Performance of Genetic Algorithm-Support Vector Machine (GA-SVM) and Autoregressive Integrated Moving Average (ARIMA) in Electric Load Forecasting

R. N. Hasanah[1*], D. Indratama[2], H. Suyono[3], M. Shidiq[4], M. Abdel-Akher[5]
Universitas Brawijaya, Indonesia[1,2,3,4,5]
Aswan University, Egypt[5]
*Corresponding Author Email: rini.hasanah@ub.ac.id

Abstract -- The main business focus of an electric power service provider is to meet the consumers' demand in time and quality as required. The increase of electrical load demand is influenced by various factors, such as the development of technology, business, region, the standard of life, climatic and weather changes, or consumer behavior. They must be considered by the power service provider to anticipate the load increase beyond the company's capability and the existing power generator capacity. This study focuses on comparing the performances of two methods in electric load demand forecasting. The Genetic Algorithm-Support Vector Machine (GA-SVM) and the Autoregressive Integrated Moving Average (ARIMA) methods are applied for the prediction of daily load, in Malang city, Indonesia, which is under the service coverage of the Indonesian national electricity provider, PT PLN Sub Unit P3B Jawa Timur-Bali. Two specific influencing factors, temperature, and precipitation are considered. The performance comparison is based on the error parameters of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). The study results indicate that the use of GA-SVM method provides better performance than the ARIMA method.

Keywords:
ARIMA
Autoregressive Integrated Moving Average
Genetic Algorithm
Load Forecasting
MAPE
RMSE
Support Vector Machine

Article History:
Received: May 19, 2020
Revised: May 24, 2020
Accepted: May 29, 2020
Published: June 3, 2020

DOI: 10.46962/forteijeeri.v1i1.07

I. INTRODUCTION

Electrical energy is vital in various sectors of daily activities. The electricity service providers always try to ensure that all installed loads of their customers can be provided by their existing power generators in time and quality as required. As the electrical load types and quantity are continuously changing, it is not easy to ensure the quality of service. The amount of electricity demand over some time cannot be calculated with certainty. A right prediction is mandatory to ensure the fulfillment of service standards. It is also to avoid the excessive and waste of electricity generation.

On the contrary, the generated and transmitted power is lower than or even does not meet customer needs. There could happen a local blackout being detrimental to the consumers. A suitable forecasting method would play an essential role in ensuring the economy and safety of power system operation, thus avoiding such undesired consequences.

The prediction of electric load demand must consider various influencing factors like the development of technology, business, region, the standard of life, climatic and weather changes, or even consumers behavior. In this study, the temperature and precipitation level factors are considered.

The purpose of forecasting is to estimate electricity load demand, to plan the power quantity to be produced and distributed from the generating plants, to plan maintenance schedules, and ultimately to ensure the security system of the electric power system operation [1].

An accurate load forecasting model is paramount in the planning and operation of the power system. It supports the electricity companies in decision making for supplying electricity, including regulating the generation, load switching, and infrastructure development [2].

Some known categories of forecasting methods are, for example, the conventional methods and the methods based on artificial intelligence theory. The artificial intelligence-based forecasting methods offer some more advantages, including the relatively easy upgrading and maintenance, the facility concerning incomplete input requirements, and the ability to produce reasoning like humans do [3].

In this study, a combination of the Genetic Algorithm (GA) and the Support Vector Machine (SVM) methods is chosen to be applied to get the best forecast results. The SVM method goes into the category of artificial intelligence-based methods, which is very popular for electrical load forecasting with excellent accuracy compared to other methods [4]. The SVM methods can perform well even for forecasting problems with a large number of inputs. The use of the GA method is applied to optimize the SVM parameter values. The GA method is known to have many successes in its use for optimization problems, besides being useful in solving a variety of complex problems, including electric power systems [5].

