Jurnal Varian
Vol. 7, No. 2, April 2024, Hal. 171–182
e-ISSN: 2581-2017

Daily Rainfall Forecasting with ARIMA Exogenous Variables and Support Vector Regression

Regita Putri Permata[1], Rifdatun Nimah [1], Andrea Tri Rian Dani[2]
[1]Data Science, School of Computing, Telkom University, Indonesia
[2]Statistics Study Program, Department of Mathematics, Faculty of Mathematics and Natural Sciences, Mulawarman University, Indonesia

Article Info

Article history:

Received : 07-24-2023
Revised : 06-12-2024
Accepted : 06-28-2024

Keywords:

ARIMAX;
Rainfall;
RMSE;
SVR.

ABSTRACT

Every year, the dry season lasts from May to October and the rainy season lasts from November to April. However, climate change makes the rainy and dry seasons unpredictable, therefore weather forecasting is very important. The aim of this research is time series modeling to predict rainfall to predict weather patterns at six observation stations in Surabaya. The methods used to predict rainfall are ARIMAX and Support Vector Regression. Seasonal rainfall patterns are an external variable. The SVR model uses ARIMAX input lag and modifications to the Kernel Radial Based Function parameters. The best model is selected by minimizing RMSE. The results of the research are that the six rainfall observation locations experience rain on average in January, February, March, April, November and December. The ARIMAX approach in this research predicts Gubeng and Wonorejo rainfall effectively. SVR input lag The ARIMAX approach predicts rainfall for Keputih, Kedung Cowek, Wonokromo, and Gunung Sari well. SVR models use input lag, while ARIMAX has significant latency. So, this research improves rainfall predictions. Nonparametric algorithms can capture highly dynamic data patterns better than parametric algorithms such as SVR, thereby providing better rainfall forecasting.

Accredited by Kemenristekdikti, Decree No: 200/M/KPT/2020
DOI: https://doi.org/10.30812/varian.v7i2.3202

Corresponding Author:

Regita Putri Permata,
Data Science, School of Computing, Telkom University
Email: regitapermata@telkomuniversity.ac.id

A. INTRODUCTION

Surabaya is the capital city of East Java and is the largest city in East Java. According to the 2016-2021 Regional Medium Term Development Plan (RPJMD) for the City of Surabaya, the total area of the City of Surabaya is approximately 326.36 km2. Topographically, most of the city of Surabaya is lowland with a height of 3-6 meters above sea level at a slope of less than 3 percent. In Surabaya, the typical dry season runs from May to October, while the typical rainy season runs from November to April. Heavy rains usually occur between December and January (Nanlohy et al., 2019).

As a result of the phenomenon of climate change on earth, the alternation of the rainy and dry seasons becomes erratic. Uncertain seasons have an impact on drought events in various regions. This is due to the weather that is too hot and prolonged drought. When the rainy season arrives, it has an impact on the potential for flooding. The city of Surabaya itself is a lowland area so the potential for flooding is quite large, especially if the rainfall is high. Having information about when there is a lot of rainfall or vice versa is very necessary to anticipate uncertain weather conditions. Rainfall forecasting is used to see natural conditions for the future and is a human effort to see the development of rainfall conditions, now and in the future, especially in terms of anticipating them by optimizing efforts to monitor, collect, analyze data, so that it becomes a form of evaluation or prediction. rainfall intensity.

The time series forecasting for rainfall that is often used is the ARIMA or Autoregressive Integrated Moving Average method (Dani et al., 2023; Nanlohy and Haumahu, 2021; Ray et al., 2021). According to Dabral and Murry (2017) in his study is to apply

SARIMA modeling to model and forecast monthly, weekly, and daily rainfall series. Trend removal, periodicity, and stochastic component techniques are better integrated into time series modeling using the SARIMA approach (Dabral and Murry, 2017). Rainfall is included in the periodic series data, therefore the ARIMAX model, which is an enlarged ARIMA model including exogenous variables, may be used for forecasting (Suhermi et al., 2019). According to Amelia et al. (2021) in his research regarding forecasting rainfall in Pangkalpinang using ARIMAX showed that the forecasting results using the ARIMAX method were better than ARIMA (Amelia et al., 2021). Exogenous variables can be used in time series modeling to increase forecasting accuracy or to produce a more meaningful model. However, rainfall data often produces nonlinear data. The use of ARIMA is often unable to model nonlinear time series (Suci and Irhamah, 2017). Another study to find the best rainfall forecasting model in Bali using ARIMAX modelling (Nisa et al., 2021).

The Support Vector Regression (SVR) method can catch nonlinear cases by adding kernel functions. SVR is part of the Support Vector Machine (SVM) which is used for regression and prediction cases. One of the most important things in the SVR method to improve forecasting accuracy is the selection of inputs. According to study by Riyani et al. (2019) on input selection for forecasting using SVR, the optimal technique is the SVR approach employing the multiplicative lag of the best ARIMA model without dummy variables (Riyani et al., 2019). In 2020 will see studies comparing the effectiveness of parametric and non-parametric forecasting techniques. According to one conclusion, empirical research has demonstrated that machine learning techniques and straightforward parametric methods are quite competitive (Gautam and Singh, 2020). Abdullah et al. (2021) conducted a forecast comparison between SARIMA and SVM (Abdullah et al., 2021). The research findings indicate that SVM produces an accurate outcome with a low MAPE. This research offers innovation by combining ARIMA with exogenous variables and applying Support Vector Regression to enhance daily rainfall prediction, while also understanding influencing factors and comparing their performance. According to the research results of ref. (Makridakis et al., 2020) Classical methods are not inferior to machine learning, but it is not uncommon for machine learning methods to be superior. But machine learning SVR will be better if it does the right input selection (Riyani et al., 2019). The gap in this research compared to previous studies is the use of input selection for exogenous variables in ARIMA and the use of PACF lags as inputs for SVR. The aim of this research is to compare the performance of using the same lag inputs in the classical ARIMA method with machine learning SVR.

