Comparison Of Support Vector Regression And Autoregressive Integrated Moving Average With Exogenous Variable On Indonesia Consumer Price Index

Abd. Fattah Al Ghofur, Yuliani Setia Dewi, Dian Anggraeni

University of Jember, Jember, Indonesia

***Abstract* – CPI is one of the most frequently used indicators to measure the inflation rate in a region. The government can maintain economic stability by knowing the CPI value in advance. Therefore, we need a suitable method to predict an accurate CPI value. In this research, we investigate the prediction of CPI based on the machine learning method, SVR, and compare it to the ARIMAX method. We use Indonesia CPI data from January 2015 to October 2021. We investigate the SVR method using four kernel functions: Radial Basis Function (RBF), Polynomial, Linear, and Sigmoid. We build the ARIMAX model through the auto ARIMA process. We divide the data into two parts with three scenarios to investigate the performance of the methods: training and testing. The results show that the partition of 80% training and 20% testing gives the best performance. The SVR method performs best using a linear kernel, with an RMSE value of 0,743 and a MAPE value of 0,684%. The best ARIMAX model is model (0,2,1) with an RMSE value of 1,928071 and a MAPE of 1,731598 %. From the plot of prediction results and indicators of RMSE and MAPE, the SVR predicts CPI data better than the ARIMAX method, with CPI in the previous one-month data (MA1) being the most influential variable on the next CPI value.**

***Keywords* – IHK, SVR, ARIMAX, MAPE, RMSE.**

1. Introduction

The Consumer Price Index (CPI) is one of the most often used measures to quantify an area's inflation rate. It can be interpreted as a comparison of the prices of a commodity packages from a group of goods or services used by households over a certain period [1]. We can use index values to compare changes from period to period, in the form of years, months, days, and units of measurement [2]. The Consumer Price Index (CPI) is often used to measure a country's inflation rate and consider adjusting salaries, wages, pensions, and other contracts [3].

Prediction is the systematic estimation of something that is most likely to happen in the future based on past and current knowledge in order to reduce errors (differences between what actually and the predicted outcome) [4]. Prediction does not have to provide a definitive response to what will occur, but rather attempts to provide an answer that is as close to what will occur as possible [5].

One of the factors in maintaining economic stability is maintaining the CPI value. A good prediction of the CPI value is essential. Therefore, we need an accurate method to predict it. There are various statistical methods for predicting CPI values. In this research, we investigate the prediction of CPI based on the machine learning method, the Support Vector Regression (SVR), and compare it to ARIMAX (Autoregressive Integrated Moving Average with Exogenous Variable) approach. The SVR method can overcome the overfitting model, where the data classification is good in the training set process [6]. SVR is also used to get the lowest error rate and good prediction results [7], [8].

DOI: 10.18421/SAR53-05
https://doi.org/10.18421/SAR53-05

Corresponding author: Yuliani Setia Dewi,
University of Jember, Jember, Indonesia.
Email: yulidewi.fmipa@unej.ac.id

Received: 09 August 2022.
Revised: 06 September 2022.
Accepted: 12 September 2022.
Published: 27 September 2022.

On the other hand, the ARIMAX method is an extension of the ARIMA (Autoregressive Integrated Moving Average) method, which has other variables or is called exogenous variables. Adding other variables to the model can be considered accurate for forecasting [9], [10]. The addition of these variables affects the data and provides good forecasting accuracy in a study [11]. Previous research using four kernel functions, namely, polynomial, linear, SPLine, and RBF, performed by [12], [13], in the SVR method, resulted in the RBF kernel giving the lowest MAPE value. CPI research conducted by [14] uses two methods, namely EEMD and SVR, and produces a good RMSE value. Another study on stock price closing predictions conducted by [15] describes the use of new variables created using existing variables by comparing the Artificial Neural Network (ANN) and Random Forest (RF) methods. ANN gives better predictive results than RF. Research using the ARIMAX model by [16] to predict dengue fever by adding an exogenous variable, namely Google Trend data, results in a MAPE value of 3%. The research by [17] examines the effect of the Consumer Price Index and inflation on interest rates. The results show that the CPI has a negative influence on interest rates. CPI research was also conducted by [18], using the Neural Network and Quantile Regression Neural Network methods on CPI data to produce a good RMSE value.