As a comparison, the implementation of the Autoregressive Integrated Moving Average (ARIMA) method will also be studied. It is a forecasting model that produces prediction based on the synthesis of historical data patterns. The ARIMA method can work well if the time series data used are of dependent properties, or statistically related to each other [6]. The advantage of the ARIMA method, which makes it very suitable to be used. Thus method predicts data because its relatively simple and easy to apply method to analyze data containing seasonal and trend patterns, and it's the ability to overcome the problem of randomness and even the cyclical nature of the analyzed time series data [7].

The GA-SVM and the ARIMA methods are applied for the prediction of the daily electricity load in Malang city, Indonesia, which is under the service coverage of the Indonesian national electricity provider, PT PLN Sub Unit P3B Jawa Timur-Bali. Two specific influencing factors, temperature, and precipitation are considered. The performance comparison is based on the error parameters of Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE).

II. METHODS

A. Power Load Forecasting

Forecasting is a prediction to estimate the occurrence of events in the future. Electricity load forecasting is an important measure to ensure the quantity of energy supplied by the provider meet the customers' demand and compensate the energy losses to be borne by the system. Prediction can be either a qualitative or quantitative measure.

Electricity load forecasting based on the period is divided into three categories [8]: long-range, medium-range, and short-range forecasting. The long-range forecasting covers the prediction time of one year or even longer duration. Principally the macro-economic problems of the company may become the main determining factor in load forecasting. The medium-range forecasting normally includes the load prediction during weeks of up to a year, whereas the short-range forecasting may be done for an hourly rate of up to weeks.

B. GA-SVM Method Concept

In a nonlinear SVM method, the general performance depends on the setting of parameters C, λ, and ε. The C parameter specifies the sanction for the estimation error. The ε parameter controls the width of the sensitive zone. The ε value may affect the number of support vectors being used to construct the regression function. The parameter λ represents the kernel function's width and has a significant influence on the forecasting performance of the SVM method.

The GA is adopted to select the SVM parameters C, λ, and ε. Each chromosome's performance in this algorithm is measured based on the respective fitness function, which is defined using the average per error (Mean Average Percentage Error/MAPE). By searching for the chromosomes, it is possible to get the minimum MAPE for SVM and the values of C, λ, and ε. The overall process of the SVM parameter optimization using the GA is illustrated in Fig. 1.

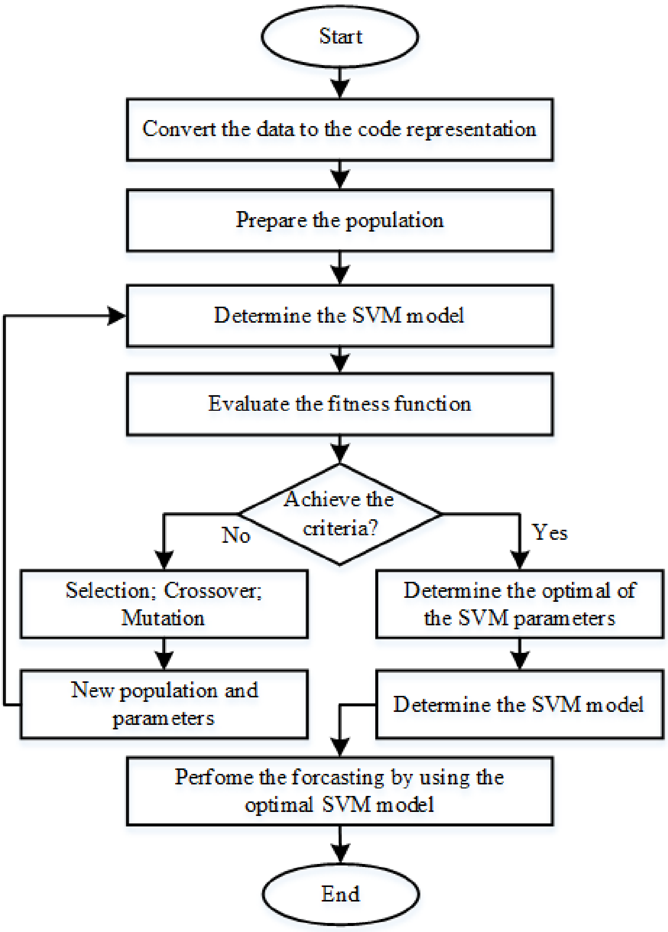

FIGURE 1.
The SVM parameters optimization using the GA method

As seen in Fig. 1, the SVM parameters optimization process using the GA method is started by coding the SVM parameters C, λ, and ε directly to produce chromosomes randomly. The chromosomes population, which is composed of parameters, is assigned randomly.

