E-ISSN : 2407 – 3911
P-ISSN : 2686 - 0333

UTILIZING THE SARIMA MODEL AND SUPPORT VECTOR REGRESSION TO FORECAST MONTHLY RAINFALL IN BANDUNG CITY

Astri Nur Innayah[1], Dwi Intan Sulistiana[2], M. Yandre Febrian[3], Fitri Kartiasih[4*]
[1,2,3] Statistical Computing Study Program, Politeknik Statistika STIS
[4] Statistics Study Program, Politeknik Statistika STIS
Jalan Otto Iskandardinata No.64C Jakarta 13330, Indonesia
E-mail: 222111928@stis.ac.id, 222111998@stis.ac.id, 222112167@stis.ac.id, fkartiasih@stis.ac.id

Abstract

As one of the largest cities in Indonesia, Bandung has varying monthly rainfall intensity. High rainfall is very dangerous for people's lives and will have an impact on various sectors such as agriculture, fisheries, tourism, and transportation. For this reason, rainfall prediction is needed as an effort for the government to make policies and the community can anticipate the possibility of high rainfall that occurs. This study compares the effectiveness of SARIMA and Support Vector Regression (SVR) models in predicting monthly rainfall objectively, with the aim of improving decision making for stakeholders. Forecasting rainfall data is carried out based on the best method of the two methods that have been compared. The results showed that the SARIMA method outperformed the SVR method in forecasting precision, as seen from the lower RMSE value of 93.2045. The results provide valuable insights into weather prediction methodologies, benefiting authorities and the public.

Keywords : rainfall, SARIMA, SVR, forecasting, climate change

I. INTRODUCTION

Bandung city, is one of the five largest cities in Indonesia. Having an area of 167.31 km^2 and a population density of 15,051 people/km2 makes this city one of the cities that is almost known throughout Indonesia (Fajri et al., 2023). Bandung is located at an altitude of 700 meters above sea level and has very varied weather. The weather in question can be rainy, hot, and even stormy. According to the World Climate Conference, weather is the state of the atmosphere measured from various aspects such as the change or development of phenomena in the air (Firdaus & Paputungan, 2022). The elements that make up the weather include rainfall, air pressure, air humidity, temperature, clouds, and wind (Firdaus & Paputungan, 2022).

One of the most dominating elements of weather formation is rainfall. Rainfall is the amount of rainwater that falls on an area in a certain period of time and is expressed in millimeters (mm) (Hakiqi et al., 2023). Throughout 2022 the city of Bandung was hit by rain with different intensities each month. October was recorded as the month with the highest rainfall reaching 366.7 mm, while August was recorded as the month with the lowest rainfall reaching 29.9 mm (BPS, 2023).

The Meteorology, Climatology, and Geophysics Agency (BMKG) as the authorized agency for weather in Indonesia divides monthly rainfall into three categories, namely low with an intensity of 0 to 100 mm, medium with an intensity of 100 to 300 mm, and high with an intensity of 300 mm - 500 mm, and very high if the intensity exceeds 500 mm. Slight changes in rainfall can cause floods, droughts, changes in food production, and other economic activities (Xiang et al., 2018). High rainfall or heavy rainfall is very dangerous for people's lives because it will have adverse effects such as floods and landslides. Floods can cause economic losses and overflow of water, which is the main factor transporting sediments, nitrates, phosphorus and other chemical compounds in watersheds, can also be harmful (Ojo & Ogunjo, 2022). Low rainfall can cause drought and will impact several water-related sectors such as agriculture, irrigation and energy (Ginting & Kartiasih, 2019; Kartiasih et al., 2022; Maulana & Kartiasih, 2017; Rachma Safitri & Kartiasih, 2019; Ojo & Ogunjo, 2022).

In addition to the sectors above, it turns out that rainfall can also have an impact on several sectors such as agriculture, fisheries, transportation, and tourism (Fransiska et al., 2019). As in agriculture, which will have an impact on crop failure due to drought or flooding (Kartiasih & Setiawan, 2019; Latifa et al., 2023; Yuliana et al., 2023; Latif et al., 2023). In the fisheries sector, which will affect the results that will be obtained by fishermen. In the tourism sector, which will affect the number of tourists who will visit (Oktaviani & Afdal, 2013), and the transportation sector, which has an impact on the intensity of weather-related accidents and delays (Latif et al., 2023).

In addition to the impacts described above, government policies such as land conversion can also indirectly exacerbate the above conditions. Overall, the Bandung City government's land conversion efforts to improve urban planning have created significant challenges related to environmental sustainability. Although the mission to expand the city's facilities and infrastructure while paying attention to spatial planning and the carrying capacity of the urban environment is recognized as a positive step, the rapid rate of population growth is a major obstacle (Prihatin, 2015). The land conversion causes the increase of Green Open Space (RTH) area to be difficult to realize. According to Law No. 26 of 2007, the minimum ideal area of urban green space is 30%, while according to BPS data in 2020 the green space in Bandung City was only 12.25%, far from the minimum green space (Prihatin, 2015). The limited RTH makes the area that should be a water catchment area when rainfall increases cannot function properly.

Therefore, rainfall prediction is needed to overcome various problems above such as farmers in determining the planting season schedule and the government in warning of disasters and risk management regarding rainfall (R. He et al., 2024). In addition, rainfall prediction is also expected to provide benefits to related parties or the community itself in taking steps such as mitigating potential floods, droughts, or other hazards due to rainfall (R. He et al., 2024). Predicting rainfall is a difficult task due to the changing nature of climate phenomena and stochastic fluctuations associated with physical processes (Mehdizadeh et al., 2018). In predicting rainfall, there are several methods including subjective methods in which the process is based on the judgment or judgment of forecasters and objective methods in which the process uses statistical procedures (Gustari et al., 2012). In addition, in predicting rainfall, data processing is very important to get good results (Fauzy et al., 2016). BMKG as the authorized agency in predicting weather in Indonesia has certainly made predictions about rainfall. However, the predictions made still use subjective methods. Meanwhile, this research will use objective methods. There are many objective methods that have been found including Autoregressive Integrated Moving Average (ARIMA) and Support Vector Regression (SVR).

The ARIMA model is a model in time series analysis (Kafara et al., 2017). This method was first introduced by Box- Jenkins in 1970 (Kafara et al., 2017). ARIMA models can produce good accuracy in predicting future data using data from the past. The ARIMA model which is part of the classic time series model will work very well on time series data that has seasonal elements and clear trends, such as monthly and annual data (Ponnoprat, 2021). In ARIMA, there are several conditions that must be met such as normal distribution, stationary in average and variance, and white noise (Simanjuntak, 2023). In time series data that has seasonal patterns in it, this ARIMA model can be developed into a Seasonal Autoregressive Integrated Moving Average (SARIMA) model (Simanjuntak, 2023).

SVR is the development of Support Vector Machine (SVM) by incorporating regression elements so that it can be implemented to predict future data (Muhamad et al., 2017). SVR was discovered by Vapnik in 1995 which is usually implemented in regression cases. SVR bases its concept on risk reduction or function estimation by reducing the upper limit of the generalization error, so that the ability of SVR to handle overfitting can be realized (Siregar, 2022). The advantage of this SVR method is that there is no need for basic assumptions and the results will be good even though the data used is small (Suci & Irhamah, 2017).

Therefore, rainfall prediction is needed to know how the weather will be in the future. In this study, a comparison will be made between modeling using SARIMA and SVR in predicting monthly rainfall in the city of Bandung. Furthermore, predictions will be made using the best model between the two models. The results of this research are expected to provide benefits to related parties or the community itself in taking steps such as mitigating potential floods, droughts, or other hazards due to rainfall.

Numerous academics have undertaken studies on rainfall prediction in Bandung City. Previously, Hakiqi et al. (2023) have conducted a study related to rainfall projections in Bandung City using the SARIMA method. The results of their research present an optimal model, namely SARIMA (0,0,0)(0,1,1)[12]. In the Bandung Regency area, Soekendro (2021) also investigated aspects of rainfall prediction. The focus of this research is to evaluate the accuracy of the SARIMA method in forecasting rainfall, and the best finding is the SARIMA (2,1,1)(1,1,0)[12].