B. RESEARCH METHOD

The objective of time series analysis is to identify the underlying pattern within historical time series data. By uncovering these patterns, analysts can gain insights into the trends and behaviors exhibited over time. This information is crucial as it allows for making informed predictions about future values based on past observations. In the following section, several time series models are presented and discussed in detail, highlighting their unique approaches and applications:

1. Indonesian Halal Certification

Time series is a collection of observations from variables that have been made continuously recorded from time to time and in chronological order of events, where each observation as a random variable is obtained based on a certain time rate (t_i) in the order $Y_1, Y_2, \ldots, Y_n$. For the process follows $ARIMA(p,d,q)(P,D,Q)^S$ model, mathematic equation form can be written in equation (1).

$$\psi_P(B)\psi_P(B^S)(1-B)^d(1-B^S)^D Y_t = \theta_q(B)^{\theta}_{Q}(B^S)_{a_t} \tag{1}$$

or

$$Y_t = \frac{\theta_q(B)}{\psi_P(B)(1-B)^d} a_t \tag{2}$$

with

$\psi_p(B) = (1 - \psi_1 B - \psi_2 B^2 - \ldots - \psi_p B^p)$

$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$

B is backshift, i.e $B^k Y_t = Y_{t-k}$.

Some observation values may suffer a sharp spike or reduction due to the presence of model-affecting factors, and this phenomenon may recur throughout time and on various timelines. By using the ARIMAX time series model, this problem may be solved. The ARIMAX model is an expansion of the ARIMA model that increases forecasting accuracy by adding more variables or external factors that are believed to have a significant influence on the data (Siswanti and Yanti, 2020). There are several types of additional variables, for example dummy variables for monthly seasonal periods. The ARIMAX models can be written in Equation (3):

$$Y_t = \beta_1 M_{1,t} + \beta_2 M_{2,t} + \ldots + \beta_{12} M_{12,t} + \frac{\theta_q(B)\Theta(B^S)}{\psi(B)\Phi(B^s)(1-B)^D} a_t \tag{3}$$

2. Support Vector Regression (SVR)

TSVR, which stands for Support Vector Regression, is a variant of Support Vector Machine (SVM) that operates by employing linear function hypotheses within a high-dimensional feature space as part of its learning process. Unlike traditional regression methods, SVR utilizes an epsilon-insensitive loss function, enabling it to effectively handle outliers in the data. Following this explanation, the subsequent section delves into the intricacies of SVR's regression function, shedding light on its underlying mechanisms and practical applications in predictive modelling can be written in equation (4) and (5):

$$f(X) = w^T \phi(x) + b \tag{4}$$

Where w is weight and b is bias. The notation $\phi(x)$ is a point in feature space ζ. It represents the mapping of x in the input space. To reduce the following risk, use the coefficients w and b:

$$R(f(x)) = \frac{1}{2}\|w\|^2 \frac{C}{n} \sum_{i=1}^{n} L_s(y_i, f(x_i)) \tag{5}$$

where:

$$L_s(y_i, f(x_i)) = \begin{cases} 0 & \text{, for } |y_i - f(x_i)| \leq \varepsilon \\ |y_i - f(x_i)| - \varepsilon & \text{, other} \end{cases} \tag{6}$$

Polynomial and Radial Based Function (RBF) kernel functions are widely used in various machine learning algorithms. The RBF kernel function, in particular, plays a crucial role in many applications due to its flexibility and ability to capture complex relationships between data points. RBF Kernel can be written in equation (7):

$$K(x, y) - \exp\left(-\frac{(x-y)^2}{2\sigma^2}\right) \tag{7}$$

For instance, before conducting an experiment, one must choose the value of the parameter for the RBF. It's crucial to choose this setting. Figure 1 shows an illustration of the SVR model in time series research.

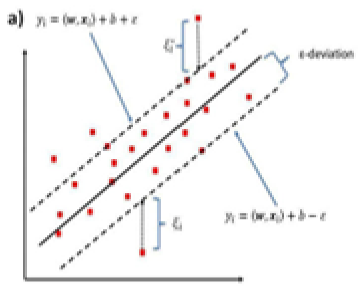

Figure 1. Support Vector Regression

The figure 1 illustrates Support Vector Regression (SVR) by showing the hyperplane, epsilon (ε) margin, and slack variables (ξ_i) used to handle prediction errors. The central black line represents the regression function $y_i = (w \cdot x_i) + b$, while the dashed lines depict the ε-deviation margins, within which no penalty is applied for errors. The red dots are data points, some within the ε-margin and others outside it. The slack variables ς_i indicate how far data points are from the ε-margin, allowing SVR to penalize data points that fall outside this margin.