In this study, we investigate the model based on the machine learning method, SVR, and ARIMAX methods in predicting the CPI values and find the factors that influence it. As for the SVR method, we use four kernel functions: linear, polynomial, radial, and sigmoid. In the ARIMAX method, we build the models using the auto ARIMAX function.

2. Method

We use the data from the website of the Central Statistics Agency (BPS), regarding the CPI Indonesia, the period from January 2015 to October 2021. The dependent variable (y) is the CPI. We use several independent variables (x) in the SVR process: interest rates, moving average CPI for one month, moving average CPI for two months, moving average CPI for three months, and moving average CPI for six months. While in the ARIMAX time series modeling, we use the interest rate as an exogenous variable. To investigate the performance of the methods, we divide the data into two parts, training and testing. This research uses three scenarios of training: testing proportion, namely 80:20, 75:25, and 90:10.

We use RMSE and MAPE indicators to compare the performance of the two methods. We also use the R program to analyze the data.

3. Results and Discussions

We build the SVR and ARIMAX models using 83 monthly CPI data in Indonesia from January 2015 to October 2021.

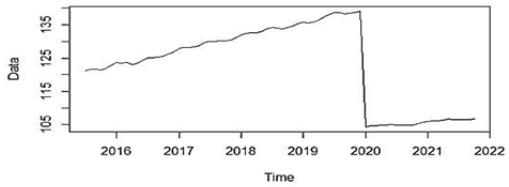

Figure 1. Plot Data

Figure 1 above shows that the CPI from July 2015 to December 2019 tends to increase. In January 2020, there was a very significant decline in CPI due to the impact of COVID-19 in Indonesia, and the economy in Indonesia was unstable.

1) SVR Prediction

a. Tune Parameter

Cost parameter adjustment aims to select cost parameters using training data to obtain the best model and prevent regression errors [19]. Cost parameters with different kernel types are given in Table 1.

Table 1. Tuning Parameter of SVR Models with Various Kernel

Cost	Kernel *Linear*	Kernel *Polynomial*	Kernel *Radials*	Kernel *Sigmoid*
0,001	75,1065	73,5223	97,6468	9,600666e+01
0,01	26,8914	44,9355	88,3935	7,434196e+01
0,1	20,9783	37,5205	60,0093	4,776243e+02
1	18,9802	38,0615	23,0206	4,839982e+02
5	18,7866	47,0768	22,8602	1,116790e+04
10	18,2174	64,4901	22,8457	4,307571e+04
100	19,0207	203,6787	22,5398	4,252946e+06

The best cost parameters for linear, polynomial, radial and sigmoid kernels are 10; 0,1; 100 and 0,1 with an error value of 18,21743; 37,52049; 22,53979 and 4,776243e+02, respectively.

b. Determining Kernel Function as the Best Parameter

We determine the best kernel function from the experimental results of training data using four kernels, namely Linear, Polynomial, Radial, and Sigmoid, with the best cost parameters given in Table 1 [20]. The smallest MAPE and RMSE values are selected as kernel functions to build the CPI model. The error percentage value in the training process producing the smallest error value is the linear kernel (Table 2). Therefore, we use the SVR method with a linear kernel *for further analysis.*

Table 2. MAPE and RMSE of SVR Model with Various Kernel

Kernel	MAPE	RMSE
Linear	**0,866**	**4,455**
Polynomial	2,377	5,287
Radial	0,983	4,302
Sigmoid	10,882	23,578

2) ARIMAX Prediction

a. Stationarity Test

From the stationary test, we obtain a p-value = 0,648. It means that the actual data is not stationary using significant level 0,05.

b. Differencing Step

In the differencing process, we use the auto ARIMA process. Based on the auto ARIMA process, we get several models listed in Table 3.