The following step is the evaluation of fitness function. The leave-one-out cross-validation is adopted in this study. The fitness function is defined as

$$\frac{1}{Y}\sum_{i=1}^{Y}\left|\frac{y_i-\hat{y}_i}{y_i}\right| \tag{1}$$

Where y_i and $\hat{y}_i$, represent the actual and validated values of C.

The next step is to replace the current population with the new population through the selection process, crossover operation, and mutation. The chromosomes which represent the possible useful solutions for recombination are chosen randomly based on the fitness function evaluation using roulette wheel selection. The genomes of two parent chromosomes are interchanged to produce new generation giving better solutions. The probability of creating a new chromosome in each pair is adjusted to be 0.8. The mutation step is then performed to change the binary code with a probability of 0.2. The whole steps are repeated until the specified stopping criteria are met.

C. ARIMA Method Concept

The ARIMA method is a combination of the autoregression (AR) method and the moving average (MA) method. The governing equation of this method can be expressed in (2).

$$Z_t = b_0 + b_1Z_{t-1} + b_2Z_{t-2} + \cdots + b_pZ_{t-p} + e_t - c_1e_{t-1} - c_2e_{t-2} - \cdots - e_{t-q} \tag{2}$$

An ARIMA(*p*,*q*) model uses the combination of the previous values and error. It is purposed to get a better model than just by using either the AR or MA models. The process of an autoregressive integrated moving average is symbolized with ARIMA(*p*,*d*,*q*). where *p* represents the autoregressive (AR) order, *d* the differencing process level, and *q* the moving average order.

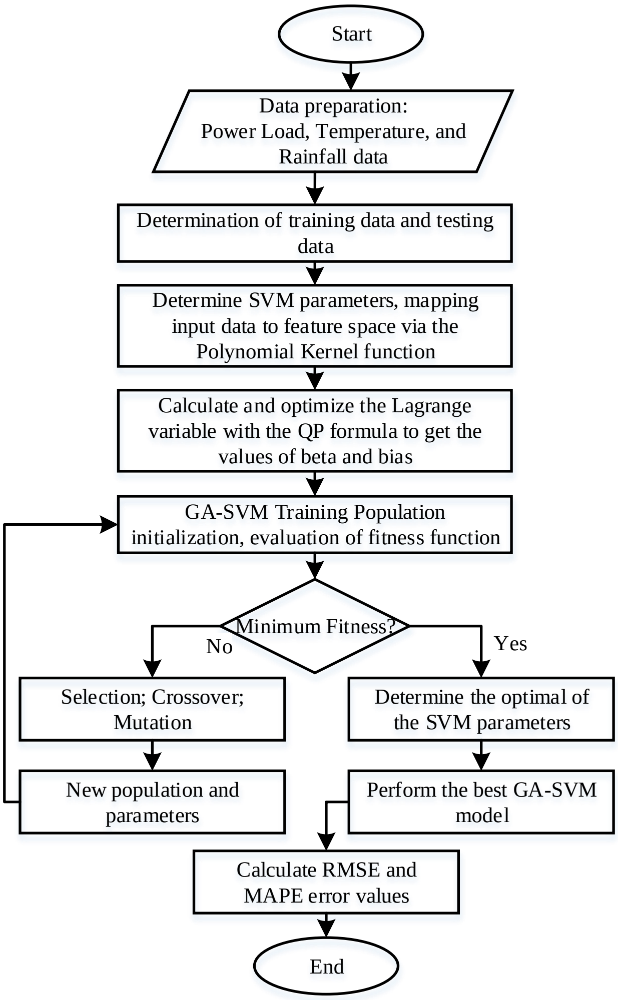

FIGURE 2.
The flowchart of GA-SVM method implementation

D. Error Calculation

Performance comparison of both forecasting methods in this study is based on the values of two types of error: the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE).