DOI: https://doi.org/10.30812/varian.v7i2.3202

3. Model Selection

We use Root Mean Square Error (RMSE) as the criteria for model selection. The formula of RMSE is given respectively as follows in equation (8):

$$RMSE = \sqrt{\frac{\sum_{l=1}^{L}\left(Y_{n+l} - \widehat{Y}_n(l)\right)^2}{L}} \tag{8}$$

where Y_{n+l} denotes the actual value, $\widehat{Y}_n$ denotes the forecast value, and L denotes forecast horizon. Root Mean Squared Error (RMSE) memberikan bobot lebih besar pada kesalahan besar karena kesalahan dikuadratkan, sehingga mendorong model untuk meminimalkan kesalahan besar dan meningkatkan akurasi prediksi

4. Data and Input Variables

The data was obtained from the Department of Public Works of Highways and Pematusan (Department of Public Works and Pematus) of Surabaya City, PSAWS Buntung Peketingan Surabaya. The data used in this task is daily rainfall data from six posts in the city of Surabaya, namely the Keputih, Kedung Cowek, Gubeng, Wonorejo, Wonokromo, and Gunung Sari rainfall observation posts. The data period is from 1 January 2009 to 31 December 2018. Data from 1 January 2009 to 30 November 2018 was used as in sample data of 3621 data and from 1 December 2018 to 31 December 2018 used as out of sample data of 31 data. Table 1 shows the sample data used in this research.

Table 1. The sample Data

Date	Keputih	Kedungcowek	Gubeng	Wonorejo	Wonokromo	Gunung Sari
1/1/2009	9	6.7	5	7	8	12
1/2/2009	0	8.384	2	4	0	0
1/3/2009	6	7.896	13	0	15	16
⋮	⋮	⋮	⋮	⋮	⋮	⋮
12/30/2018	0	0	0	0	0	0
12/31/2018	0	0	0	0	0	0

This is input for ARIMA with exogeneous variables from dummy month. This input series includes exogenous variables, using a dummy month indicator, where the value is 1 for the current month and 0 for other months.

Input series:

$$M_{k,t} = \begin{cases} 1 & \text{, for this month} \\ 0 & \text{, others month} \end{cases}$$

Output series: each prediction of rainfall observation post in Surabaya, $t = 1, 2, \ldots, 6$.

The purpose of including important input identification variables and lag variables is to effectively capture the essential elements of the data producing process in a concise manner. The identification of trend and seasonality components in time series modeling is crucial, since it involves capturing the deterministic behavior of these components in the lags of the dependent variable (Riyani et al., 2019). Input selection for Support Vector Regression is lags obtained from the ARIMAX model which already has significant parameters.

This research flowchart describes the methodological steps and data analysis used in the research following the Figure 2.

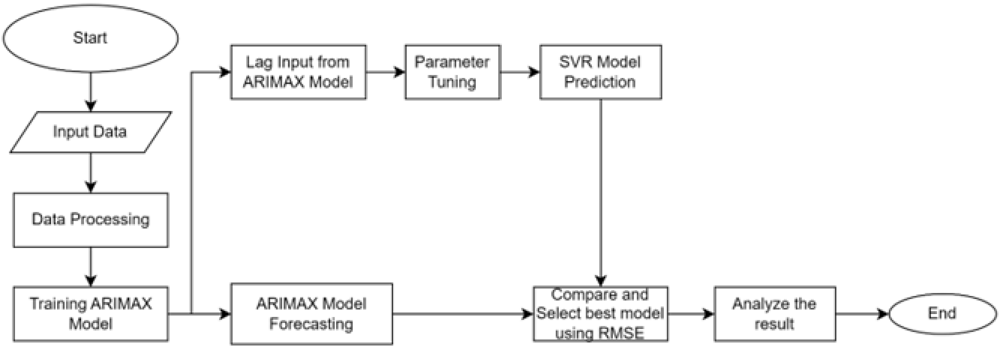

Figure 2. Flow chart of research

The steps used to analyze the data in this research is described as follows:

1. Input data: Perform data division into two groups in-sample and out-sample.
   Data from January 1, 2009 to November 30, 2018 was used as data in the sample of 3621 data. Data from December 1, 2018 to December 31, 2018 were used as out of sample data of 31 data.
2. Data processing: perform data transformation to facilitate input into the model.
3. Modeling and Forecasting Rainfall Using the ARIMA Method with Exogenous Variables (ARIMAX).
4. From the trained ARIMAX model, optimal time lags are identified and extracted. These lags will be used as inputs for the SVR (Support Vector Regression) model.
5. Adjust SVR model parameters to improve performance.
6. Compare prediction results from both models and select the best one based on RMSE.
7. Analyze prediction results to understand patterns and trends in the data.
8. Conclusions.

C. RESULT AND DISCUSSION

1. Characteristics at six Rainfall Intensity Observation Posts in Surabaya City

It is important to analyze the description of rainfall to find out the special characteristics of rainfall at the six observation posts in Surabaya. Recording of rainfall data at each post is carried out daily, so that there are two seasons, namely the dry season and the rainy season. The type of data exploration that is carried out is using descriptive statistics both in the form of numeric and graphical. The results of the descriptive statistics of rainfall at the six observation posts are shown in Table 2.

Table 2. Descriptive Statistics

Variable	Mean	Standard Deviation	Minimum	Maximum
Y_1	5.513	13.195	0	164
Y_2	5.132	11.401	0	123
Y_3	6.487	14.009	0	116
Y_4	5.86	13.782	0	122
Y_5	6.419	14.104	0	114
Y_6	6.361	14.052	0	120

According to data.bmkg.go.id information, the normal criteria for rainfall are divided into four categories, namely light (5-20 mm/day), moderate (20-50 mm/day), heavy (50-100 mm/day), and very dense (>100 mm/day). This criterion will be used as a reference in analyzing the characteristics of the six rainfall observation posts in Surabaya. Table 2 informs that during the last 9 years from 2009 to 2018, the average rainfall from six rainfall observation posts in Surabaya, namely Keputih, Gubeng, Wonorejo, Kedung Cowek, Wonokromo and Gunung Sari respectively 5,513; 5.132; 6,487; 5,860; 6,419; 6.361 millimeters per day is classified as light rain. Even though it is in the rainy season interval, the rainfall per day is relatively light.