Ray et al. (2021) also conducted rainfall forecasting using the SARIMA approach. This study tries to examine the behavior of monthly average rainfall and temperature in South Asian countries using the SARIMA model. The best model selection is done using the Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), Root Mean Square Error (RMSE), and Mean Absolute Error (MAE) values. The best model is the one with the lowest values for AIC, BIC, RMSE, and MAE. From the results of this study, it can be concluded that climate change occurs in both rainfall and temperature in South Asian countries. Dabral & Murry (2017) carried out research on SARIMA in Northeast India as well. The modeling of monthly, weekly, and daily rainfall data is the main goal of this study. The Autocorrelation Function (ACF), the Partial Autocorrelation Function (PACF), and the lowest values of AIC and SBC are used to determine which model is the best fit. The SARIMA method approach shows that there is an improvement in terms of integration of trend removal, periodicity, and stochastic components of the time series modeling approach.

In addition to using the SARIMA method which is widely used by researchers in predicting rainfall, methods that utilize machine learning are also widely used, one of which is Support Vector Regression or SVR. Muhamad et al. (2017) compared the SVR method with SVR optimized with IPSO to find out which method is better at predicting rainfall. The result obtained is that the SVR method that has been optimized using IPSO is better. Research using SVR was also done by Hasan et al. (2015), which showed that the proposed technique is better than the conventional framework in terms of accuracy and processing time, where it managed to achieve a maximum prediction of 99.92%.

Rainfall prediction is complex due to complicated and unpredictable factors. However, in recent times, SVR and Recurrent SVR methods have been successfully used to overcome the challenges of time series prediction in various fields. However, the use of RSVR in rainfall forecasting has not been widely explored. Pai & Hong (2007) conducted a study with the aim of improving rainfall prediction accuracy by utilizing the advantages of SVR, genetic algorithms, and RSVR models.

Aghelpour et al. (2019) conducted a comparison between the SARIMA, SVR, and SVR-FA models. The study used monthly temperature data of five stations from various climates in Iran from 1951-2011, with a proportion of 75% training data and 25% testing data. Despite using the SVR model and its meta-innovative type, SVR-FA, the results showed that the SARIMA model still produced better performance in forecasting long-term temperatures.

From a number of related literatures, this research is important because it adds unmet knowledge to the literature related to rainfall prediction. Previously, many studies only focused on the application of one method, such as Support Vector Regression (SVR) or Seasonal Autoregressive Integrated Moving Average (ARIMA), without a direct comparison between the two. Therefore, this study aims to complement the existing knowledge by conducting a direct comparison between SVR and ARIMA in the context of rainfall prediction. Thus, it is hoped that the results of this research can provide clearer insights into the effectiveness of both as well as enrich the understanding of the application of prediction methods in this field.

Rainfall

Precipitation measurement, expressed in millimeters, denotes the level of rainwater accumulated in a rain gauge situated on a non-absorbent, non-seeping, and non-flowing flat surface (III, 2022). When one millimeter of rain falls on a one-square-meter flat surface, it signifies the collection of one liter of rainwater.

Rainfall intensity pertains to the quantity of rainfall within a specific time frame. According to BMKG's classification, rainfall is categorized into three levels: low (0 - 100 mm), moderate (100 - 300 mm), high (300 - 500 mm), and very high (>500 mm).

Time Series Data

Time series data are the values of individual variables that are organized sequentially according to the order of time. Time series data includes daily, weekly, monthly, quarterly, and annual data. The purpose of periodic series analysis is to understand and model the stochastic processes that occur in a series of observations and predict future values based on the information contained in the time series (Hakiqi et al., 2023). Time series data patterns can be trend, cyclical, seasonal, and irregular.

Data Stationarity

Time series data is considered stationary when its mean and variance values do not change structurally over a certain time span. The stationarity of time series data can be divided into two, namely stationarity in average and stationarity in variance. To determine the stationarity of a time series data can be through Graphical Analysis, Autocorrelation Function & Correlogram, and Unit Root Test.

Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF)

The Autocorrelation Function (ACF) assesses the averaging of data points in a time series concerning preceding data points. ACF serves the purpose of detecting the existence of Moving Average (MA) traits within a time series, and in the ARIMA model, it corresponds to the q parameter. When there are indications of MA characteristics, the typical range for the q value is generally 1 or 2 (Soekendro, n.d.).

$$\gamma_k = \mathrm{corr}(Y_t, Y_{t-k}) = \frac{\sum_{t=1}^{n-1}(Y_t - \bar{Y})}{\sum_{t=1}^{n-1}(Y_t - \bar{Y})^2} \quad \ldots [1]$$

The Partial Auto Correlation Function (PACF) is employed to assess the correlation between Yt and Yt-k after eliminating the influences of other time lags (Makridakis et al., 1984). Broadly, PACF serves the purpose of detecting Auto Regressive (AR) traits within a time series, typically denoted by the parameter p. If AR characteristics are present, the value of p is usually 1 or 2.

Correlogram

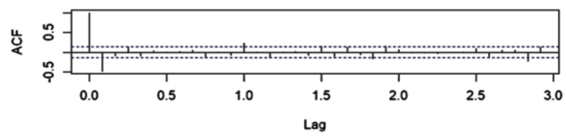


Figure 1. ACF Plot.

A correlogram employs the autocorrelation function (ACF) to detect stationarity in time series data. This function elucidates stochastic processes by examining the correlation between successive data points (Yt). Stationarity in a dataset is indicated when the correlogram exhibits a notable decrease as the value of k, representing lag, increases. Conversely, if the correlogram fails to approach zero (does not decrease) with the rising value of k, it suggests that the data lacks stationarity.

Unit Root Test

One of the established approaches for assessing the stationarity of data involves employing the unit root test. The Augmented Dickey-Fuller (ADF) Test, devised by David Dickey and Wayne Fuller, has gained widespread recognition in this context. When a time series lacks stationarity at the zero-order, I(0), the examination of data stationarity can be conducted at subsequent orders, such as the first difference or I(1), second difference or I(2), and so forth.

Here are some models to perform the ADF test including:

$\Delta Y_t = \delta Y_{t-1} + u_t$ (without intercept)
$\Delta Y_t = \beta + \delta Y_{t-1} + u_t$ (with intercept)
$\Delta Y_t = \beta_1 + \beta_{2t} + \delta Y_{t-1} + u_t$ (intercept with time trend)
$\Delta =$ first difference of the variable used
$t =$ trend variable

The proposed hypothesis is as follows:
$H_0: \delta = 0$ (here is a unit root, not stationary)
$H_1: \delta \neq 0$ (no unit root, stationary)

Data will be stationary if p-value < alpha.

Differencing

Differencing is employed when dealing with data that exhibits instability around the mean. The presence of such instability in the average can be detected using the Augmented Dickey-Fuller (ADF) test. If the test indicates instability, the subsequent

… [2]

action involves applying differencing through the following formula:

$$\Delta Y_t = Y_t - Y_{t-1}$$

Where:
Y_t : data at time t
Y_{t-1} : data at time (t-1)
ΔY_t : first level differencing

If the first level differencing has not made the data more stable, the next step is to find the difference at the second level. This process is iterative and aims to achieve stability in the data.

Transformation

Transformation is a technique employed on data lacking consistent variance properties. In the realm of data analysis, the Box-Cox transformation is frequently utilized for this purpose.

$$T(Y_t) = \begin{cases} \frac{Y_t^{\lambda}-1}{\lambda}, for\ \lambda \neq 0 \\ \lim_{\lambda \to 0} \frac{Y_t^{\lambda}-1}{\lambda} = \ln(Y_t), for\ \lambda = 0 \end{cases} \quad \ldots [3]$$

ARIMA (Autoregressive Integrated Moving Average) Model

In this model, the Box-Jenkins technique is applied. This technique includes three types of models, namely Autoregressive (AR), Moving Average (MA), Autoregressive Moving Average (ARMA), and Autoregressive Integrated Moving Average (ARIMA) combined models. The ARIMA model considers current and lagged observations (AR), current and previous random disturbances (MA), and the level of integration (I) (K. He et al., 2021). As an extension of ARIMA, the Seasonal ARIMA (SARIMA) model considers seasonal adjustment (S) in addition to ARIMA. To use ARIMA in modeling, it is assumed that the data must be stationary in mean and variance. If it is not stationary, a differencing process will be performed. It should be emphasized that, since AR, MA, and ARMA models do not have the ability to explain the meaning of differencing, the differencing process can only be applied to ARIMA models with the following model formulation.