The RMSE has been commonly used as a standard statistical metric to measure the model performance during the load forecasting. It is calculated using (3).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} e_i^2} \tag{3}$$

where *n* represents the number of data while e_i is the error value (the predicted value – the observed value).

The calculation of Mean Absolute Percentage Error (MAPE) is based on each period’s absolute error with respect to the observed value. The respective mean average value of the absolute error is then found out. This approach is advantageous when the forecasting variables size is important to consider for forecasting accuracy. MAPE indicates the degree of forecasting error concerning to the actual value. It is calculated using (4)

$$MAPE = \frac{100}{n}\sum_{i=1}^{n} \frac{|f_i - t_i|}{t_i} \tag{4}$$

where *n* represents the number of data while f_i is the forecasted value, t_i is the actual value.

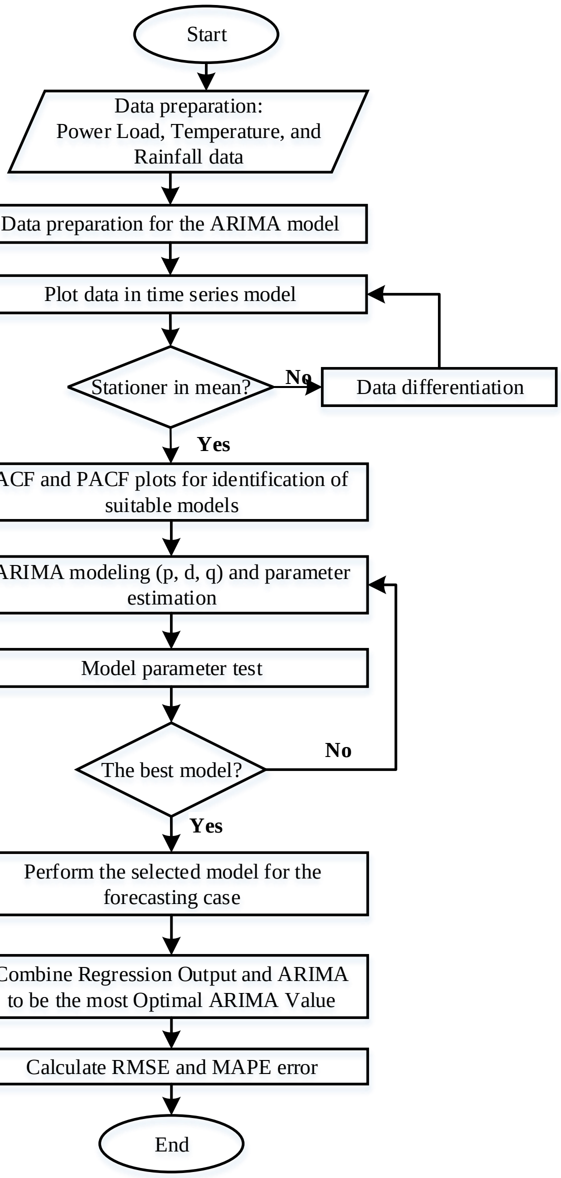

FIGURE 3.
The flowchart of ARIMA method implementation

E. Implementations on Load Forecasting

The process of electric power load forecasting has been divided into three steps. It started with building the forecasting model using the training data, followed with model validation using the testing data, and ended with the forecasting implementation. The training data has been taken from the daily load in Malang city, Indonesia, which is under the service coverage of the Indonesian national electricity provider, *PT PLN Sub Unit P3B Jawa Timur-Bali*, during the period of January-December 2018.

The application of the GA-SVM method for electric power load forecasting can be described using the flowchart given in Fig. 2, whereas that of the ARIMA method is given in Fig. 3.