Figure 3 shows the visualization of rainfall data at six observation posts over a 9-year daily period. Figure 3 shows that there are certain patterns that distinguish the dry and rainy seasons. The dry season is characterized by zero rainfall data. Figure 3 shows that the minimum data is zero where there is no rain or drought. Rain occurred at the six observation posts from 2009

DOI: https://doi.org/10.30812/varian.v7i2.3202

to 2018 in November to June, while the dry season was in May to October. The intensity of the heaviest rainfall at the Keputih observation post occurred on 30 May 2016 at 164 mm/day which resulted in flooding in the Keputih area. In addition, in the near future in 2017 at the Gubeng observation post there will be heavy rains with an intensity of 116 mm/day on November 24 2017.

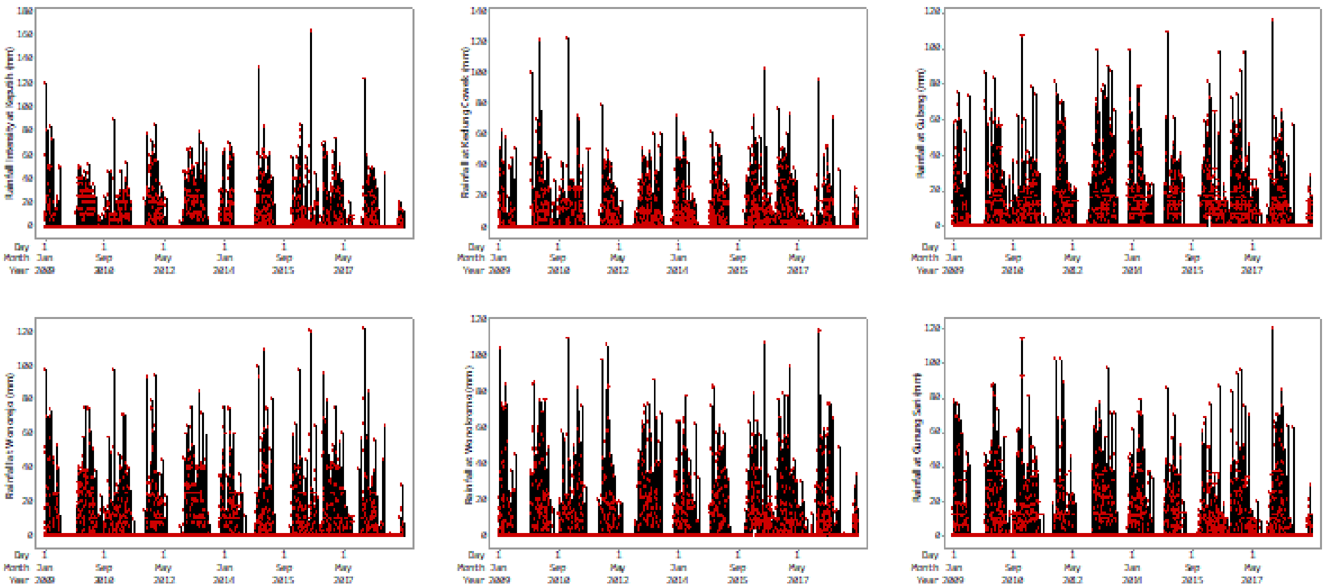

Figure 3. Time Series Rainfall Plot for six Observation Posts

Furthermore, to see the monthly rainfall conditions at the six observation posts, it can be seen visually by using plot intervals as shown in Figure 4. Figure 4 shows that January, February, March, April, November and December have high rainfall values. Meanwhile, May and June have low average rainfall values or there is a change in season from rainy to dry. While July, August, September indicate the dry season because the average data is zero and this pattern occurs on average at the six rainfall observation posts in Surabaya.