AR (p) model :

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \cdots + \emptyset_p Y_{t-p} + e_t \quad \ldots [4]$$

MA (q) model :

$$Y_t = \delta + e_t - \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} \quad \ldots [5]$$

ARMA (p,q) model :

$$Y_t = \delta + \emptyset_1 Y_{t-1} + \cdots + \emptyset_p Y_{t-p} + e_t - \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} \quad \ldots [6]$$

ARIMA model :

$$Y_t = \delta + (1+\emptyset_1) Y_{t-1} + \cdots + (\emptyset_p - \emptyset_{p-1}) Y_{t-p} - \emptyset_p Y_{t-p-1} + e_t - \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} \quad \ldots [7]$$

SARIMA Model

An alternative form of the ARIMA model is the seasonal ARIMA, abbreviated as SARIMA. SARIMA is applied to data exhibiting both a seasonal pattern and the ARIMA effect (K. He et al., 2021). SARIMA can be represented as SARIMA $(p,d,q)(P,D,Q)^m$, where the formulation is:

$$\emptyset_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^S)\alpha_t \quad \ldots [8]$$

Model Diagnostic Test

Despite the fact that the chosen model could appear to be the best of those under consideration, diagnostic testing must be done to ensure that it is adequate. To find any patterns that are not explained, this entails carefully examining the residuals. Although it is more difficult to calculate residuals in an ARIMA model than in a regression model, residuals are nonetheless produced as part of the estimate process for ARIMA models.

The residuals that remain after modeling should have properties similar to white noise in order to produce a dependable forecasting model. As a result, the ACF and PACF of the residuals are expected to be devoid of both major and partial autocorrelations. The Ljung-Box test is a useful tool for validating the white noise assumption.

Hypothesis:
$H_0: \rho_1 = \rho_2 = \cdots = \rho_k = 0$
$H_1: at\ least\ one\ \rho_i\ is\ not\ equal\ to\ zero, i = 1,2,\ldots,k$

Test Statistics:

$$Q = n(n+2)\sum_{k=1}^{k} \frac{\widehat{\rho_k^2}}{(n-k)}, n > k \quad \ldots [9]$$

Rejection region: Reject H0 if $Q > x^2(\alpha; K-p-q)$ or p-value<alpha.

Terasvirta Nonlinearity Test

To determine whether the data exhibits a linear or nonlinear pattern, an assessment is carried out through nonlinearity testing. Various methods, including the Terasvirta test, Ramsey-Reset test, and White Test, have been developed for pattern identification (Purwoko et al., 2023). It is acknowledged that the Terasvirta test outperforms other methods in detecting nonlinearity. The Terasvirta test falls within the category of the Lagrange Multiplier (LM) test, extended through Taylor expansion. The formulation of the nonlinear neural network model is as follows:

$$Z_t = \varphi(\gamma' w_t) + \beta' w_t + \varepsilon_t \qquad \dots [10]$$

with β $'\boldsymbol{w}$, $\varphi(\gamma$ $'w)$, γ $'$, β $'$ are the linear component, nonlinear component, weights in the neural network model from the input layer to the hidden layer for the nonlinear component, and weights in the neural network model from the input layer to the output layer for the linear component, while φ is a sigmoid activation function. The hypothesis for the Terasvirta test can be formulated as follows:

H_0: $model\ linier$

H_1: $model\ nonlinier$

Support Vector Regression (SVR)

Support Vector Regression (SVR) constitutes a segment of the Support Vector Machine (SVM) framework, pioneered by Vapnik in 1995. SVM operates as a learning system employing a hypothesis space represented by a linear function within a high-dimensional feature space. The SVM training procedure encompasses learning algorithms rooted in optimization theory, incorporating biased learning. The fundamental concept of SVM employs a loss function known as the ε-insensitive loss function. Regression analysis and classification are the main applications of SVM, a machine learning technique based on statistical learning theory (Guo et al., 2017). Put simply, SVM's core principle involves elevating dimensionality and achieving linearization. SVM can be extended or generalized to encompass a functional approach termed Support Vector Regression (SVR). The key advantage of SVR lies in its ability to mitigate overfitting, thereby ensuring optimal performance.

The secret to successful SVR is choosing the right kernel function. Forecasting models with varying kernel functions have varying learning and generalization capacities (Lu et al., 2019). Tan et al. (2020) have identified some kernel function options that are frequently utilized in SVR techniques.

Linear kernel functions can quickly solve linear problems with relatively few parameters (Tan et al., 2020). The expression of the product in the linear kernel function is as follows:

$$K(x_1, x_2) = < \mathrm{x}_1, x_2 > \qquad \dots [11]$$

The RBF (Radial Basis Function) kernel function is more stable in processing non-linear problems (Jiang et al., 2017) and provides an ideal fitting effect after selecting accurate parameters. The expression of the product in the RBF kernel function is as follows:

$$K(x_1, x_2) = \exp\left(-\frac{||x_1 - x_2||^2}{2\sigma^2}\right) \qquad \dots [12]$$

Polynomial (Poly) kernel functions solve non-linear problems and are suitable especially for orthogonally normalized data. The expression of the product in the Polynomial kernel function is as follows:

$$K(x_1, x_2) = (\alpha \mathrm{x}_1^T x_2 + c)^d \qquad \dots [13]$$

The sigmoid kernel function handles non-linear problems. In this case, SVR is a type of multilayer perceptron neural network. In convex quadratic optimization problems, it prevents falling into a local optimum. The expression of the product in the sigmoid kernel function is as follows:

$$K(x_1, x_2) = \tanh(\alpha \mathrm{x}^{\mathrm{T}} y + c) \qquad \dots [14]$$

In the context of SVR, there are three independent parameters C, ε, and γ. Configuring these parameter values is pivotal in attaining optimal outcomes, specifically in obtaining a model with minimal error rates. Hence, the meticulous selection of an appropriate range for these parameters is essential to achieve optimal results.

Best Model Selection

The identification of the optimal model is achieved through the evaluation of the Root Mean Square (RMSE) value. RMSE serves as a metric for gauging the disparity between a model's predicted values and the actual values. A lower RMSE suggests that the model's output closely aligns with the measured values' variation (Makridakis et al., 1984). The formulation of RMSE is expressed as follows.

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}\left(Y_t - \hat{Y}_t\right)^2} = \sqrt{MSE} \qquad \ldots [15]$$

Where:
n = amount of data
Y_t = Actual data at time t
$\hat{Y}_t$ = Predicted data at time t

Forecasting

Forecasting is a process that aims to estimate future conditions based on information that has been recorded in the past. The forecasting process is needed to estimate various aspects such as quantity, quality, time, and location of needs that will arise in the future (Muhamad et al., 2017). As a tool in efficient and effective planning, forecasting can also be interpreted as a method that supports the planning process (Soekendro, n.d.). However, the results of predictions do not always match reality, such as in forecasting the level of product demand or forecasting the price of basic necessities in a certain period in the future (Nuraini Anitasari, 2022).

Stochastic and deterministic methods are two popular methodologies used in forecasting (Soekendro, n.d.). A stochastic model, characterized by a quantifiable level of certainty in its mathematical structure, tends to exhibit instability. This model allows for the calculation of the probability of each event and is frequently applied in contexts such as queuing theory and game theory. Conversely, a deterministic model is a form of simulation that disregards randomness, yielding a predictable outcome when specific inputs and relationships are applied. Despite the absence of randomness in deterministic models, the consistency of results remains intact even with repeated processing of inputs (Soekendro, n.d.).

II. DATA AND METHODOLOGY

This research uses monthly rainfall data in Bandung City. The data period is from January 2008 to October 2023, with a total of 190 data points. The data used includes secondary data obtained from two sources, namely data from the Central Statistics Agency (BPS), which can be downloaded at https://bandungkota.bps.go.id/, and data from Ogimet, which can be accessed using the R package "climate." Monthly rainfall data for the period January 2008 to December 2022 was obtained from the Central Bureau of Statistics (BPS), while data from January 2023 to October 2023 was obtained from Ogimet. Data from the Central Bureau of Statistics and Ogimet have the same data source, which comes from the observation station of the Meteorology, Climatology, and Geophysics Agency (BMKG) in Bandung City. The data collection flow is shown in Figure 1.