III. RESULTS AND DISCUSSION

A. Results of Load Forecasting using the GA-SVM Method

Using the GA-SVM method, the C values have been specified in the range of 1-100, λ values in the range of 0.0001-0.01, whereas ε values in the range of 0.0001-0.01. Iteration has been done ten times.

The computation results using the GA-SVM method showed that minimum of error values of RMSE and MAPE had been achieved with the values of parameters C=85.208, λ = 0.0099 and ε = 0.0087, giving the RMSE=15.5947 and MAPE=0.0830, as seen in Table I. In contrast, the plot of forecasting results are given in Fig. 4.

TABLE I
The obtained RMSE and MAPE values during the implementation of GA-SVM method in load forecasting

RMSE	MAPE
15.5947 MWh	0.0830 MWh

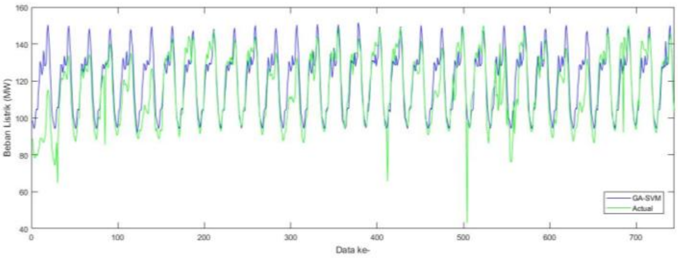

FIGURE. 4.
The plot of the load forecasting results using the GA-SVM method

B. Results of Load Forecasting using the ARIMA Method

The ARIMA method implementation uses the regression results as inputs. The regression is adopted to perform the electrical load prediction using multi inputs. The regression method computation has been implemented using the mathematical equation presented in (4).

$$Y = 1.5269 + 4.9277X_1 + 4.8056X_2 \tag{4}$$

Equation (4) has been used to predict the electrical power load during a specific period. Y variable implies that the load prediction is influenced by X_1 and X_2 variables, representing the predicted temperature and precipitation in the future. The regression results will be considered as L_s, which is combined with the ARIMA method prediction. The results of regression on the power load, temperature, and precipitation are presented in Fig. 5.

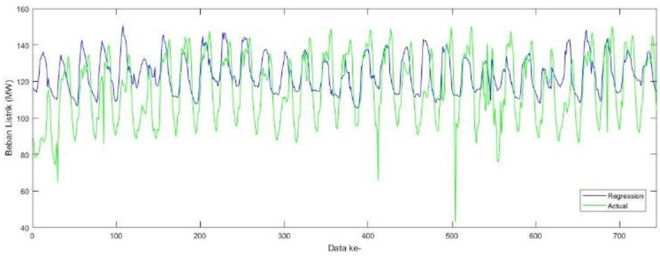

FIGURE 5.
The plot of regression results

Stationary testing has been performed using the Augmented Dickey-Fuller Test (ADF Test). The testing criteria require that data are considered stationary. Whenever the value of p is less than or equal to the level of significance (alpha (α) = 5%), they are not stationary whenever the value of p is higher than the level of significance (alpha (α) = 5%), which would require the transformation to be done on the data (differencing). The results of stationary testing can be seen in Table II.

TABLE II.
Augmented Dickey-Fuller Test results

Data Integration	Statistical ADF	*p* value
I(0)	-7.1265	0.001

As observed from Table II, it is known that the statistical ADF value at the level I(0) was -7.1265 with *p* value of 0.001. This result indicates that the *p* value is less than the level of significance (alpha (α) = 5%), which means that the data are of stationary type and does not need any differencing.

The ARIMA model identification is aimed at getting the ARIMA(*p*,*d*,*q*) model from the stationary data. The model can be found from the number of significant lags. They are outside Bartlet's limits (±1/√n). The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) values. The significant lags of ACF would form the MA(*q*) model, whereas the significant lags of PACF would form the AR(*p*). The number of differences being not utilized to transform the non-stationary data was stated as the *d* order. The resulted ACF and PACF values of the ARIMA model can be found, as displayed in Fig. 6 and Fig. 7.