Table 3. Model ARIMAX p,d,q

Model	AIC	*differencing*
ARIMAX (1,0,1)	367,2074	d = 0
ARIMAX (0,0,0)	455,9017	
ARIMAX (1,0,0)	364,8922	
ARIMAX (0,0,1)	412,3677	
ARIMAX (2,0,0)	367,2011	
ARIMAX (1,1,1)	361,8976	d = 1
ARIMAX (0,1,0)	357,2331	
ARIMAX (1,1,0)	359,514	
ARIMAX (0,1,1)	359,5137	
ARIMAX (1,2,1)	358,5043	d = 2
ARIMAX (0,2,0)	392,8505	
ARIMAX (1,2,0)	377,0763	
ARIMAX (0,2,1)	**356,2336**	
ARIMAX (0,2,2)	358,5029	
ARIMAX (0,3,0)	450,1384	d = 3
ARIMAX (1,3,0)	418,5686	
ARIMAX (2,3,0)	404,9812	
ARIMAX (3,3,0)	397,2886	
ARIMAX (4,3,0)	393,2892	
ARIMAX (5,3,0)	385,1011	
ARIMAX (0,4,0)	518,5866	d = 4
ARIMAX (1,4,0)	474,0774	
ARIMAX (2,4,0)	452,1915	
ARIMAX (3,4,0)	438,7911	
ARIMAX (4,4,0)	426,9303	
ARIMAX (0,5,0)	578,4288	d = 5
ARIMAX (1,5,0)	523,839	
ARIMAX (2,5,0)	494,7588	
ARIMAX (3,5,0)	474,2947	

Based on Table 3, the ARIMAX (0,2,1) is the best model; it has the smallest AIC value [21]. The model gives the ARIMAX (0,2,1) model estimation coefficient, as shown in Table 4 below:

Table 4. Coefficient of ARIMAX (0,2,1) Model

	ma1	Xreg
	-0,9550	-0,3970
s,e	0,0616	2,6344

3) *Comparison of SVR and ARIMAX*

We find the best model using the linear kernel in the SVR method, while in ARIMAX; we find the ARIMAX model (0,2,1) as the best model. Prediction results with 80% training and 20% testing data provide the best performance compared to other training and testing data scenarios (Table 5).

Table 5. SVR and ARIMAX performance comparison

Training-Testing	SVR		ARIMAX	
	RMSE	MAPE	RMSE	MAPE
75:25	1,1073	1,035%	2,8397	2,614%
90:10	0,8339	0,7337%	0,8511	0,791%
80:10	0,743	0,684%	1,9281	1,731%

Visually, the prediction results are given in Figure 2.

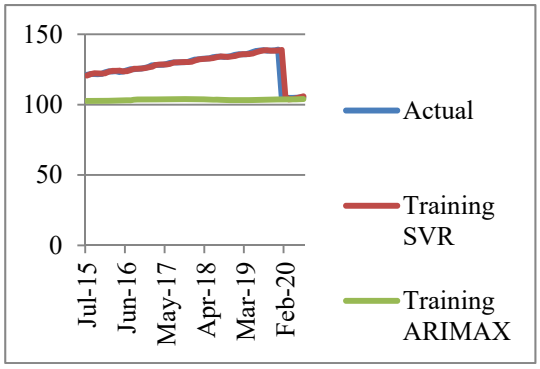

Figure 2. Prediction of training data

The prediction from SVR method results RMSE = 4,455 and MAPE = 0,866%. Figure 2 shows that the SVR prediction is close to the actual data. While predictions using ARIMAX results RMSE = 25,80315 and MAPE = 18,26541%. The pattern obtained in the prediction using ARIMAX does not seem to follow the actual data pattern; the pattern is increasing slowly. The prediction of testing data is given in Figure 3.

