International Journal of Data and Network Science 6 (2022). 1309–1318

Contents lists available at GrowingScience

International Journal of Data and Network Science

homepage: www.GrowingScience.com/ijds

The implementation of the ARIMA-ARCH model using data mining for forecasting rainfall in Bandung city

Putri Monika[a*], Budi Nurani Ruchjana[b] and Atje Setiawan Abdullah[c]

[a]*Magister Student at Mathematics Postgraduate Program, Universitas Padjadjaran, Sumedang, 45363, Indonesia*
[b]*Department of Mathematics, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, 45363, Indonesia*
[c]*Department of Computer Science, Faculty of Mathematics and Natural Sciences, Universitas Padjadjaran, Sumedang, 45363, Indonesia*

C H R O N I C L E

Article history:
Received: February 18, 2022
Received in revised format: March 27, 2022
Accepted: May 30, 2022
Available online: June 2 2022

Keywords:
ARIMA-ARCH
Data Mining
KDD
Forecasting
Rainfall

A B S T R A C T

A time series is a stochastic process which is arranged by time simultaneously. In this article, a time series model is used in accordance with Box-Jenkins' procedure. The Box-Jenkins procedure consists in identifying the model, estimating the parameters and diagnostic checking. The time series model is differentiated according to the number of variables, i.e. univariate and multivariate. The univariate method for the time series model that is often used is the Autoregressive Integrated Moving Average (ARIMA) model and the multivariate time series model is the Vector Autoregressive Integrated Moving Average (VARIMA) model. In this research, we studied the ARIMA model which is studied with a non-constant error variance. In this case, the Autoregressive Conditional Heteroscedasticity (ARCH) model is applied to outgrow the non-constant error variance. Selection of the best model by examining the minimum AIC for each model. The ARIMA-ARCH model is implemented on rainfall data in Bandung city with Knowledge Discovery in Database (KDD) in Data Mining. The methodology in the KDD process, including pre-processing, data mining process, and post-processing. Based on the results of model fitting, the best model is the ARIMA (2,1,4)-ARCH (1) model. The result of forecasting rainfall in Bandung shows a MAPE value is 11%, which has a similar pattern with actual data for short time 2-4 days. From these results, we conclude that the ARIMA-ARCH model is a good model for forecasting the rainfall in Bandung city.

1. Introduction

Climate events can be described as changes in the atmosphere over a long period of time. Climate parameters include rainfall, temperature, humidity, air pressure, etc. Climate change has the potential to affect areas of human life. The impact on the climate of the western region of Java is an increase of rainfall, which affects food centers or agriculture (BMKG, 2022). The rainfall changes are important to discuss because they influence food or agriculture in Bandung city. Rainfall prediction has been done by many researchers using various models and methods. Nayak et al. (2013) has done rainfall prediction with Artificial Neural Network (ANN) for nonlinear systems. A survey conducted shows that rainfall prediction has a better prediction using ANN than traditional statistical methods. Mishra et al. (2018); Sahai et al. (2000) have done rainfall prediction using ANN with Northern India dataset. Lee et al. (1998) performed a rainfall prediction at 367 locations which divide and conquer each location into four sub-areas. RBF network (Neural Network) approach is used for two larger areas. Prediction in two smaller areas is used as a simple linear regression. Comparison between the two results shows that RBF has a good prediction. Many researchers have developed the Artificial Neural Network (ANN) model for rainfall prediction, including: The Multilayered Artificial Neural Network with back-propagation algorithm (Abhishek et al., 2012), a hybrid model with

* Corresponding author.
E-mail address: putri17001@mail.unpad.ac.id (P. Monika)

doi: 10.5267/j.ijdns.2022.6.004

Artificial Neural Network and Support Vector Regression (Chau & Wu, 2010), and the wavelet technique and Artificial Neural Network (Venkata et al., 2013). Another model used in the rainfall prediction is self-organizing map (SOM), backpropagation neural network (BPNN) and fuzzy rule systems (Wong et al., 2003). Toth et al. (2000) discuss the comparison of the rainfall prediction using two techniques such as autoregressive moving average (ARMA), Artificial Neural Network (ANN) and the non-parametric nearest-neighbours method.