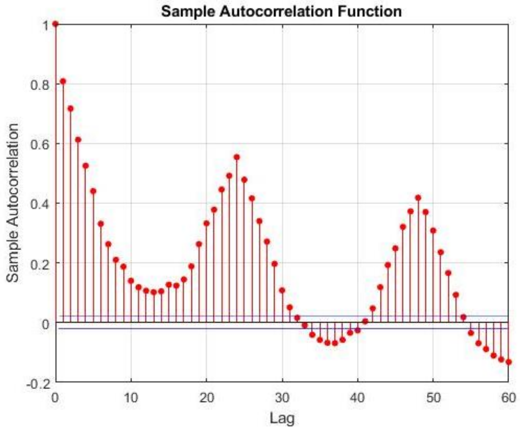

FIGURE 6.
Plot of ACF

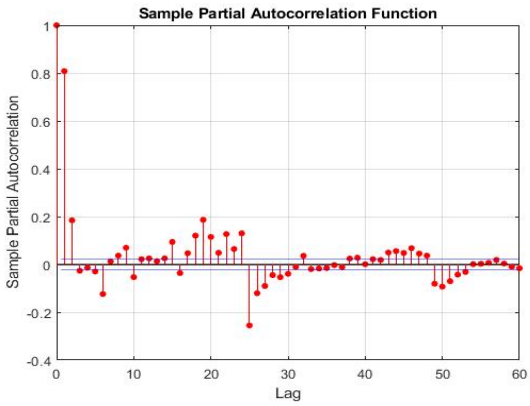

FIGURE 7.
Plot of PACF

As shown in Fig. 6, the whole resulted plot of ACF indicates an exponential form, whereas in Fig. 7 the plot of PACF implies the cut off on the seasonal lags. Overall, it confirms the occurrence of the Seasonal Autoregressive (SAR) process.

The plots of ACF and PACF have been used during the initial model identification. As seen in the ACF plot, a line indicating the seasonal pattern has been identified exactly on the 24th lag, giving the value of *q*=1. On the PACF plot, significant lines have been found on the 25th lag and the 1st lag, giving the value of *p*=1. Based on the observation of the ACF and PACF plots, the estimated ARIMA model was ARIMA(25,1,1)$(1,0,12)_{24}$, as shown in Table III.

TABLE III.
Estimation results using ARIMA model (25,1,1)$(1,0,12)_{24}$

Parameter	Estimation	Standard error	Statistical T	*p* value
θ_0	0	0	NaN	NaN
θ_1	-0.48963	0.0034241	-143	0
θ_2	0.0093655	0.0060891	1.5381	0.12403
θ_3	-0.058019	0.0045546	-12.739	3.6098e^{-37}

The related empirical model of ARIMA(25,1,1)(1,0,12)$_{24}$ is expressed using in (5).

$$\Delta Y_t = \theta_0 + \phi_1 \Delta Y_{t-25} - \theta_1\, \varepsilon_{t-1} - \theta_1\, \varepsilon_{t-12} + \varepsilon_t \qquad (5)$$

By introducing the respective values in Table III, (6) is obtained.

$$\Delta Y_t = 0 - 0.058019\Delta Y_{t-25} + 0.48963\varepsilon_{t-1} - \ 0.0093655\varepsilon_{t-12} + \varepsilon_t \qquad (6)$$

Finally, based on Table III and (6) it is found out that the parameters θ_1 and ϕ_{25} resulted in the *p* value less than the level of significance (alpha (α) = 5%), which indicated that the influence of the parameters was significant.

The identified best model of ARIMA(25,1,1)(1,0,12)$_{24}$ was used to forecast the load demand during the coming 31 days. The plot of forecasting results using the ARIMA(25,1,1)(1,0,12)$_{24}$model is presented in Fig. 8.