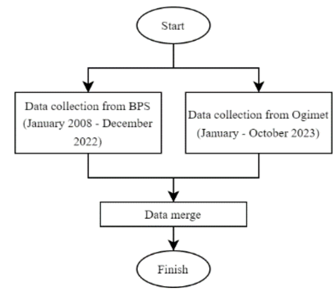


Figure 2. Data collection flow.

In the descriptive analysis of rainfall data, there are important steps involved. First, a descriptive statistical analysis is conducted, which involves calculating the rainfall mean, maximum value, minimum value, variance, and standard error. This information provides an overview of the distribution of the data and its general characteristics. In addition, this analysis helps in understanding the level of rainfall variation in a region.

Furthermore, in addition to the descriptive statistical analysis, a visual presentation was also conducted using plots of monthly rainfall data in Bandung City. This plot provides a graphical representation of the fluctuations in rainfall over a certain period. By looking at the graph, seasonal patterns and trends in the rainfall data can be identified.

Analysis with SARIMA Method

SARIMA (Seasonal Autoregressive Integrated Moving Average) is an ARIMA model developed to account for seasonal patterns (Abdullah et al., 2021). ARIMA is the most popular classical time series model, both for stationary and non-stationary data (Abdullah et al., 2021). For SARIMA, the general multiplicative model is expressed by SARIMA (p, d, q)(P, D, Q)[s], where p, d, and q are AR terms, differentiation, and MA terms, respectively, and P, D, and Q are seasonal AR terms, seasonal differentiation, and seasonal MA terms, respectively, and s is the seasonal period (Ray et al., 2021). The SARIMA method is easy to use for predicting time series data that has a seasonal pattern (Prianda & Widodo, 2021). The steps to perform rainfall forecasting using this method include: 1) Make a decomposition plot to determine the presence of seasonal patterns in the data. 2) Identify data stationarity. This is seen from the average and variance of the data using ACF examination and ADF testing. If it is not stationary, differencing is carried out to fulfill the assumption of data stationarity. 3) Identify the ARIMA model seen from the ACF and PACF plots and use the "auto.arima" function in R, which can automatically select the best ARIMA or SARIMA model. 4) Estimate parameters and perform parameter significance tests on the best model. 5) Perform diagnostic tests for model evaluation. This diagnostic test is to ensure that the residuals (errors) of forecasting are random. In this diagnostic test, the Ljung-Box statistics indicator test is used. If the p-value obtained is greater than the specified significance level, it can be shown that the residuals are random or that the model obtained is adequate.

Analysis with SVR Method

An SVM model called SVR (Support Vector Regression) was created specifically for use in regression scenarios. SVR bases its approach on the concept of risk minimization, where the goal is to estimate a function by finding the minimum value of an upper bound on the common error (S.R. Gunn, 1998). Thus, SVR has the ability to overcome overfitting (S.R. Gunn, 1998). The SVR algorithm is also suitable for data that has random values and is able to overcome the case of non-linearly modeled data (Suci & Irhamah, 2017). SVR is an efficient learning system based on constrained optimization theory that utilizes the inductive principle of minimum structural risk, leading to an overall optimal response. On the other hand, SVM is a group of related supervised learning methods used for classification and regression analysis purposes (Aghelpour et al., 2019). The SVR method can recognize time-series data patterns and provide good forecasting results (Puspita et al., 2022).

The steps to perform rainfall forecasting using this method are as follows: 1) Test nonlinearity using the Terasvirta test to see nonlinearity in the data and determine the kernel function used. 2) Divide the data into training data and testing data. At this stage, the 80:20 data partition is carried out, where 80% is training data and 20% is testing data. 3) Determining the range of parameter values C, ε, and γ for optimization in training data. 4) Perform modeling using the SVR method based on the best parameter values.

Selection of the Best Approach

Comparing the SARIMA and SVR algorithms is the first step in selecting the optimal approach. The RMSE value for each approach is taken into consideration during this comparison, and the model with the lowest RMSE value is selected as the best (Pemayun et al., 2024; Subian et al., 2024). After getting the best model, the next step is to use the model to perform monthly rainfall forecasting.

Forecasting Rainfall Data

After comparing the best method, a forecast or prediction of monthly rainfall intensity data is carried out. Data prediction is carried out for the next 14 months, from November 2023 to December 2024. This forecasting is done based on the model of the best method obtained.

III. RESULT AND DISCUSSION

Rainfall data in Bandung City from January 2008 to October 2023 has an average of 195.22 mm/month. According to the Meteorology, Climatology and Geophysics Agency, the rainfall is categorized as moderate rainfall. The maximum rainfall is above the average of 637 mm/month, while the minimum rainfall shows no rainy days in a month. This rainfall data has a high variance of 17926.95 with a standard error of 133.8916.

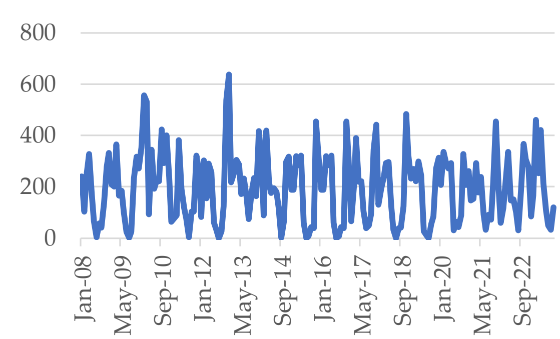


Figure 3. Plot of rainfall in Bandung City.

Source: BPS and Ogimet (2023)

Figure 2 shows that the data has a seasonal pattern. This can be seen from the pattern that repeats at certain periods, where an increase in rainfall occurs at the end of the year, especially in November and December.

III.1 SARIMA Model Analysis

Previously, we have explored descriptive statistics and time series plots that provide an overview of the patterns and variations in rainfall data. Now, our focus turns to analyzing the SARIMA model to understand more about this data. The decomposition plot (Figure 2) is the first step to dig deeper into the components of the data. Further analysis at the stationarity test stage will validate the stationarity of the data, an important step before we enter the SARIMA modeling stage.

Decomposition Plot

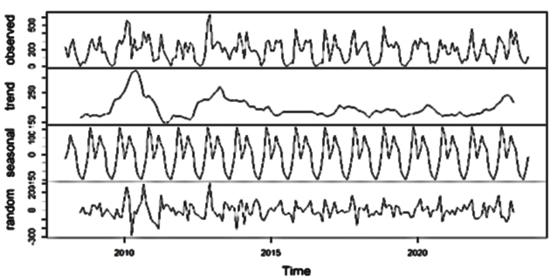


Figure 4. Decomposition plot of rainfall data in Bandung City.

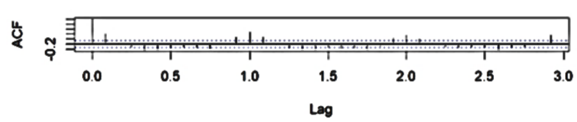
Source: BPS and Ogimet, processed (2023)

Judging from Figure 3, the rainfall data of Bandung City from January 2008 to October 2023 does not show a significant trend because the trend does not show a trend line that is almost straight up. So that in January 2010 to October 2023 rainfall in Bandung City there is no growth or increase phenomenon. In addition, there is an additive seasonal component that can change over time. This seasonal component is constant each year, as seen from the same seasonal pattern plot each year.

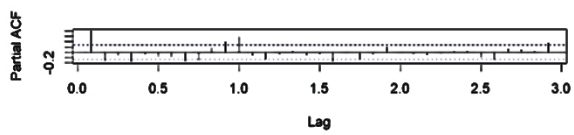
The residual component that is left over after trend, seasonal, and other distinguishable characteristics have been eliminated from the time series data is referred to as "random" in time series decomposition. This component is the part of the data that cannot be explained by other components in the decomposition and usually contains random fluctuations or noise. The rainfall data for Bandung City from January 2008 to October 2023 has a random residual component because the component does not show a clear pattern or trend, indicating that the fluctuations in the data are random and cannot be explained by trends or seasonality.

Stationarity Test

The importance of stationarity test in time series data analysis lies in the stage of determining the existence of unit root among variables, so that the validity of the relationship between variables in an equation can be ensured. Fulfillment of stationarity conditions is required to perform modeling using SARIMA. Stationarity can be seen in various ways, such as through the examination of the ACF plot and with the ADF Test to further ensure the stationarity of the data.