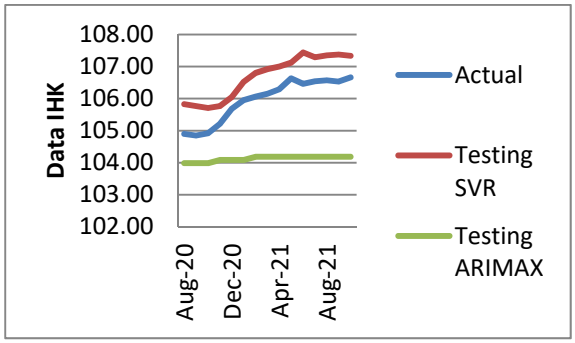

Figure 3. Prediction of testing data

The prediction results of testing data using the SVR method yield RMSE = 0,743 and MAPE = 0,684%. While the RMSE and MAPE using the ARIMAX method are 1,928071 and 1,731598%, respectively. The prediction results pattern from SVR is closer to the actual data than the results from the ARIMAX model.

3.1. Discussion

The results show that the SVR model provides better predictive results than ARIMAX based on RMSE and MAPE values. Based on the SVR model, several independent variables contribute to the CPI prediction (variable importance), with the order of the contributions shown in Figure 4.

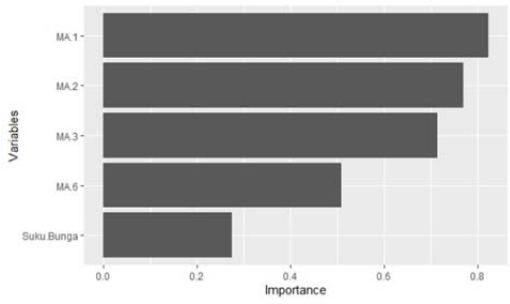

Figure 4. Variabel impotance of SVR Model

The 1month moving average (MA1) contributes the most to building the SVR model. The 2month moving average (MA2) and 3month moving average (MA3) have almost the same value as the MA1 value compared to other variables. Meanwhile, the 6month moving average (MA6) and interest rates have values that are not too high but have sufficient contribution to building the CPI model.

4. Conclusion

This research concludes that the best model of the SVR method is by using a linear kernel. It has the smallest error value among other SVR models. While in ARIMAX, the best result is the model (0,2,1). It has the smallest AIC value among other ARIMAX models. The SVR and ARIMAX models perform well based on RMSE and MAPE values. However, SVR performs better than ARIMAX, with smaller RMSE and MAPE; the predictions are close to their actual values. Based on the SVR model, the most influential variable on the CPI is the 1month moving average, which means that the CPI value in the previous month more influences the current CPI than other variables.

References

[1]. Permatasari, N. P., Chotimah, H., Permana, P., Tarigan, W. S., Toharudin, T., & Ruchjana, B. N. (2022). Application of GSTARI (1, 1, 1) Model for Forecasting the Consumer Price Index (CPI) in Three Cities in Central Java. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, *6*(1), 134-143.

[2]. OECD, The World Bank, The United Nations Economic Commission for Europe, Statistical Office of the European Communities, & Luxembourg. (2004). *Consumer price index manual: Theory and practice* . International Labor Organization.

[3]. Lesnussa, Y. A., Patty, H. M., Mahu, A. N., & Matdoan, M. Y. (2018). Analisis Indeks Harga Konsumen Terhadap Indeks Harga Sandang dan Pangan di Kota Ambon. *Euclid*, *5*(1), 100-107.

[4]. Chamdani, M., Mahmudah, U., & Fatimah, S. (2019). Prediction of Illiteracy Rates in Indonesia Using Time Series. *International Journal of Education*, *12*(1), 34-41.

[5]. Wanto, A., Zarlis, M., & Hartama, D. (2017, December). Analysis of Artificial Neural Network Backpropagation Using Conjugate Gradient Fletcher Reeves in the Predicting Process. In *Journal of Physics: Conference Series* (Vol. 930, No. 1, p. 012018). IOP Publishing.