The ARIMA method has a univariate forecasting characteristic, which can only be used to predict one variable. A regression method is needed to enable the ARIMA method to be used for load demand forecasting with temperature and precipitation levels as predictor variables. In this way, the combination of the ARIMA method and regression results in better prediction with higher accuracy and smaller or tolerated error values.

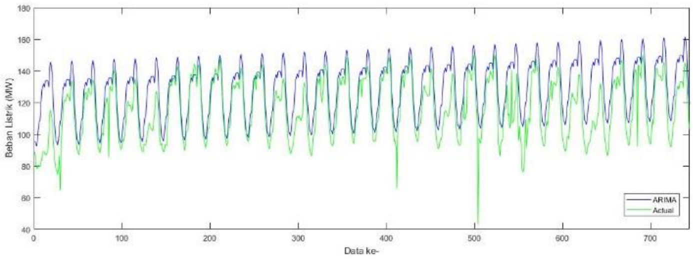

FIGURE 8.
Prediction results using ARIMA model of (25,1,1)(1,0,12)$_{24}$

The combination of both forecasting methods can be expressed using (7) with L_t the average of L_h and L_s. L_h represents the forecasting of historical load, whereas L_s is the result of load forecasting by considering the temperature and precipitation level.

$$L_t = \frac{L_h + L_s}{2} + e \qquad (7)$$

The results of the combination between the regression process and ARIMA are then used to calculate RMSE and MAPE's error values, as given in Table IV.

TABLE IV.
The obtained RMSE and MAPE values during the implementation of the ARIMA method in load forecasting

RMSE	**MAPE**
19.0165 MWh	1.7422 MWh

As seen, the implementation of the ARIMA method produced the error values of 19.0165 for RMSE and 1.7422 for MAPE.

The forecasting results using the ARIMA method being combined with the regression for the load demand, temperature, and precipitation level are shown in Fig. 9.

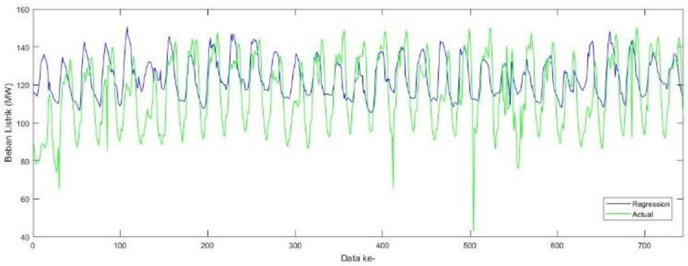

FIGURE 9.
Load forecasting results using the ARIMA method

C. Results Comparison of Load Forecasting using the GA-SVM Method and the ARIMA Method

The purpose of the comparison was to determine which method would provide better forecasting results, indicating the smallest error values. The Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE) have been chosen as parameters for comparison. Being implemented to predict the daily load of the national electricity provider in Indonesia, *PT PLN Sub Unit P3B Jawa Timur-Bali*, the performances of both methods have been investigated.

The smallest values of RMSE and MAPE using the GA-SVM method have been obtained using the parameters of C=85.208, λ = 0.0099, and ε = 0.0087. During the implementation of the ARIMA method, the respective error values have been achieved using the ARIMA model of $(25,1,1)(1,0,12)_{24}$. A comparison of the error values obtained using both methods is given in Table V.

TABLE V.
Comparison of errors during the implementation of both the GA-SVM method and the ARIMA method on the load forecasting

Error	**Methods**	
	GA-SVM	**ARIMA**
RMSE	15.5947 MWh	19.0165 MWh
MAPE	0.0830 MWh	1.7422 MWh

As can be observed in Table V, the implementation of the GA-SVM method to predict the daily electricity load of PLN Indonesia resulted in better performance, indicating smaller error values of RMSE and MAPE than the use of ARIMA method. The obtained MAPE values during the implementation of both methods fulfill the standard required by the PLN, which is 0.2.

The value is compared to the actual data; the prediction results of both methods based on the data used for testing are shown graphically in Fig. 10.