(a)



(b)

Figure 5. (a) ACF plot and (b) PACF plot of rainfall data in Bandung City.

Source: BPS and Ogimet, processed (2023)

From the ACF plot seen in Figure 4, it can be concluded that the data is stationary due to the correlogram pattern that approaches zero quickly. For a more convincing test, it can be done with the ADF Test with a 5% test level. The results of the ADF Test obtained a t-Statistic of -8.7481 with a p-value of 0.000.

By using the ADF Test, it is found that the resulting p-value is less than the specified significance level of 5%. So it can be concluded that the decision is to reject H0. So with a significance level of 5%, it is obtained that the rainfall data for Bandung City from January 2008 to October 2023 is stationary so there is no need for transformation and can be directly used for SARIMA modeling.

SARIMA Model Identification

Using the ACF and PACF plots of the stationarized data, the SARIMA model identification process is carried out to find the appropriate model. The SARIMA model can be determined based on the characteristics of the ACF and PACF plots. In the PACF plot, it can be seen that there is a drastic change (cut-off) in the early lags, precisely after the first lag, which indicates the presence of AR. For seasonal elements, it can be seen from the lag pattern with a multiple of 12 because the data used is monthly data with S = 12. In the ACF and PACF plots, the presence of a Seasonal Moving Average (SMA) component can be indicated by the presence of a significant lag-12. Therefore, the initial conjecture model generated is ARIMA (1,0,0)(0,0,1)[12]. In addition, by using auto.arima in R, several candidate models can be obtained for comparison with the previous estimated model. The candidate models are listed in table 1.

Table 1. Candidate models obtained with the R function.

Model Candidate	
ARIMA(1,0,0)(0,0,1)12	ARIMA(0,0,0)(1,0,0)12
ARIMA(1,0,0)(1,0,0)12	ARIMA(0,0,1)(1,0,0)12
ARIMA(0,0,1)(0,0,1)12	ARIMA (0,0,1)
ARIMA (1,0,0)	ARIMA(0,0,1)(2,0,0)12
ARIMA(1,0,0)(2,0,0)12	ARIMA(0,0,1)(1,0,1)12
ARIMA(1,0,0)(1,0,1)12	ARIMA(0,0,1)(2,0,1)12
ARIMA(1,0,0)(2,0,1)12	

Source: BPS and Ogimet, processed (2023)

Parameter Significance Test

After obtaining a provisional SARIMA model, parameter significance testing is carried out on the candidate model. Parameter significance is evaluated based on the p-value of the model, where the model is considered significant if the p-value is less than the alpha set at 0.05.

Table 2. Parameter significance of candidate models.

Model Candidate	Parameters	Estimation	P-value	Description
ARIMA(1,0,0)(0,0,1)12	AR(1)	0.3600	0.0000	Significant model
	SMA(1)	0.3559	0.0000	
ARIMA(1,0,0)(1,0,0)12	AR(1)	0.3242	0.0000	Significant model
	SAR(1)	0.4258	0.0000	
ARIMA(0,0,1)(0,0,1)12	MA(1)	0.4026	0.0000	Significant model
	SMA(1)	0.3508	0.0000	
ARIMA (1,0,0)	AR(1)	0.4168	0.0000	Significant model
ARIMA(1,0,0)(2,0,0)12	AR(1)	0.3012	0.0000	Significant model
	SAR(1)	0.3573	0.0000	
	SAR(2)	0.1658	0.0000	
ARIMA(1,0,0)(1,0,1)12	AR(1)	0.2356	0.0009	Significant model
	SAR(1)	0.9994	0.0000	
	SMA(1)	-0.9693	0.0000	
ARIMA(1,0,0)(0,0,1)12	AR(1)	0.3600	0.0000	Significant model
	SMA(1)	0.3559	0.0000	
ARIMA(1,0,0)(2,0,1)12	AR(1)	0.2343	0.0011	Model is not significant
	SAR(1)	1.014	0.0000	
	SAR(2)	-0.0143	0.8572	
	SMA(1)	-0.9807	0.0000	
ARIMA(0,0,0)(1,0,0)12	SAR(1)	0.4953	0.0000	Significant model
ARIMA(0,0,1)(1,0,0)12	MA(1)	0.3583	0.0000	Significant model
	SAR(1)	0.4147	0.0000	
ARIMA (0,0,1)	MA(1)	0.4626	0.0000	Significant model
ARIMA(0,0,1)(2,0,0)12	MA(1)	0.3249	0.0000	Significant model
	SAR(1)	0.3551	0.0000	
	SAR(2)	0.1547	0.0434	
ARIMA(0,0,1)(1,0,1)12	MA(1)	0.2463	0.0010	Significant model
	SAR(1)	0.9995	0.0000	
	SMA(1)	-0.9728	0.0000	

Source: BPS and Ogimet, processed (2023)

Table 2 displays the parameter significance results of the candidate models, which include parameter estimates, p-values, and related information. From the table, it can be seen that there are twelve models that show significance in the parameters tested. The twelve models' residuals were then examined to verify their forecasting validity and consistency based on the analysis's findings. In the context of time series data forecasting, this phase is crucial to ensuring that the models that have been chosen can produce accurate and trustworthy prediction outcomes.

Model Diagnostic Test

After obtaining a significant model, the residuals are tested. There is an assumption that must be met, namely white noise, or the assumption that the residues are identical and independent. Testing the white noise assumption can be done by visualizing the ACF and PACF plots of the residuals. Another option is feasibility testing using the Ljung Box p-value. If the p-value is greater than 0.05 then the white noise assumption is met.

Table 3. Ljung Box test of models with significant parameters.

Model Candidate	Lag	P-value	Description
ARIMA(1,0,0)(0,0,1)12	12	0.1566	No white noise
	24	0.0018	
	36	0.0000	
	48	0.0000	
ARIMA(1,0,0)(1,0,0)12	12	0.2935	No white noise
	24	0.1050	
	36	0.0131	
	48	0.0000	
ARIMA(0,0,1)(0,0,1)12	12	0.2286	No white noise
	24	0.0032	
	36	0.0000	
	48	0.0000	
ARIMA (1,0,0)	12	0.0000	No white noise
	24	0.0000	
	36	0.0000	
	48	0.0000	
ARIMA(1,0,0)(2,0,0)12	12	0.3817	No white noise
	24	0.1162	
	36	0.0316	
	48	0.0273	
ARIMA(1,0,0)(1,0,1)12	12	0.7258	White noise
	24	0.4925	
	36	0.4186	
	48	0.7699	
ARIMA(1,0,0)(0,0,1)12	12	0.1566	No white noise
	24	0.0018	
	36	0.0000	
	48	0.0000	
ARIMA(0,0,0)(1,0,0)12	12	0.0030	No white noise
	24	0.0010	
	36	0.0000	
	48	0.0000	
ARIMA(0,0,1)(1,0,0)12	12	0.3799	No white noise
	24	0.0934	
	36	0.0111	
	48	0.0034	
ARIMA (0,0,1)	12	0.0000	No white noise
	24	0.0000	
	36	0.0000	
	48	0.0000	
ARIMA(0,0,1)(2,0,0)12	12	0.4344	No white noise
	24	0.0970	
	36	0.0259	
	48	0.0165	
ARIMA(0,0,1)(1,0,1)12	12	0.7361	White noise
	24	0.4563	
	36	0.3859	
	48	0.7333	

Source: BPS and Ogimet, processed (2023)

Table 3 shows the results of diagnostic testing of significant models. Based on table 3, only two models are identified as white noise, which means that the residuals are random or uncorrelated.

Selection of the best SARIMA Model

The optimal model will be chosen once a substantial tentative estimation model has been obtained and the white noise assumption has been met. The goal of model selection is to identify the model that can predict outcomes better than all others. Based on the RMSE value, the optimal model is chosen in this study. The Root Mean Square Error (RMSE) metric is used to evaluate two ARIMA models. The first model, ARIMA (1,0,0)(1,0,1)[12], has an RMSE value of 93.3720. Meanwhile, the second model, ARIMA (0,0,1)(1,0,1)[12], shows an RMSE value of 93.2045. The model is said to be good if it has a small Root Mean Square. So between the 2 models, the best model for predicting rainfall data in Bandung City is the ARIMA (0,0,1)(1,0,1)[12] model with an RMSE value of 93.2045.