DOI: https://doi.org/10.30812/varian.v7i2.3202

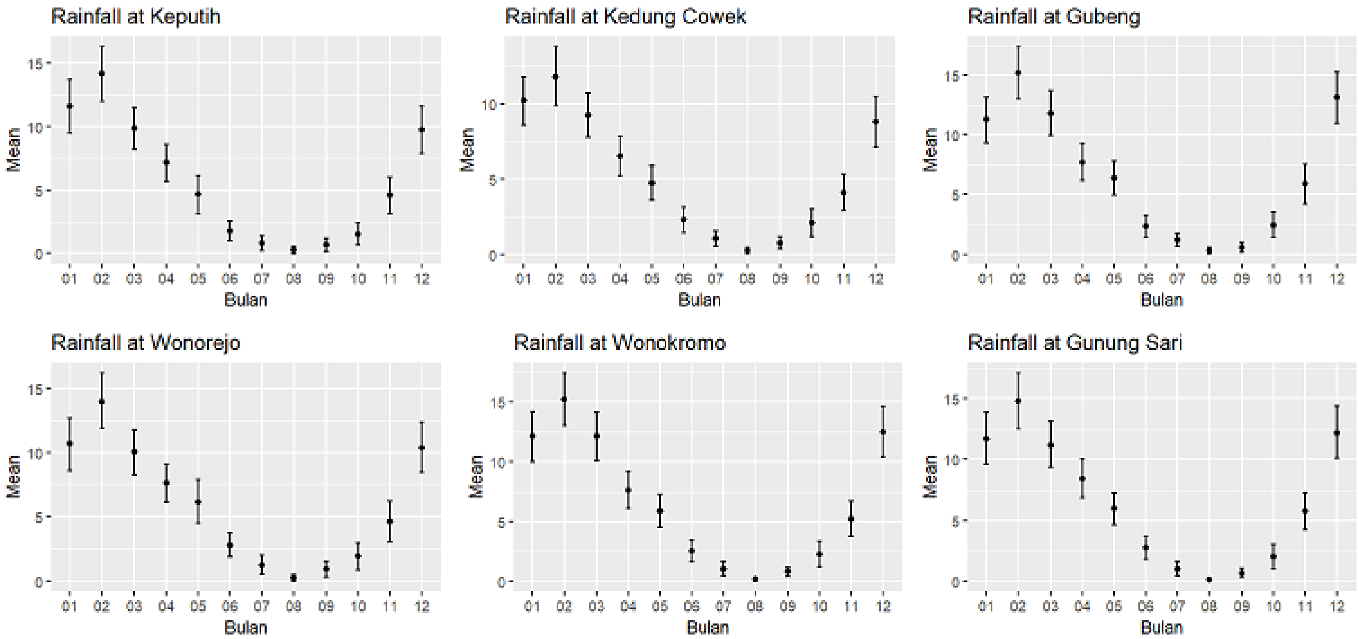

Figure 4. Monthly Rainfall plot intervals

2. Modeling and Forecasting Rainfall Using the ARIMA Method with Exogenous Variables

Rainfall modeling at six observation posts in Surabaya uses ARIMA with the addition of an exogenous variable in the form of a month dummy to assist in checking the presence of rainfall patterns in each month according to the plot interval visualization that has been done. So, the method used is ARIMAX. The exogenous variable used is the dummy month in order to be able to capture the rainfall pattern for that month. Modeling of rainfall data at six rainfall observation posts with ARIMAX is done by modeling the data using time series regression with a month dummy variable. The residuals from time series regression modeling are modeled using ARIMA so that an ARIMAX model with residual white noise is formed. The following is a PACF image of the residual time series regression results from six rainfall observation posts in Surabaya in Figure 5.

DOI: https://doi.org/10.30812/varian.v7i2.3202

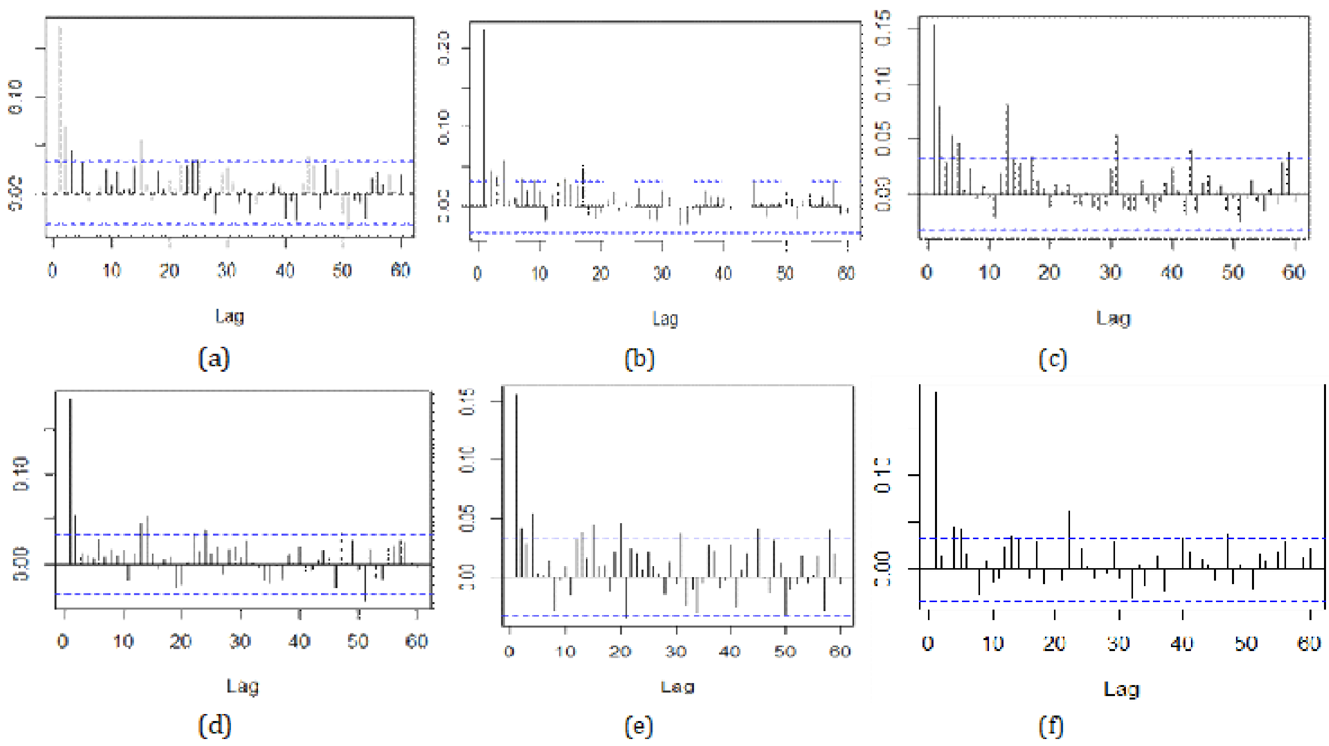

Figure 5. PACF Residual Rainfall Data (a) Keputih (b) Kedung Cowek (c) Gubeng (d) Wonorejo (e) Wonokromo (f) Gunung Sari

Figure 5 shows a significant PACF on the residual results of the time series regression model. The presence of this significant PACF suggests potential autocorrelation in the residuals, this helps to capture the lag pattern of rainfall. This PACF significant lag plot will be used to determine the ARIMAX model. From the PACF results, the significant lag that will be modeled using ARIMA is shown in Table 3.