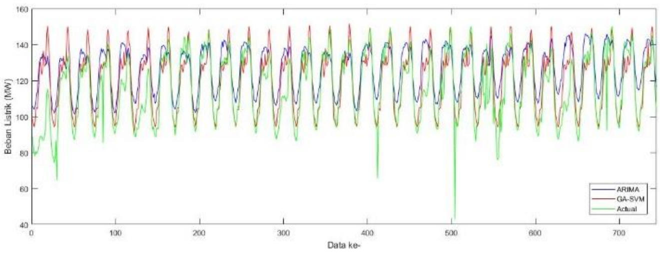

FIGURE 10.
A comparison of both forecasting methods results the actual condition

As seen, the load forecasting results using both the GA-SVM and ARIMA methods are almost similar. The predicted results approached well the actual data. It is also shown that both methods are convenient to be used by PLN, as the results enabled PLN to conform to the adopted standard.

IV. Conclusions

The analyses on the forecasting results implementing the Genetic Algorithm-Support Vector Machine method and the Autoregressive Moving Average method bring some conclusions as follows. The comparison of both methods can be based on the Root Mean Square Error (RMSE) and Mean Absolute Percentage Error (MAPE). Based on the RMSE and MAPE parameters, this study indicated the excellence of the GA-SVM method concerning the ARIMA method for load forecasting implementation. The obtained

RMSE and MAPE values were smaller during the implementation of the GA-SVM method than during the implementation of the ARIMA method. However, it seems that the number of data being used during the training process of forecasting implementation would influence the resulted error. Data of a longer duration, and comparison to other forecasting methods like the artificial intelligence-based ones are suggested in future research.

V. References

[1] X. Jin, Y. Dong, J. Wu and J. Wang, "An Improved Combined Forecasting Method for Electric Power Load Based on Autoregressive Integrated Moving Average Model," *2010 International Conference of Information Science and Management Engineering*, Xi'an, 2010, pp. 476-480, doi: 10.1109/ISME.2010.124.

[2] M-G. Zhang, "Short-term load forecasting based on support vector machines regression," *2005 International Conference on Machine Learning and Cybernetics*, Guangzhou, China, 2005, pp. 4310-4314 Vol. 7, doi: 10.1109/ICMLC.2005.1527695.

[3] S.A. Soliman and A.M. Al-Kandari, *Electrical load forecasting: modeling and model construction.* Burlington, MA: Elsevier, 2010.

[4] X. Li, D. Gong, L. Li and C. Sun, "Next day load forecasting using SVM," *ISNN'05: Proceedings of the Second international conference on Advances in Neural Networks* - Volume Part III, May 2005, Pages 634–639, https://doi.org/10.1007/11427469_101.

[5] M. A. A. Amin and M. A. Hoque, "Comparison of ARIMA and SVM for Short-term Load Forecasting," *2019 9th Annual Information Technology, Electromechanical Engineering and Microelectronics Conference (IEMECON)*, Jaipur, India, 2019, pp. 1-6, doi: 10.1109/IEMECONX.2019.8877077.

[6] W. Sun, "A Novel Hybrid GA Based SVM Short Term Load Forecasting Model," 2009 Second International Symposium on Knowledge Acquisition and Modeling, Wuhan, 2009, pp. 227-229, doi: 10.1109/KAM.2009.31.

[7] L. Wei and Z. Zhen-gang, "Based on Time Sequence of ARIMA Model in the Application of Short-Term Electricity Load Forecasting," *2009 International Conference on Research Challenges in Computer Science*, Shanghai, 2009, pp. 11-14, doi: 10.1109/ICRCCS.2009.12.

[8] S. N. Dodamani, V. J. Shetty and R. B. Magadum, "Short term load forecast based on time series analysis: A case study," *2015 International Conference on Technological Advancements in Power and Energy (TAP Energy)*, Kollam, 2015, pp. 299-303, doi: 10.1109/TAPENERGY.2015.7229635.