The following is the equation of the ARIMA (0,0,1)(1,0,1)[12] model :

$$(1-\Phi_1 B^{12})Y_t = \alpha + (1+\theta_1 B)(1+\Theta_1 B^{12})e_t$$

$$Y_t = \alpha + \Phi Y_{t-12} + e_t + \theta_1 e_{t-1} + \Theta_1 e_{t-12} + \theta_1 \Theta_1 e_{t-13}$$

$$Y_t = 195.7317 + 0.9995Y_{t-12} + e_t + 0.2463e_{t-1}$$
$$-0.9728e_{t-12} - 0.2396e_{t-13}$$

According to the model above, rainfall data from the previous 12 months, forecast error from the previous month, and rainfall data from the 12 and 13 months prior all have an impact on Bandung City's rainfall prediction for month t. This model takes into account significant variables in the time series data to offer a robust framework for rainfall forecasting in Bandung City.

III.2 SVR Model Analysis

After obtaining the best model through the analysis of the SARIMA method, the next step is to find out the best model using the SVR method. The goal is to compare the best method between the two methods. Thus, a more comprehensive understanding of the effectiveness of each method in forecasting can be obtained.

Nonlinearity Test

Before performing the Support Vector Model method, it is necessary to test the nonlinearity of rainfall data for Bandung City from January 2008 to October 2023. Then nonlinearity testing is carried out to determine the kernel function in the SVR method. In this test, the Terasvirta test will be used. If the p-value is less than 0.05, the data is considered nonlinear.

Table 4. Nonlinearity test using Terasvirta test.

Lag	t-Statistic	Prob.
1	10.5373	0.0052
2	38.6120	0.0000
3	54.8421	0.0000
4	72.85678	0.0000
5	113.0719	0.0000
10	1087.4360	0.0000

Source: BPS and Ogimet, processed (2023)

Table 4 shows the results of the Terasvirta test conducted on the data. The Terasvirta test is carried out at several lags, namely lags 1, 2, 3, 4, 5, and 10. The p-value is smaller than alpha (alpha = 5%) at each lag. Based on the results in Table 4, the rainfall data for Bandung City from January 2008 to October 2023 is nonlinear. So that the Radial Basis Function (RBF) kernel function will be used in forecasting.

Selection of the Best SVR Model

The applied SVR approach involves the use of a Radial Basis Function kernel function by utilizing three variables, namely C (cost), γ, and ε. These parameters are determined in order to obtain optimal results. Trial-and-error variations or combinations of several parameter interval values are used in parameter settings in this study.

To facilitate the selection of parameters, it is necessary to identify the optimal range of the ε parameter first. In trial and error to find the epsilon parameter range, the value for the C parameter is 0.01–10 and for γ is 0.5–1.5. The results of the trial of the range value of parameter ε are shown in the table below.

Table 5. Experimental range value of epsilon parameter when the range value of C parameter is 0.01-10.

Trial to	Parameter Range	RMSE training	RMSE testing
1	0.0001-0.001	137.5117	116.3676
2	0.001-0.01	137.5175	116.3690
3	0.01-0.1	137.5663	116.3930
4	0.1-0.5	137.5553	116.7369
5	0.5-1	138.0736	116.8523

Source: BPS and Ogimet, processed (2023)

Based on table 5, the optimum epsilon parameter interval is obtained when the values of C and γ are constant, namely 0.0001-0.001. The optimum range selection comes from the smaller RMSE testing value compared to other ranges. Furthermore, to determine the optimum γ parameter, the value of the ε parameter obtained from the previous experimental process is 0.0001-0.001, and the C parameter used is 0.01–10. The results obtained from the parameter interval value experiment are in the table below.

Table 6. Range parameter value experiment when the epsilon parameter range value is 0.0001-0.001.

Trial to	Parameter Range	RMSE training	RMSE testing
1	0.05-0.1	137.5285	116.3568
2	0.1-0.5	137.5287	116.3563
3	0.5-1	137.5117	116.3676
4	0.6-1.5	137.5182	116.3721
5	1-1.5	137.5369	116.3827

Source: BPS and Ogimet, processed (2023)

Based on table 6, the results of the parameter interval are optimum when the values of C and ε are constant, namely 0.1–0.5. The selection of the optimum range comes from the smaller RMSE testing value compared to other ranges. Furthermore, to determine the optimum parameter C parameter, the optimum parameter values ε and γ were obtained from the previous experimental process, namely 0.0001-0.001 and 0.1-0.5. The results obtained from the parameter interval value experiment C are in the table below.

Table 7. Experiment of the range parameter value of C when the range parameter value γ is 0.1-0.5.

Trial to	Parameter Range C	RMSE training	RMSE testing
1	0.01-10	137.5287	116.3563
2	1-10	137.9983	120.0731
3	5-50	137.9135	121.0695
4	10-100	137.8786	121.2511
5	100-150	137.4116	118.9790

Source: BPS and Ogimet, processed (2023)

Based on Table 7, the optimum C parameter range when the epsilon parameter value is in the interval 0.0001-0.001 and the gamma parameter value is in the interval 0.1-0.5 based on previous trial and error is 0.01–10. The selection of the most optimal range comes from the smaller RMSE value compared to other ranges. In the Bandung City rainfall data from January 2008 to October 2023, the C parameter value in the interval 0.01–10, the epsilon parameter value in the interval 0.0001-0.001, and the gamma parameter value in the interval 0.1–0.5 obtained an RMSE of 116.3563.

III.3 Selection of the Best Approach

According to the primary goals of this study, the performance of the two approaches will be compared based on the RMSE value once the optimal model using the SARIMA and SVR methods has been obtained from the rainfall intensity data in Bandung City from January 2008 to October 2023. The RMSE value is used to compare the two methods because the rainfall data used has a value of 0, which, if using MAPE, can cause division by zero, which can produce an infinite or undefined value. Because the RMSE value does not require dividing by the actual value, it is therefore more prudent to use it in this particular case.

From the calculation results, it is found that the SARIMA method provides an RMSE value of 93.2045, while the SVR method has an RMSE value of 116.3563. With a lower RMSE value, the SARIMA method clearly shows better performance in forecasting rainfall compared to the SVR method. Because RMSE gives an indication of the extent of the difference between predicted and actual values, it can be concluded that the best model for forecasting rainfall intensity data for Bandung City in that period is ARIMA (0,0,1)(1,0,1) [12].

III.4 Forecasting

After obtaining the best model, forecasting the rainfall intensity in Bandung City can be done. Table 8 illustrates the results of rainfall forecasting in Bandung City from November 2023 to December 2024, which is based on the best model ARIMA (0,0,1)(1,0,1) [12]. Evaluation of the increase or decrease in rainfall can be observed from month to month. In November 2023, there was a significant increase, with rainfall reaching 334.4 mm, a high category. Meanwhile, December 2023 showed a decrease to 298.3 mm in the medium category. The period from January to October 2024 reflects seasonal fluctuations with generally medium to low rainfall. However, in November 2024, there is a significant increase to 345.7 mm (high), then a return to the medium category in December 2024 with 298.3 mm. This dynamic reflects potential seasonal changes or the influence of other factors that can affect rainfall intensity.

Table 8. Rainfall forecasting from November 2023-December 2024.

Period	Rainfall	Category
November 2023	334.4	High
December 2023	298.3	Medium
January 2024	177.7	Medium
February 2024	213.1	Medium
March 2024	300.2	High
April 2024	256.9	Medium
May 2024	236.9	Medium
June 2024	114.9	Medium
July 2024	75.4	Low
August 2024	60.1	Low
September 2024	97.6	Low
October 2024	180.3	Medium
November 2024	345.7	High
December 2024	298.3	Medium

Source: BPS and Ogimet, processed (2023)

Table 8 shows the prediction results of rainfall intensity in Bandung City in the period November 2023 to December 2024. It can be seen that rainfall in Bandung will be high or low in certain months, according to the existing seasonal pattern. Rainfall in Bandung City is predicted to be high in November 2023, March 2024, and November 2024. From the results of this prediction, a number of relevant suggestions can be made for the government in making policies and for various affected sectors, such as agriculture, so that the presence of high rainfall intensity does not have a significant impact on various existing sectors.