Table 3. Descriptive Statistics

Variable	Model	Residual White Noise
Y_1	ARIMAX ([1,2,3,15],0,0)	Yes
Y_2	ARIMAX ([1,2,4,17],0,0)	Yes
Y_3	ARIMAX ([1,2,4,5,13],0,0)	Yes
Y_4	ARIMAX ([1,2,4,13,14],0,0)	Yes
Y_5	ARIMAX ([1,2,4,13,15],0,0)	Yes
Y_6	ARIMAX ([1,2,4,13,22],0,0)	Yes

Table 3 presents the ARIMA model applied to residual time series regression at six rainfall observation posts in Surabaya, which adheres to the residual white noise assumption. With the existing modeling results, Figure 6 depicts a comparison plot between actual data and predictions produced by the ARIMAX model using test data.

These findings provide insight into the model's performance and its ability to capture underlying patterns in rainfall data.

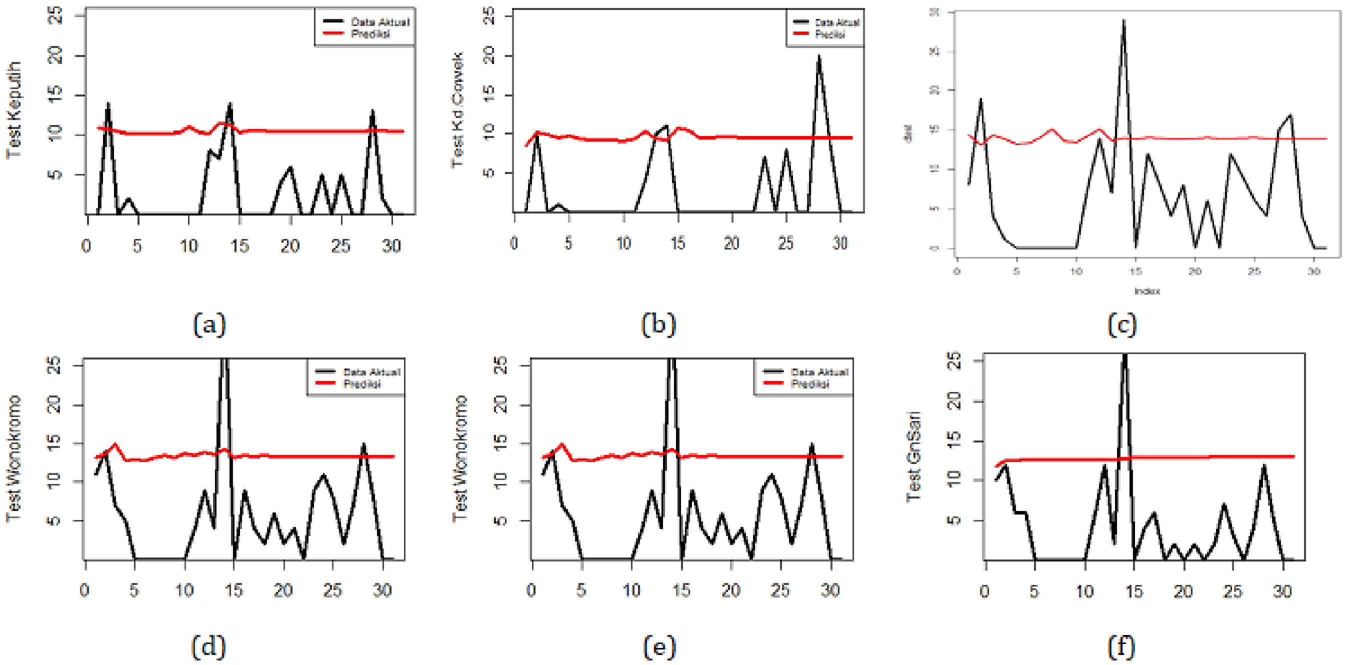

Figure 6. Forecasting Method ARIMAX Rainfall Data Testing; (a) Keputih (b) Kedung Cowek (c) Gubeng (d) Wonorejo (e) Wonokromo (f) Gunung Sari

Figure 6 reveals a noticeable disparity between the predicted data and the actual data during the testing phase. The discrepancy observed in the plot is attributed to the limitation of ARIMAX in accurately capturing the intricate rain patterns. These results emphasize the need for alternative modeling approaches to better account for the complexity of rainfall patterns.

3. Modeling and Forecasting Rainfall Using SVR

Forecasting with the SVR method will use the input lags obtained from the ARIMAX model which already has significant parameters. Parameter tuning in the SVR method uses a RBF with a gamma between 1/8 to 1/16, and an epsilon of 0.01 to 0.1. The following is the optimum result of tuning parameters shown in Table 4.

Table 4. Parameter Tuning Optimum

Variable	Sigma	Epsilon
Y_1	0.8	0.1
Y_2	16	0.1
Y_3	16	0.1
Y_4	8	0.1
Y_5	8	0.1
Y_6	16	0.1

Furthermore, SVR modeling was conducted for each observation post using the optimal RBF kernel parameters listed in Table 4. The comparison between actual data and SVR-predicted data using this method is illustrated in Figure 7.