In this research, the rainfall arranged simultaneously by a time index is called time series. In this research, we studied a time series model which is the ARIMA model. The ARIMA model uses the Box-Jenkins procedure to carry out the application of the model. The ARIMA model assumes a constant error variance. In some phenomena, these assumptions are irrelevant because there is an error variance that is not constant. To treat the problem of non-constant error variance (heteroscedasticity), we used the Autoregressive Conditional Heteroscedasticity (ARCH) model (Engle, 1982). The application of the Hybrid ARIMA – ARCH model has been carried out to predict stock prices and gold price volatility. Yusof et al. (2013) used the ARIMA-GARCH hybrid model to forecast rainfall in Malaysia. Yan & Chen (2018) predicts China's mobile marketing using the ARIMA-ARCH model. In this study, forecasting of climate phenomena in the city of Bandung city was carried out using the ARIMA-ARCH model.

In this research, the data source used is the data from NASA satellite observation (POWER NASA) which is big data that has 5 characteristics, including value, volume, variety, velocity, and veracity (Anuradha, 2015). Climate data collected on POWER NASA's website have many climate variables and various locations. The methodology used is the KDD process in Data Mining for describing and predicting. Nikam & Meshram (2013) have done rainfall prediction using Bayesian approach with data mining technique. Based on the description above, this research is the ARIMA-ARCH model by means of the Data Mining approach to predict rainfall in Bandung city.

2. Materials

2.1. Data Mining

Knowledge Discovery in Databases (KDD) is a process of finding patterns and relationships between variables in large data. The procedure in the KDD process in data mining is as follows (Han et al., 2012):

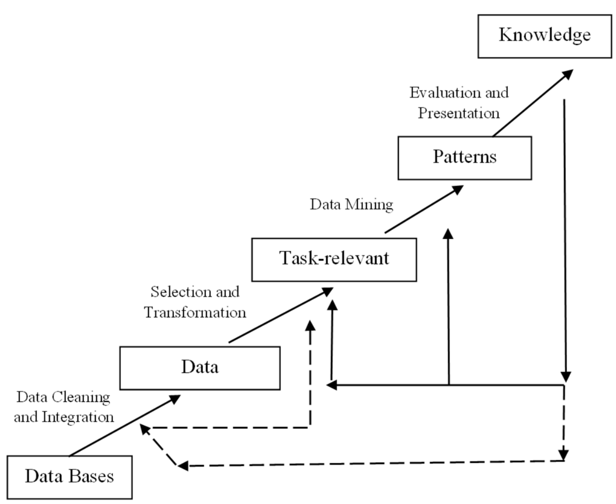

Fig 1. KDD Process in Data Mining

A. Pre-Processing

Pre-processing data is the stage of processing raw data into data that is ready to be processed in the data mining process. Pre-processing data includes:

a. Data Cleaning is the process of cleaning all missing values and noisy data in the database.
b. Data Selection is the process of selecting data and variables that can support research.
c. Data Transformation is a process where data is transformed in several ways so that the data is in accordance with the mining process.
d. Data Integration is combining and grouping data from different sources.

B. Data Mining Process

In the data mining process, there are two steps: Mining Process and Pattern Evaluation. Mining Process is an important process that is carried out with many different techniques and algorithms to extract useful data patterns so as to obtain good output and answers. Pattern evaluation is a stage to obtain the final result of the mining process by identifying models and patterns.

C. Post-Processing

Knowledge Presentation is the last process by presenting the knowledge found through summarization and visualization techniques.

2.2. Time Series Model

A time series is a sequential observation with the same time interval (Wei, 2006). Based on the number of variables, time-series data analysis is divided into two, namely univariate time series and multivariate time series (Wei, 2006). One of the univariate stationary time series models is the Autoregressive Moving Average (ARMA) model and the univariate non-stationary time series model is the Autoregressive Integrated Moving Average (ARIMA) model. Meanwhile, the stationary model with a multivariate method is the VARMA model and the multivariate non-stationary model is the VARIMA model.

2.2.1. Stationary Time Series Model

Stationary data in the time series model are data that do not have a significant change. In general, a time-series data $Z(t)$ is said to be stationary if:

a. Mean: $E[Z(t)] = \mu(t)$
b. Variance: $Var[Z(t)] = E[Z(t) - \mu(t)]^2 = \sigma(t)^2$
c. Covariance: $Cov\left[Z(t_1), Z(t_2)\right] = E[Z(t_1) - \mu(t_1)][Z(t_2) - \mu(t_2)]$

The statistical test used to determine whether it is stationary or not is the Augmented Dickey Fuller (ADF) test.

