR parameters used is \(C\) of 0.1 to 1000, epsilon of 0.0001 to 1, and gamma of 0.0001 to 1. The test result of the generation number shows in Figure 7.

![Figure 7 Test result of the generation number](image)
Figure 7 shows the generation number as much as 3500 generates the highest average of fitness value. In Figure 10, the generation of 100 to 500 has increased the fitness value, and decreased in the generation of 1000 and increased in the generation of 1500. In the generation of 2000, the fitness value has stable and decreased in the generation 2500. The generation of 3000 has increased the fitness value but it is not too significant. Next, in the generation of 3500, the fitness value has increased as the highest fitness value. In the generation of 4000, the fitness value has decreased as the lowest fitness value. If the number of generations is more, the longer computation time needful and the resulting solutions are not necessarily optimal [23].

Based on the testing of population size, the combination of cr and mr, and the generation number, the test result obtained the error value of population size is 10, the error value of the combination of cr is 0.5 and mr is 0.5, and the error value of the generation number is 3500. The testing result of population size, the combination of cr and mr, and the generation number used to find the best SVR parameters with the range used for each parameter are; C is 0.1 to 1000, epsilon is 0.0001 to 1 and gamma is 0.0001 to 1. After getting the value from each parameter, the next step is to predict using SVR model. The predicting result is MSE. The comparison of SVR and GA-SVR is shown in Table 2.

<table>
<thead>
<tr>
<th></th>
<th>SVR</th>
<th>GA-SVR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSE</td>
<td>0.05315841</td>
<td>0.03767636</td>
</tr>
<tr>
<td>Computational Time</td>
<td>0.007996s</td>
<td>691.7s</td>
</tr>
</tbody>
</table>

From Table 2, it can be seen that the computational time of GA-SVR is longer than SVR basic. However, the result of the error value by GA-SVR is smaller as 0.03767 than SVR basic as 0.05315. It is because the GA-SVR has a search space to generate the optimal solution. The more parameters that used, the more feasible the solution that produced.

4 Conclusion

Based on the experiment that has been done, it can be concluded that although the computational time required by GA-SVR is longer than SVR basic, the error value of GA-SVR is smaller than SVR basic. It shows that GA-SVR can solve the problems effectively by providing the optimal result than SVR basic. Therefore, GA-SVR can use as an alternative to predict the inflation rate in Indonesia.

References