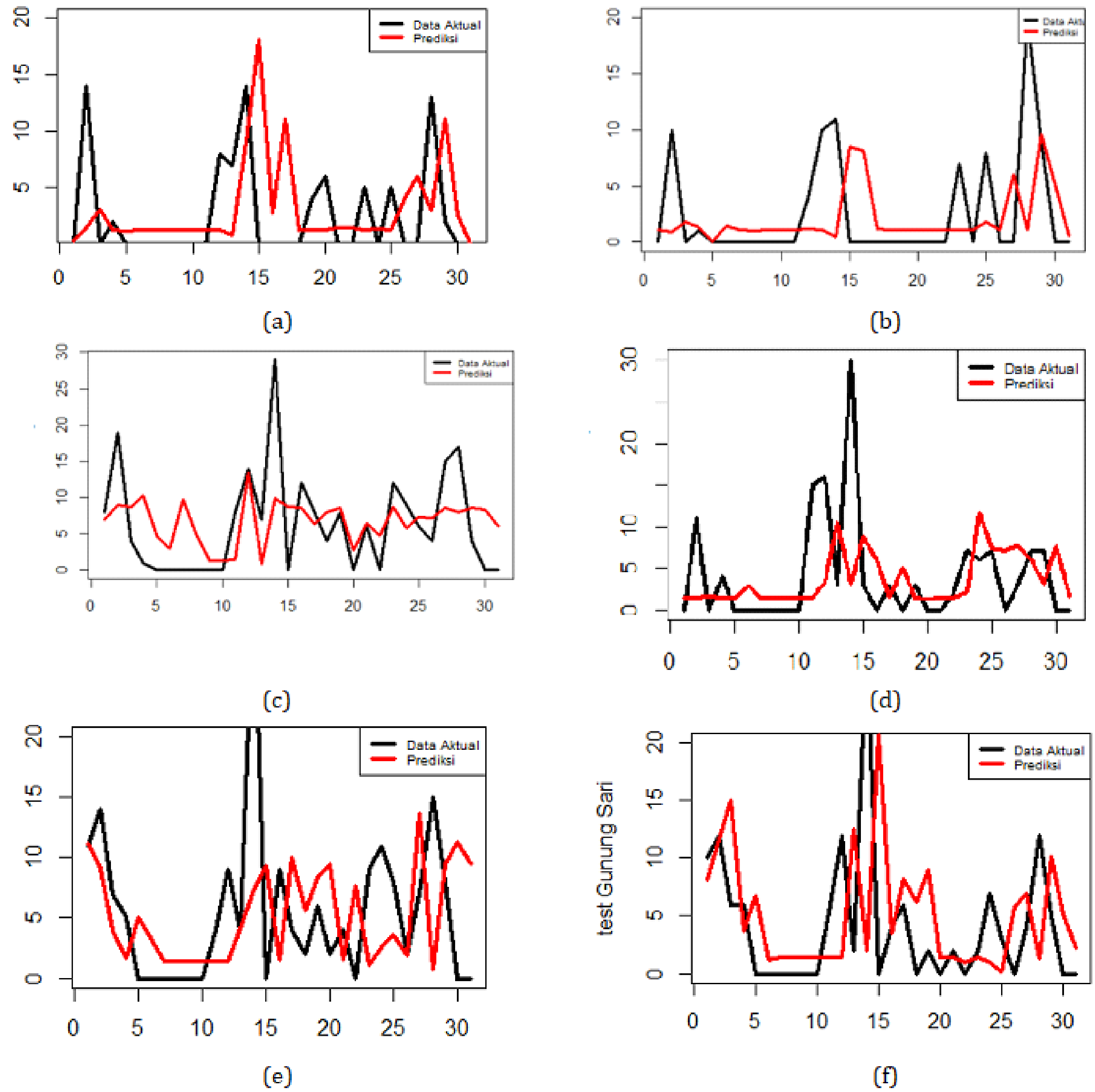

Figure 7. Forecasting SVR Rainfall Data Testing Method

Figure 7 shows that predictive data can follow actual data patterns. When compared with the predicted results on the ARIMAX model visually the SVR method is better. The findings of the research indicate that the forecasting conducted exhibits a lagging trend, as seen by the results depicted in Figure 5. It is observed that the projected trend aligns with the actual trend, but with a time delay, hence indicating a lack of accuracy in the prediction outcomes.

4. Comparison of Models Accuracy

After obtaining the best model for each method at each observation post, a comparison was made between the ARIMAX and SVR methods using the Root Mean Square Error (RMSE). Then the best method is selected based on the smallest RMSE on Testing. The following is a comparison of the accuracy values of all models presented in Table 5.

Table 5. RMSE comparison of the ARIMAX method and the SVR method

Variable	ARIMAX		SVR	
	RMSE in	RMSE out	RMSE in	RMSE out
Y_1	12.145	7.942	10.446	5.809
Y_2	10.415	6.974	8.769	5.48
Y_3	12.824	4.743	10.116	6.311
Y_4	12.824	6.76	10.577	7.063
Y_5	12.968	7.722	10.558	6.896
Y_6	12.968	9.012	10.369	7.732

In Table 5 the Keputih post (Y_1), the Kedung Cowek post (Y_2), the Wonokromo post (Y_5) and the Gunung Sari post (Y_6) follow the SVR model because the RMSE is the smallest outsample data. Meanwhile, the Gubeng post (Y_3) and Wonokromo post (Y_4) follow the ARIMAX model because the RMSE is the smallest out of sample data.