As is known, the high rate of land conversion in Bandung City has caused the water catchment area to continue to decrease, while the intensity of rainfall in Bandung Regency is quite high. Therefore, there is a need for stricter policies by the local government. The government could consider tightening land use change permits in areas that have the potential to experience significant impacts due to flooding or landslides. In addition, the government should also improve the supervision and enforcement of regulations related to land conversion. At the same time, it is also important to involve relevant parties, including local communities and environmental experts, in decision-making related to land conversion. The active involvement of various stakeholders can create policies that are more effective and accepted by the entire community.

In addition, high rainfall intensity can also impact the agricultural sector in Bandung City. With the information on predicted high and medium rainfall in Bandung City, a number of suggestions can be made that are relevant for agriculture in the area. In high rainfall conditions, it is recommended to plant crops that have tolerance for excess water. During the medium rainfall season, choosing crops that are able to cope with variations in rain intensity can be an appropriate strategy, such as corn. In addition, providing education on the benefits of planting crops that are suitable for certain rainfall conditions can have an impact on the optimal results obtained by farmers. These efforts are expected to increase agricultural productivity and reduce the potential impacts of rainfall fluctuations in Bandung city.

Another policy that can be carried out by the Bandung City government is to strengthen and improve the early warning system for high rainfall potential based on the forecasting results above for November to December 2024. Early warning of high rainfall potential can be done through various official social media platforms owned by the Bandung City government, such as Instagram and other platforms. With this effort, it is hoped that notifications regarding predictions of the possibility of high-intensity rainfall can be known by the wider community so that people can anticipate these events and minimize the risks of these events.

IV. CONCLUSION

Based on rainfall modeling that has been done with SARIMA and SVR, it is found that the best SARIMA model in modeling rainfall in Bandung City from January 2008 to October 2023 is ARIMA (0,0,1)(1,0,1)[12] with an RMSE value of 93.2045 while modeling done with SVR using 3 parameters namely C of 0.01-10, γ of 0.1 - 0.5, and ε of 0.0001 - 0.001 resulted in an RMSE value of 116.3563. Based on the above results, it is found that the best model in modeling rainfall in Bandung City is ARIMA (0,0,1)(1,0,1)[12] with an RMSE value of 93.2045.

After predicting rainfall using the ARIMA (0,0,1)(1,0,10)[12] model, it was found that the highest rainfall intensity will occur in November 2024 with an intensity of 345.7 mm, while the lowest rainfall will occur in August 2024 with an intensity of 60.1 mm. In addition, if grouped by rainfall category according to BMKG, it is found that high rainfall intensity will occur in November 2023, March 2024, and November 2024, moderate rainfall intensity is predicted to occur in December 2023 to February 2024, April 2024 to June 2024, and in October and December 2024, while low rainfall intensity will occur in July 2024 to August 2024.

Predictions of rainfall in Bandung City are expected to be the basis for proposing policies to mitigate potential flooding. Since the rainfall pattern does not show a decrease, a strict policy on land conversion is needed to reduce the risk of flooding because if more land is converted, the potential for flooding will increase. In addition, it is recommended that agricultural policies are responsive to rainfall fluctuations, including education for farmers. The government also needs to strengthen early warning systems through official social media to increase public awareness and preparedness for heavy rainfall. Overall, a holistic and collaborative approach is needed to manage the impacts of rainfall in Bandung City so that the impacts can be minimized.

REFERENCE

Abdullah, A. S., Ruchjana, B. N., Jaya, I. G. N. M., & Soemartini. (2021). Comparison of SARIMA and SVM models for rainfall forecasting in Bogor city, Indonesia. Journal of Physics:

Conference Series, 1722(1), 012061. https://doi.org/10.1088/1742-6596/1722/1/012061

Aghelpour, P., Mohammadi, B., & Biazar, S. M. (2019). Long-term monthly average temperature forecasting in some climate types of Iran, using the models SARIMA, SVR, and SVR-FA. Theoretical and Applied Climatology, 138(3-4), 1471-1480. https://doi.org/10.1007/s00704-019-02905-w

BPS. (2023). Bandung City in Figures. BPS Bandung City.

Dabral, P. P., & Murry, M. Z. (2017). Modeling and Forecasting of Rainfall Time Series Using SARIMA. Environmental Processes, 4(2), 399-419. https://doi.org/10.1007/s40710-017-0226-y

Farikhul Firdaus, R., & Paputungan, I. V. (2022). Rainfall Prediction in Bandung City Using Long Short Term Memory Method. Journal of Innovative Research, 2(3), 453-460. https://doi.org/10.54082/jupin.99

Fauzy, M., Rahmat Saleh, K. W., & Asror, I. (2016). Application of Association Rule Method Using Apriori Algorithm on Simulation of Rain Prediction in Bandung City Area. Scientific Journal of Applied Information Technology, II (2), 221-227.

Fransiska, H., Novianti, P., & Agustina, D. (2019). Modeling Monthly Rainfall in Bengkulu City with Seasonal Autoregressive Integrated Moving Average (Sarima) Case Study: Climatology Station in Bengkulu. National Seminar on Official Statistics 2019: Development of Official Statistics in Supporting the Implementation of SDG's, 390-395.

Ginting, C. P., & Kartiasih, F. (2019). Analisis Ekspor Kopi Indonesia Ke Negara-Negara Asean. *Jurnal Ilmiah Ekonomi Dan Bisnis*, *16*(2), 143–157. https://doi.org/10.31849/jieb.v16i2.2922

Guo, Q., Feng, Y., Sun, X., & Zhang, L. (2017). Power Demand Forecasting and Application based on SVR. Procedia Computer Science, 122, 269-275. https://doi.org/10.1016/j.procs.2017.11.369

Gustari, I., Wahyu Hadi, T., Hadi, S., & Renggono, F. (2012). Accuracy of Operational Daily Rainfall Prediction in Jabodetabek: Comparison with WRF Model, Journal of Meteorology and Geophysics, 13(2), 119-130.

Hakiqi, M. I., Firmansyah, A., & Arisanti, R. (2023). Forecasting Rainfall in Bandung City with SARIMA (Seasonal Autoregressive Integrated Moving Average) Method. National Seminar on Statistics XI 2022, 1(1), 23-29. https://doi.org/10.12962/j27213862.v1i1.19119

Hasan, N., Nath, N. C., & Rasel, R. I. (2015). A support vector regression model for forecasting rainfall. 2015 2nd International Conference on Electrical Information and Communication Technologies (EICT), 554-559. https://doi.org/10.1109/EICT.2015.7392014

He, K., Ji, L., Wu, C. W. D., & Tso, K. F. G. (2021). Using SARIMA-CNN-LSTM approach to forecast daily tourism demand. Journal of Hospitality and Tourism Management, 49, 25-33. https://doi.org/10.1016/j.jhtm.2021.08.022

He, R., Zhang, L., & Chew, A. W. Z. (2024). Data-driven multi-step prediction and analysis of monthly rainfall using explainable deep learning. Expert Systems with Applications, 235, 121160. https://doi.org/10.1016/j.eswa.2023.121160

Jiang, P., Liu, F., & Song, Y. (2017). A hybrid forecasting model based on date-framework strategy and improved feature selection technology for short-term load forecasting. Energy, 119, 694-709. https://doi.org/10.1016/j.energy.2016.11.034

Kafara, Z., Rumlawang, F. Y., & Sinay, L. J. (2017). Rainfall Forecasting with Seasonal Autoregressive Integrated Moving Average (SARIMA) Approach. Journal of Mathematical and Applied Sciences, 11(1), 63-74.

Kartiasih, F., Rizky Ramadhani, A., Anisya Fitri, K., & Aselnino, P. (2022). Faktor-Faktor yang Mempengaruhi Volume Impor Jagung

Indonesia dari Lima Negara Eksportir Terbesar tahun 2009-2021. *INOVASI: Jurnal Ekonomi, Keuangan Dan Manajemen*, *18*(4), 936–946.

Kartiasih, F., & Setiawan, A. (2019). Efisiensi Teknis Usaha Tani Padi di Provinsi Kepulauan Bangka Belitung. *Analisis Kebijakan Pertanian*, *17*(2), 139. https://doi.org/10.21082/akp.v17n2.2019.139-148

Latifa, A., Primadani, A. D. P., Fitriyyah, N. R., & Kartiasih, F. (2023). Mapping and Estimating the Impact of Drought on Food Crop Farmers Using Remote Sensing in East Nusa Tenggara Province. *TheJournalish: Social and Government*, *4*(5), 309–335.