[6]. Riofrío, J., Chang, O., Revelo-Fuelagán, E. J., & Peluffo-Ordóñez, D. H. (2020). Forecasting the Consumer Price Index (CPI) of Ecuador: A comparative study of predictive models. *International Journal on Advanced Science, Engineering and Information Technology*, *10*(3), 1078-1084.

[7]. Utami, T. M. A. N., Novitasari, D. C. R., Setiawan, F., Ulinnuha, N., & Farida, Y. (2021). Tide Prediction in Prigi Beach using Support Vector Regression (SVR) Method. *Scientific Journal of Informatics*, *8*(2), 194-201.

[8]. Yigit, M. E., Günel, T., & Günel, G. Ö. (2018). PSO based approach to the synthesis of a cylindrical-rectangular ring microstrip conformal antenna using SVR models with RBF and wavelet kernels. In *2018 18th Mediterranean Microwave Symposium (MMS)* (pp. 13-16). IEEE.

[9]. Chiu, L. Y., Rustia, D. J. A., Lu, C. Y., & Lin, T. T. (2019). Modelling and forecasting of greenhouse whitefly incidence using time-series and ARIMAX analysis. *IFAC-PapersOnLine*, *52*(30), 196-201.

[10]. A Adli, K. (2020). Forecasting steel prices using ARIMAX model: A case study of Turkey. *The International Journal of Business Management and Technology*.

[11]. Amelia, R., Dalimunthe, D. Y., Kustiawan, E., & Sulistiana, I. (2021, November). ARIMAX model for rainfall forecasting in Pangkalpinang, Indonesia. In *IOP Conference Series: Earth and Environmental Science* (Vol. 926, No. 1, p. 012034). IOP Publishing.

[12]. Rohmah, M. F., Putra, I. K. G. D., Hartati, R. S., & Ardiantoro, L. (2021). Comparison Four Kernels of SVR to Predict Consumer Price Index. In *Journal of Physics: Conference Series* (Vol. 1737, No. 1, p. 012018). IOP Publishing.

[13]. Rohmah, M. F., Putra, K. G. D., Hartati, R. S., & Ardiantoro, L. (2020). Minimum wage correlation with consumer price index predictions using support vector regression. *International Journal of Scientific and Technology Research, 9*(3), 4596-4602.

[14]. Qin, X., Sun, M., Dong, X., & Zhang, Y. (2018, September). Forecasting of China consumer price index based on EEMD and SVR method. In *2018 2nd International Conference on Data Science and Business Analytics (ICDSBA)* (pp. 329-333). IEEE.

[15]. Vijh, M., Chandola, D., Tikkiwal, V. A., & Kumar, A. (2020). Stock closing price prediction using machine learning techniques. *Procedia computer science*, *167*, 599-606.

[16]. Anggraeni, W., & Aristiani, L. (2016, October). Using Google Trend data in forecasting number of dengue fever cases with ARIMAX method case study: Surabaya, Indonesia. In *2016 International Conference on Information & Communication Technology and Systems (ICTS)* (pp. 114-118). IEEE.

[17]. Cristanti, I. L., Ismanto, B., & Sitorus, D. S. (2020). Pengaruh Indeks Harga Konsumen (IHK) dan Inflasi Terhadap Suku Bunga Tahun 2008-2018 Indonesia. *Ecodunamika*, *3*(2).

[18]. Rantini, D., Octavanny, M. A. D., Kruba, R., Kuswanto, H., & Fithriasari, K. (2018). Quantile Regression Neural Network Model For Forecasting Consumer Price Index In Indonesia. *Inferensi*, *1*(1), 39-43.

[19]. Maulana, N. D. (2018). *Implementasi Metode Support Vector Regression (SVR) Dalam Peramalan Penjualan Roti (Studi Kasus: Harum Bakery)* (Doctoral dissertation, Universitas Brawijaya).

[20]. Ghumman, A. R., Ahmad, S., & Hashmi, H. N. (2018). Performance assessment of artificial neural networks and support vector regression models for stream flow predictions. *Environmental monitoring and assessment*, *190*(12), 1-20.

[21]. Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.