The results of the research show that machine learning models are better able to read trends and patterns of rainfall data. This is because the input lag used in the SVR model adjusts to the significant input lag of the ARIMAX model. So, this research is in accordance with the results of M4 forecasting research (Makridakis et al., 2020) and SVR input for forecasting (Riyani et al., 2019). As a result of this research to predict rainfall more accurately.

D. CONCLUSION AND SUGGESTION

The analysis that has been carried out finds that the average occurrence of rainy months at the six rainfall observation posts occurs in January, February, March, April, November and December. In the forecasting analysis that has been done, the best modeling is obtained by looking at the minimum RMSE value. The ARIMAX method in this analysis is used to predict rainfall in Gubeng and rainfall in Wonorejo. For the SVR method, the ARIMAX input lag is used to predict rainfall for Keputih, Kedung Cowek, Wonokromo and Gunung Sari. This is that the input delay applied in the SVR model aligns with the substantial input lag found in the ARIMAX model. this research leads to more precise rainfall predictions. Nonparametric methods are better used to forecast rainfall data because they are able to capture data patterns with greater volatility than parametric methods, one of which is the SVR method. There is an expectation that next research endeavors would prioritize the examination of intermittent impacts, namely the occurrence of empty records, in order to enhance the precision of rainfall forecasts.

REFERENCES

Abdullah, A., Ruchjana, B., Jaya, I., et al. (2021). Comparison of sarima and svm model for rainfall forecasting in bogor city, indonesia. In *Journal of Physics: Conference Series*, volume 1722, page 012061. IOP Publishing. https://doi.org/10.1088/1742-6596/1722/1/012061.

Amelia, R., Dalimunthe, D., Kustiawan, E., and Sulistiana, I. (2021). Arimax model for rainfall forecasting in pangkalpinang, indonesia. In *IOP Conference Series: Earth and Environmental Science*, volume 926, page 012034. IOP Publishing. https://doi.org/10.1088/1755-1315/926/1/012034.

Dabral, P. and Murry, M. Z. (2017). Modelling and forecasting of rainfall time series using sarima. *Environmental Processes*, 4(2):399–419. https://doi.org/10.1007/s40710-017-0226-y.

Dani, A. T. R., Wahyuningsih, S., Putra, F. B., Fauziyah, M., Wigantono, S., Sandariria, H., A'yun, Q. Q., and Zen, M. A. (2023). Aplikasi model arimax dengan efek variasi kalender untuk peramalan trend pencarian kata kunci zalora pada data google trends. *Inferensi*, 6(2):107–116. https://doi.org/10.12962/j27213862.v6i2.15793.

Gautam, A. and Singh, V. (2020). Parametric versus non-parametric time series forecasting methods: A review. *Journal of Engineering Science & Technology Review*, 13(3). https://doi.org/10.25103/jestr.133.18.

Makridakis, S., Spiliotis, E., and Assimakopoulos, V. (2020). The m4 competition: 100,000 time series and 61 forecasting methods. *International Journal of Forecasting*, 36(1):54–74. https://doi.org/10.1016/j.ijforecast.2019.04.014.

Nanlohy, Y. W. A., Brodjol Sutijo, S., et al. (2019). Model fungsi transfer multi input untuk peramalan curah hujan di kota surabaya. *VARIANCE: Journal of Statistics and Its Applications*, 1(2):82–92. https://doi.org/10.30598/variancevol1iss2page82-92.

Nanlohy, Y. W. A. and Haumahu, G. (2021). Model autoregressive integrated moving average (arima) untuk peramalan curah hujan di kota surabaya. *Tensor: Pure and Applied Mathematics Journal*, 2(1):25–32. https://doi.org/10.30598/tensorvol2iss1pp25-32.

Nisa, C., Sumarjaya, I. W., and Srinadi, I. (2021). Penggunaan model arimax untuk meramalkan data curah hujan bulanan di bali. *EJ. Mat*, 10(4). https://doi.org/10.24843/MTK.2021.v10.i04.p341.

Ray, S., Das, S. S., Mishra, P., and Al Khatib, A. M. G. (2021). Time series sarima modelling and forecasting of monthly rainfall and temperature in the south asian countries. *Earth Systems and Environment*, 5:531–546. https://doi.org/10.1007/s41748-021-00205-w.

Riyani, D., Prastyo, D. D., and Suhartono, S. (2019). Input selection in support vector regression for univariate time series forecasting. In *AIP Conference Proceedings*, volume 2194. AIP Publishing. https://doi.org/10.1063/1.5139837.

Siswanti, T. E. and Yanti, T. S. (2020). Pemodelan arimax (autoregressive integrated moving average with exogenous variable). *Prosiding Statistika*, 6(2):113–118. https://dx.doi.org/10.29313/.v6i2.

Suci, K. W. and Irhamah, I. (2017). Peramalan curah hujan sebagai pendukung kalender tanam padi di pos kedungadem bojonegoro menggunakan arima, support vector regression dan genetic algorithm-svr. *Jurnal Sains dan Seni ITS*, 6(1):D157–D163. https://doi.org/10.12962/j23373520.v6i1.22092.

Suhermi, N., Suhartono, Permata, R. P., and Rahayu, S. P. (2019). Forecasting the search trend of muslim clothing in indonesia on google trends data using arimax and neural network. In *Soft Computing in Data Science: 5th International Conference, SCDS 2019, Iizuka, Japan, August 28–29, 2019, Proceedings 5*, pages 272–286. Springer. https://doi.org/10.1007/978-981-15-0399-3_22.