2.2.2. Differencing Process

According to Wei (2006), the differencing process is a process to overcome modeling of non-stationary data of the average by differentiating or reducing the value of the observation $Z(t)$ with the value of the previous observation $Z(t-1)$. A time series data that is not stationary can be carried out by a first-order differencing process with the following equation:

$$\Delta = Z(t) - Z(t-1) \quad (1)$$

where Δ represents the first difference series, $Z(t)$ and $Z(t-1)$ denote the observation values at time *t* and *t-1,* respectively.

2.2.3. Autocorrelation Function and Partial Autocorrelation Function

Autocorrelation function (ACF) and Partial Autocorrelation Function (PACF) can be used for identifying the univariate time series model (Wei, 2006). Basically, ACF and PACF are used to describe patterns that exist in a univariate time series. The ACF formula is:

$$\hat{\rho}_k = \frac{\sum_{t=1}^{T-k}(Z(t)-\bar{Z})(Z(t+k)-\bar{Z})}{\sum_{t=1}^{T}(Z(t+k)-\bar{Z})^2} = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}, k = 0,1,2, \ldots \quad (2)$$

The formula for calculating PACF is:

$$\hat{P}_k = \frac{Cov\left[\left(Z(t)-\hat{Z}(t)\right), \left(Z(t+k)-\hat{Z}(t+k)\right)\right]}{\sqrt{Var\left(Z(t)-\hat{Z}(t)\right)}\sqrt{Var\left(Z(t+k)-\hat{Z}(t+k)\right)}} \quad (3)$$

The process of identifying the time series model using ACF and PACF is presented in Table 1 as follows:

Table 1
Identify models using ACF and PACF

Model	ACF	PACF
AR($\boldsymbol{p}$)	dies down	cut off after lag-p
MA($\boldsymbol{q}$)	cut off after lag-q	dies down
ARMA($\boldsymbol{p},\boldsymbol{q}$)	dies down after lag $(q-p)$	dies down after lag $(p-q)$

2.2.4. Autoregressive Integrated Moving Average Model (ARIMA)

Autoregressive (AR) model is a time series model that shows the relationship of a random variable with itself at some previous time (lag) and is added to an error. The AR model for order p is as follows:

$$Z(t) = \mu + \sum_{k=1}^{p} \varphi_k Z(t-k) + e(t), \tag{4}$$

where $e(t) \overset{iid}{\sim} N(0, \sigma^2)$. AR model for order 1/ AR (1) can be described as follows:

$$Z(t) = \varphi_1 Z(t-1) + e(t). \tag{5}$$

The AR (1) model which is not stationary, then the differencing process is carried out until the stationary data becomes the ARI (1,1) model. The ARI (1,1) model is stated as follows:

$$Y(t) = \varphi_1 Y(t-1) + e(t) \tag{6}$$

where, $Y(t) = Z(t) - Z(t-1)$, $Y(t-1) = Z(t-1) - Z(t-2)$, and $e(t) \overset{iid}{\sim} N(0, \sigma^2)$ (Wei, 2006).

The Moving Average/MA (q) model is shown by the equation below:

$$Z(t) = \mu + \sum_{k=1}^{p} \theta_k e(t-k) + e(t) \tag{7}$$

where $e(t) \overset{iid}{\sim} N(0, \sigma^2)$. The ARIMA model is a combination of the AR model and the MA model. The ARIMA model (p, d, q) is shown by the equation below:

$$Z(t) = \mu + \sum_{k=1}^{p} \varphi_k Z(t-k) + \sum_{k=1}^{q} \theta_k e(t-k) + e(t) \tag{8}$$

where $e(t) \overset{iid}{\sim} N(0, \sigma^2)$ (Wei, 2006).

2.2.5. Autoregressive Conditional Heteroscedasticity Model (ARCH)

The ARCH model is a time series process with a zero mean and the conditional variance is not constant, but the unconditional variance is constant. ARCH model is a time series model that can model heteroscedastic properties (Engle, 1982).