Latif, S. D., Alyaa Binti Hazrin, N., Hoon Koo, C., Lin Ng, J., Chaplot, B., Feng Huang, Y., El-Shafie, A., & Najah Ahmed, A. (2023). Assessing rainfall prediction models: Exploring the advantages of machine learning and remote sensing approaches. Alexandria Engineering Journal, 82, 16-25. https://doi.org/10.1016/j.aej.2023.09.060

Lu, H., Azimi, M., & Iseley, T. (2019). Short-term load forecasting of urban gas using a hybrid model based on improved fruit fly optimization algorithm and support vector machine. Energy Reports, 5, 666-677. https://doi.org/10.1016/j.egyr.2019.06.003

Makridakis, S. G., Wheelwright, S. C., & Hyndman, R. J. (1984). The Forecasting Perspective.

Maulana, A., & Kartiasih, F. (2017). Analisis Ekspor Kakao Olahan Indonesia ke Sembilan Negara Tujuan Tahun 2000–2014. *Jurnal Ekonomi Dan Pembangunan Indonesia*, *17*(2), 103–117. https://doi.org/10.21002/jepi.v17i2.664

Mehdizadeh, S., Behmanesh, J., & Khalili, K. (2018). New Approaches for Estimation of Monthly Rainfall Based on GEP-ARCH and ANN-ARCH Hybrid Models. Water Resources Management, 32(2), 527-545. https://doi.org/10.1007/s11269-017-1825-0

Muhamad, H., Cholissodin, I., & Setiawan, B. D. (2017). Support Vector Regression (SVR) Optimization Using Improved-Particle Swarm Optimization (IPSO) Algorithm for Rainfall Forecasting (Vol. 1, Issue 11). http://j-ptiik.ub.ac.id

Novitasari, D. C. R., Rohayani, H., Suwanto, Arnita, Rico, Junaidi, R., Setyowati, R. D. N., Pramulya, R., & Setiawan, F. (2020). Weather Parameters Forecasting as Variables for Rainfall Prediction using Adaptive Neuro Fuzzy Inference System (ANFIS) and Support Vector Regression (SVR). Journal of Physics: Conference Series, 1501(1), 012012. https://doi.org/10.1088/1742-6596/1501/1/012012

Nuraini Anitasari. (2022, January 13). Getting to Know Forecasting Methods and Their Benefits for Business - Zahir. Zahir Accounting Blog. https://zahiraccounting.com/id/blog/forecasting-dalam-sektor-industri/

Ojo, O. S., & Ogunjo, S. T. (2022). Machine learning models for prediction of rainfall over Nigeria. Scientific African, 16, e01246. https://doi.org/10.1016/j.sciaf.2022.e01246

Oktaviani, C., & Afdal. (2013). Monthly Rainfall Prediction Using Artificial Neural Network with Multiple Backpropagation Training Functions. Journal of Physics Unand, 2(4), 228-237.

Pai, P., & Hong, W. (2007). A recurrent support vector regression model in rainfall forecasting. Hydrological Processes, 21(6), 819-827. https://doi.org/10.1002/hyp.6323

Pemayun, A. A. G. R. B. D., Azizi, M. Z., Daulay, N. A., Apriliani, N. H., & Kartiasih, F. (2024). Estimation of Java GRDP in Regency/City Level: Satellite Imagery and Machine Learning Approaches. *JURTEKSI (Jurnal Teknologi Dan Sistem Informasi)*, *X*(2), 379–386. http://dx.doi.org/10.33330/jurteksi.v10i2.2993

Ponnoprat, D. (2021). Short-term daily precipitation forecasting with seasonally-integrated autoencoder. Applied Soft Computing, 102, 107083. https://doi.org/10.1016/j.asoc.2021.107083

Prianda, B. G., & Widodo, E. (2021). Comparison of Seasonal Arima and Extreme Learning Machine Methods in Forecasting the Number of Foreign Tourists to Bali. BAREKENG:

Journal of Mathematical and Applied Sciences, 15(4), 639-650. https://doi.org/10.30598/barekengvol15iss4pp639-650

Prihatin, R. (2015). Land Use Change in Urban Areas (Case Studies in Bandung and Yogyakarta). Aspirasi: Journal of Social Issues, 6(2), 105-118.

Purwoko, C. F. F., Sediono, S., Saifudin, T., & Mardianto, M. F. F. (2023). Prediction of Non-Oil and Gas Export Prices in Indonesia Based on Fourier Series Estimator and Support Vector Regression Methods. Inference, 6(1), 45. https://doi.org/10.12962/j27213862.v6i1.15558

Puspita, N., Afendi, F. M., & Sartono, B. (2022). Comparison Of SARIMA, SVR, and GA-SVR Methods Forecasting the Number of Rainy Days in Bengkulu City. BAREKENG: Journal of Mathematical and Applied Sciences, 16(1), 355-362. https://doi.org/10.30598/barekengvol16iss1pp353-360

Rachma Safitri, V., & Kartiasih, F. (2019). Daya Saing dan Faktor-Faktor yang Mempengaruhi Ekspor Nanas Indonesia. *Jurnal Hortikultura Indonesia*, *10*(1), 63–73. https://doi.org/10.29244/jhi.10.1.63-73

Ray, S., Das, S. S., Mishra, P., & Al Khatib, A. M. G. (2021). Time Series SARIMA Modeling and Forecasting of Monthly Rainfall and Temperature in the South Asian Countries. Earth Systems and Environment, 5(3), 531-546. https://doi.org/10.1007/s41748-021-00205-w

S.R. Gunn. (1998). Support Vector Machines for Classification and Regression. https://eprints.soton.ac.uk/256459/

Sajidul Fajri, Kurniati, E., & Suhaedi, D. (2023). Modeling Rainfall in Bandung City Using Seasonal Autoregressive Integrated Moving Average Model on Time Series Data with Minitab Assistance. Bandung Conference Series: Mathematics, 3(1), 7-17. https://doi.org/10.29313/bcsm.v3i1.6121

Simanjuntak, P. P. (2023). Comparison of Output Performance of Artificial Neural Network and SARIMA Models for Predicting the Beginning of the Rainy Season in Pangkalpinang City. J Statistics, 16(1), 407-423.

Siregar, N. A. (2022). Forecasting Rainfall in Medan City using the Support Vector Regression Method. Journal of Informatics and Data Science (J-IDS, 1(1).

Soekendro, C. A. (n.d.). Prediction of Rainfall in Bandung Regency with Time Series Analysis, Using the Sarima (Seasonal Autoregressive Integrated Moving Average) Model.

Subian, A. R., Mulkan, D. A., Ahmady, H. H., & Kartiasih, F. (2024). Comparison Methods of Machine Learning and Deep Learning to Forecast The GDP of Indonesia. *SISTEMASI: Jurnal Sistem Informasi*, *13*(1), 149–166. https://doi.org/10.32520/stmsi.v13i1.3445

Suci, K. W., & Irhamah. (2017). Forecasting Rainfall as a Supporter of the Rice Planting Calendar at Kedungadem Bojonegoro Post Using ARIMA, Support Vector Regression and Genetic Algorithm-SVR. ITS Journal of Science and Arts, 6(1), D55-D-61.

Supu, I., Usman, B., & Basri, S. (2016). Effect of Temperature on Heat Transfer in Different Materials.

Tan, Z., Zhang, J., He, Y., Zhang, Y., Xiong, G., & Liu, Y. (2020). Short-Term Load Forecasting Based on Integration of SVR and Stacking. IEEE Access, 8, 227719-227728. https://doi.org/10.1109/ACCESS.2020.3041779

Xiang, Y., Gou, L., He, L., Xia, S., & Wang, W. (2018). A SVR-ANN combined model based on ensemble EMD for rainfall prediction. Applied Soft Computing, 73, 874-883. https://doi.org/10.1016/j.asoc.2018.09.018

Yuliana, R., Aulia, W. D., & Kartiasih, F. (2023). Benarkah Beras Tidak Tergantikan Sebagai Makanan Pokok di Indonesia? *Prosiding Kontribusi Forum Akademisi SAC*, *1*(September), 340–356.