The ARCH (m) model can be written in the following equation:

$$h_t = \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \cdots + \alpha_m e_{t-m}^2 \tag{9}$$

where h_t, α_0 and $\alpha_1, \alpha_2, \dots, \alpha_m$ are conditional variance at *t,* intercept and ARCH model parameters, respectively with $\alpha_0 > 0$ and $\alpha_i \geq 0$

2.2.6. Hybrid ARIMA-ARCH Model

The ARIMA (p, d, q)-ARCH (m) univariate model is expressed as follows:

$$\left.\begin{array}{c} Z(t) = \mu + \sum_{k=1}^{p} \varphi_k Z(t-k) + \sum_{k=1}^{q} \theta_k e(t-k) + e(t) \\ e(t) = \eta_t \sqrt{h_t},\ \eta_t \sim iid.N(0,1) \\ h_t = \sigma_t^2 = \alpha_0 + \alpha_1 e_{t-1}^2 + \alpha_2 e_{t-2}^2 + \cdots + \alpha_q e_{t-m}^2 \end{array}\right\} \tag{10}$$

2.2.7. Akaike's Information Criterion (AIC)

AIC is a criterion for selecting the best model by considering the number of parameters in the model (Wei, 2006). Selection of the best model based on the minimum AIC. AIC calculation can be done with the formula:

$$IC(M) = 2M - 2\, l\, n[maximum\ likelihood], \tag{11}$$

where M is the number of parameters in model.

2.2.8. Akaike's Information Criterion Corrected (AICc)

AICc is the corrected AIC calculated value. The use of the AICc value is better than the AIC value. The calculation formula is:

$$AICc = AIC + \frac{2M(M+1)}{n-M-1}, \quad (12)$$

where n the number of observations (Akaike, 1973).

2.2.9. Mean Percentage Error (MAPE)

MAPE is the percentage of the average absolute residual price in each period divided by the actual value in that period. The MAPE calculation is expressed in the following equation (Wei, 2019):

$$MAPE = \left[\frac{1}{N}\sum_{t=1}^{N}\left|\frac{e(t)}{Z(t)}\right|\right] \times 100\% \quad (13)$$

where $Z(t), e(t)$ and N are actual value of period $-t,$ error value of period $-t$ and number of time series observations, respectively. According to Lewis (1982) in Lawrence et al. (2009), the MAPE criteria in the level of forecasting accuracy are explained as follows:

< 10% : very accurate forecasting ability

10-20% : accurate forecasting ability

20-50% : forecasting ability is quite accurate

>50% : less accurate forecasting ability

3. Methodology

The research methodology follows the process of Knowledge Discovery in Databases (KDD) (Han et al., 2012) which goes through three stages of the process, including pre-processing, data mining process (using: ARIMA-ARCH Model), and post-processing. The data mining function in this research is the description and prediction. The research methodology as shown in Fig 2.

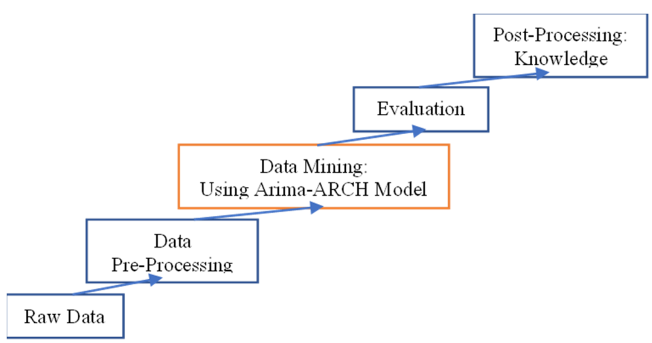

Fig 2. Research Methodology by Follow KDD Process

3.1. Raw Data

The data used in this research is the climate observation data on NASA satellites (NASA POWER) which is accessed online via https://power.larc.nasa.gov/data-access-viewer/. The climate data on POWER NASA consists of a wide variety of climate parameters across locations around the world. The climate data at POWER NASA is big data, so tools are needed to mine the data in order to obtain useful data.

3.2. Data Pre-Processing

Raw data from POWER NASA database goes through pre-processing with the following steps:

a. Selected the climate parameters, in this research, we used rainfall parameters.
b. Input latitude and longitude to select the location observed, we observed Bandung city.

c. Determine the observation date, from August 21, 2021 to March 1, 2022.
d. Cleaning data from missing values/missing data.
e. Divide the data into two parts, in-sample data 94% and out-sample data 6%.

3.3. Data Mining Process

The data mining process carried out in this research uses the ARIMA-ARCH model with the stages shown in Fig. 3:

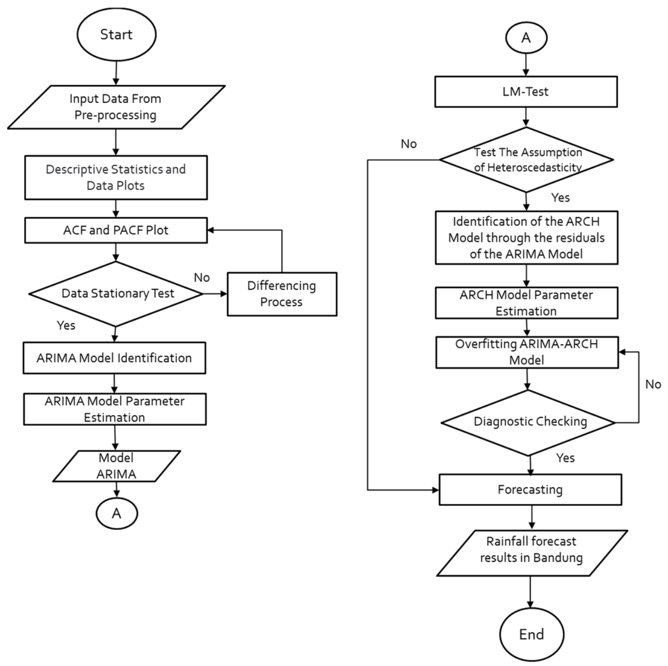

Fig 3. Data Mining Process using ARIMA-ARCH Model

3.4. Post-Processing

Visualization and interpretation of rainfall forecasting results in Bandung city which produces knowledge.

4. Result and Discussion

The ARIMA model in forecasting rainfall in Bandung city is carried out by following the Box-Jenkins procedures, namely identification model, parameter estimation, and diagnostic checking. In diagnostic checking the ARIMA Model, the assumption of heteroscedasticity was tested on the ARIMA residuals. In overcoming the assumption of non-constant error variance (heteroscedasticity), modeling of the ARIMA residuals is carried out using the ARCH model. The parameter estimation of the Hybrid ARIMA-ARCH model is obtained to forecast rainfall in Bandung city.

4.1. Pre-Processing

Data Mining approach is needed in this modelling with ARIMA-ARCH. The data used is sourced from rainfall data in Bandung city with coordinate of latitude = -6.914846 and coordinate of longitude = 107.608238. The data is sourced from secondary data of the POWER NASA website. The data recorded with daily time intervals. Table 2 shows the result of pre-processing data for rainfall data in Bandung city.

Table 2
Data Pre-Processing for Rainfall in Bandung City

Date	Rainfall (mm)
August 21, 2021	1.04
August 22, 2021	0.76
August 23, 2021	1.78
August 24, 2021	1.07
August 25, 2021	0.21
August 26, 2021	0.12
August 27, 2021	0.31
⋮	⋮
March 1, 2022	6.22

4.2. *Data Mining using ARIMA-ARCH Model*

4.2.1. *Descriptive Statistics and Plot Data*

Rainfall data in Bandung city are plotted using RStudio. The plot results are shown in Fig. 4 Based on the plot, the climate data in Bandung City have an up and down trend. Descriptive Statistics show that the minimum value of the data is 0.11, maximum value is 189.83, and the mean of the data is 14.85.

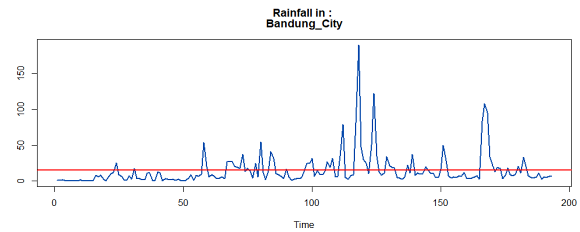

Fig 4. Plot Time Series for Rainfall Data in Bandung City

4.2.2. *Stationary Data Test*

Stationary test of rainfall data was carried out using the Augmented Dickey Fuller (ADF) test through RStudio. Table 3 shows the results of the ADF test on rainfall data in Bandung city. Based on the table, it can be seen that p-value = 0.1191 > 0.05 so that it is concluded that the data is not stationary, then by doing differencing once, it is obtained p-value = 0.01 <0.05 so that it can be concluded that the data is stationary.

Table 3
ADF Test for Stationary Data

Before Differencing		After Differencing	
ρ-value	Conclusion	ρ-value	Conclusion
0,1191	Non-Stationary	0,01	Stationary

Fig. 5 shown the time series plot of rainfall data in Bandung city which has a first order differencing. It can be seen that the rainfall data is stationary by average.

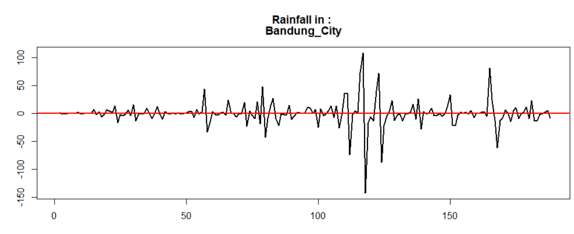

Fig 5. The Plot Time Series Data with First Order Difference

4.2.3. *ACF and PACF Plots*

Initial identification of the ARIMA model was carried out using ACF and PACF plot(s) calculated using Eqs .(2-3). The model is identified from the results of the ACF and PACF plot(s) by the criteria in Table 1. Fig. 6 shows the ACF and PACF plot(s) on the in-sample data. The ACF and PACF plots show a dies-down pattern, so the initial assumption of the model is the ARIMA model with a first order difference.

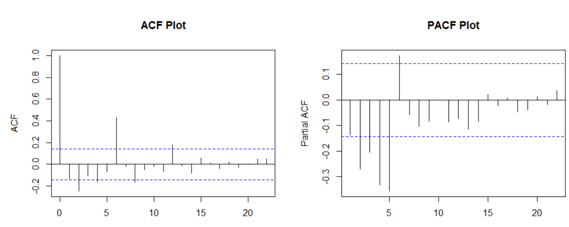

Fig 6. ACF and PACF Plots

4.2.4. ARIMA Model Identification

Based on the results of the ACF and PACF plots, a trace is carried out to see if ARIMA models can be used. Table 4 shows the ARIMA model obtained along with the AICc value calculated by equation (12). The best model selection is done by selecting the model with the minimum AICc value. Based on the result, the ARIMA (2,1,4) model was selected as the best model and then parameter estimation was carried out.

Table 4
Identification Model

Model	AICc Value	Model	AICc Value
ARIMA(0,1,0)	1699.374	ARIMA(2,1,2)	1651.25
ARIMA(0,1,1)	1684.684	ARIMA(2,1,3)	1649.092
ARIMA(1,1,1)	1652.568	ARIMA(2,1,4)	1641.344
ARIMA(2,1,0)	1687.666	ARIMA(3,1,1)	1652.433
ARIMA(2,1,1)	1649.232	ARIMA(3,1,4)	1643.984

4.2.5. Estimation Parameter for ARIMA Model

The selected model is based on the information criteria, then parameter estimation is carried out. Table 5 shows the results of the parameter estimates for the ARIMA (2,1,4) model on rainfall data in Bandung city.

Table 5
Result of Estimation Parameter for ARIMA (2,1,4) Model

Coefficients	ar1	ar2	ma1	ma2	ma3	ma4
Estimates	0.5403	-0.7908	-0.9565	0.7215	-0.3622	-0.3055
S.E	0.1094	0.1088	0.1328	0.1542	0.0996	0.0892

4.2.6. LM-Test for ARCH Effect

Lagrange multiplier (LM) test is used to see the error variance in the ARIMA model. The LM test for ARIMA residuals are given in Fig. 7 and Table 6. Based on Fig. 7, the residual and square residual have non-constant error variances. Then it was clarified again through the LM test which showed the p-value < 0.05. So, it is concluded that the residual ARIMA model contains ARCH effects.

Table 6
ARCH LM Test for Residual ARIMA (2,1,4) Model

Order	LM	p-value
4	141.2	0.00e+00
8	44.5	1.74e-07
12	27.6	3.66e-03
16	19.1	2.09e-01
20	14.4	7.59e-01

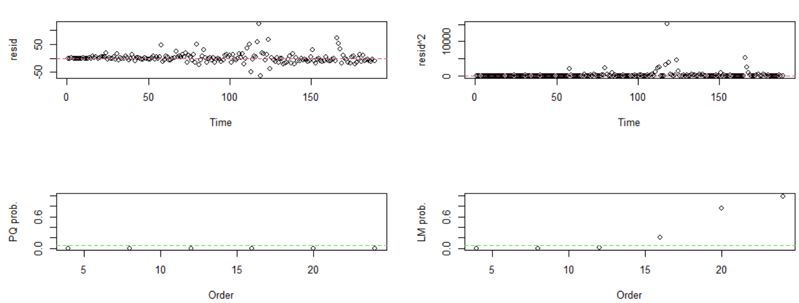

Fig 7. LM-Test for Residual ARIMA (2,1,4) Model

4.2.7. Parameter Estimation for ARIMA-ARCH Model

The parameter estimates for the ARIMA (2,1,4) and ARCH (1) models for rainfall data in Bandung city are shown in Table 7. The standardization residual test is presented in Table 8. Based on the results of the LM-test in Table 8, it can be seen that there is no ARCH effect on the residual ARIMA (2,1,4)-ARCH (1) model with p-value = 0.78 > 0.05.

Table 7
The Result of Parameter Estimation for ARIMA-ARCH Model

Coefficient(s)	Estimate	Std. Error	t value
ar1	0.13681	0.13936	0.982
ar2	-0.54242	0.11421	-4.749
ma1	-0.58713	0.09692	-6.058
ma2	0.31558	0.12072	2.614
ma3	0.30132	0.08228	-3.662
ma4	-0.39125	0.04569	-8.563
alpha1	1.00000	0.33837	2.955

Table 8
Standardized Residual Tests

Test(s)	Residual	Statistics	p-value
Ljung-Box Test	R	11.49596	0.3202041
Ljung-Box Test	R^2	7.976963	0.6310872
LM ARCH Test	R	8.061777	0.7802843

4.2.8. Diagnostics Checking

Next, diagnostic checking is carried out to see whether the model assumptions have been met or not. The Q-Q plot is shown in Fig. 8 Based on the results of the plot, the residual forecast results are close to the normal distribution. And the results of the error test using the Ljung-Box have a p-value of more than 5% (0.05). This means that the forecasting results in Bandung city have uncorrelated errors.

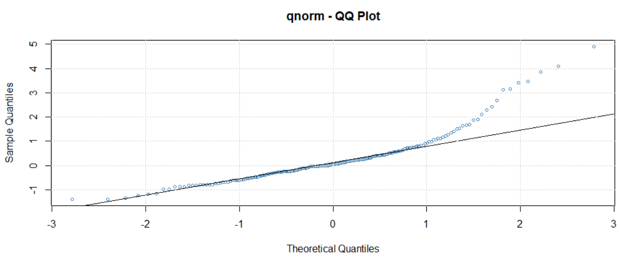

Fig. 8. Q-Q Plot of ARIMA-ARCH Model

Table 9
Forecasting using ARIMA-ARCH Model

Actual	Forecast	Error
5.96	5.503478	0.4565218
4.95	4.563255	0.3867452
6.25	5.017553	1.2324469
6.22	6.780652	-0.5606518
MAPE		11.0514%

4.3. Post-Processing

Rainfall forecasting using the ARIMA-ARCH model is applied for forecast the next four days. Table 9 shows the actual data, forecast result, and error. The accuracy of the model in forecasting is calculated using MAPE based on Eq. (13). MAPE results are 11%, which shows good forecasting results.

5. Conclusion

This paper discusses rainfall forecasting in Bandung city through the ARIMA-ARCH model in terms of overcoming the assumption of an error variance that is not constant. The methodology of the research follows the KDD Data Mining process. The data mining process uses the Hybrid ARIMA-ARCH model following the Box Jenkins procedure. Based on the stages of the research conducted, the results show that the forecasting of rainfall in Bandung city only satisfied for a short term i.e. for the next four days. It still has a good result for forecasting because it has a small MAPE, which is 11%.

Acknowledgement

We thank the Rector of Universitas Padjadjaran, Indonesia, who have provided funding through the Padjadjaran Excellence Fastrack Scholarship (BUPP) with a contract number: 1595/UN6.3.1/PT.00/2021. Also, thanks to the reviewers for the valuable review of this paper.

References

Abhishek, K., Kumar, A., Ranjan, R., & Kumar, S. (2012, July). A rainfall prediction model using artificial neural network. In *2012 IEEE Control and System Graduate Research Colloquium* (pp. 82-87). IEEE.

Agboola, A. H., Gabriel, A. J., Aliyu, E. O., & Alese, B. K. (2013). Development of a fuzzy logic based rainfall prediction model. *International journal of Engineering and Technology*, *3*(4), 427-435.

Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia computer science*, *48*, 319-324.

BMKG (2022). https://www.bmkg.go.id/iklim/dinamika-atmosfir.bmkg. (Online: Accessed March 1, 2022).

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, *31*(3), 307-327.

Bonar, H., Ruchjana, B. N., & Darmawan, G. (2017, March). Development of generalized space time autoregressive integrated with ARCH error (GSTARI–ARCH) model based on consumer price index phenomenon at several cities in North Sumatera province. In *AIP Conference Proceedings* (Vol. 1827, No. 1, p. 020009). AIP Publishing LLC.

Box, G. E., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American statistical Association*, *65*(332), 1509-1526.

Box, G. E., & Jenkins, G. M. (1976). Time series analysis: Forecasting and control. San Francisco. *Calif: Holden-Day*.

Chau, K. W., & Wu, C. L. (2010). A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. *Journal of Hydroinformatics*, *12*(4), 458-473.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the econometric society*, 987-1007.

Han, J., Kamber, M., and Pei,J., (2012) '*Data Mining: Concept and Techniques'*. Elsevier,Inc.

Hernández, E., Sanchez-Anguix, V., Julian, V., Palanca, J., & Duque, N. (2016, April). Rainfall prediction: A deep learning approach. In *International Conference on Hybrid Artificial Intelligence Systems* (pp. 151-162). Springer, Cham.

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.

Lawrence, K. D., Klimberg, R. K., & Lawrence, S. M. (2009). *Fundamentals of forecasting using excel*. Industrial Press Inc.

Lee, S., Cho, S., & Wong, P. M. (1998). Rainfall prediction using artificial neural networks. *journal of geographic information and Decision Analysis*, *2*(2), 233-242.

Ling, S., & McAleer, M. (2003). Asymptotic theory for a vector ARMA-GARCH model. *Econometric theory*, *19*(2), 280-310.

Ljung, G. M. (1986). Diagnostic testing of univariate time series models. *Biometrika*, *73*(3), 725-730.

Mishra, N., Soni, H. K., Sharma, S., & Upadhyay, A. K. (2018). Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data. *International Journal of Intelligent Systems & Applications*, *10*(1).

Nainggolan, N., Ruchjana, B. N., Darwis, S., & Siregar, R. E. (2010). GSTAR Models with ARCH Errors and The Simulation. In *Proceeding of the Third Internastional Conference on Mathematics and Natural Sciences (ICMNS2010)* (pp. 1075-1084).

Nayak, D. R., Mahapatra, A., & Mishra, P. (2013). A survey on rainfall prediction using artificial neural network. *International Journal of Computer Applications*, *72*(16).

Nikam, V. B., & Meshram, B. B. (2013, September). Modeling rainfall prediction using data mining method: A Bayesian approach. In *2013 Fifth International Conference on Computational Intelligence, Modelling and Simulation*, 132-136.

Power NASA (2022). Prediction of Worldwide Energy Resource. https://power.larc.nasa.gov/data-access-viewer/. (Online: Accessed March 1, 2022).

Sahai, A. K., Soman, M. K., & Satyan, V. (2000). All India summer monsoon rainfall prediction using an artificial neural network. *Climate dynamics*, *16*(4), 291-302.

Sethi, N., & Garg, K. (2014). Exploiting data mining technique for rainfall prediction. *International Journal of Computer Science and Information Technologies*, *5*(3), 3982-3984.

Toth, E., Brath, A., & Montanari, A. (2000). Comparison of short-term rainfall prediction models for real-time flood forecasting. *Journal of hydrology*, *239*(1-4), 132-147.

Venkata Ramana, R., Krishna, B., Kumar, S. R., & Pandey, N. G. (2013). Monthly rainfall prediction using wavelet neural network analysis. *Water resources management*, *27*(10), 3697-3711.

Wei, W. (2006). *Time Series Analysis: Univariate and Multivariate Methods (2nd Edition)*. In Pearson: Addison Wesley.

Wei, W. (2019). *Multivariate Time Series Analysis and Applications*. John Wiley & Son Ltd.

Wong, K. W., Wong, P. M., Gedeon, T. D., & Fung, C. C. (2003). Rainfall prediction model using soft computing technique. *Soft Computing*, *7*(6), 434-438.

Yan, B., & Chen, Z. (2018). A prediction approach for precise marketing based on ARIMA-ARCH Model: A case of China Mobile. *Communications in Statistics-Theory and Methods*, *47*(16), 4042-4058.

Yusof, F., Kane, I. L., & Yusop, Z. (2013). Hybrid of ARIMA-GARCH modeling in rainfall time series. *Jurnal Teknologi*, *63*(2).

Zainudin, S., Jasim, D. S., & Bakar, A. A. (2016). Comparative analysis of data mining techniques for Malaysian rainfall prediction. *International Journal of Advanced Science Engineering Information Technology*, *6*(6), 1148-1153.