

Rainfall prediction model in Semarang City using machine learning

Carissa Devina Usman¹, Aris Puji Widodo², Kusworo Adi³, Rahmat Gernowo³

¹Department of Mathematics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

²Department of Informatics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

³Department of Physics, Faculty of Science and Mathematics, Diponegoro University, Semarang, Indonesia

Article Info

Article history:

Received Jun 16, 2022

Revised Jan 20, 2023

Accepted Jan 23, 2023

Keywords:

Artificial neural network
Machine learning
Multiple linear regression
Rainfall prediction
Random forest regression

ABSTRACT

The erratic distribution of rainfall greatly affects people's daily activities, especially in Semarang City, so it is necessary to predict rainfall. Correct prediction of rainfall can improve community preparedness in dealing with natural disasters. Algorithms for machine learning and data mining have been extensively utilized in research involving rainfall data from various regions. The primary objectives of this study are to find the best regression algorithm and use machine learning algorithms to predict rainfall in Semarang. The dataset used is daily rainfall data for the City of Semarang from the meteorological, climatological, and geophysical agency (BMKG). Machine learning algorithms such as multiple linear regression, random forest regression, and replicated neural networks will be used to conduct regression analysis on this dataset. The mean absolute error and Root mean squared error techniques are utilized to evaluate the performance of machine learning algorithms. With an error rate of 13.055 for root mean squared error (RMSE) and 6.621 for mean absolute error (MAE), the results of the research indicate that the performance of the neural network algorithm is superior to that of other algorithms.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Aris Puji Widodo

Department of Informatics, Faculty of Science and Mathematics, Diponegoro University

Semarang, Indonesia

Email: arispw@gmail.com

1. INTRODUCTION

Residents' day-to-day activities are significantly impacted by rainfall. Additionally, a variety of natural disasters that are certain to cause harm to residents can be brought on by severe rains. According to disaster data from the Semarang City regional disaster management body (BPBD) for the years 2018, 2019, and 2020, excessive rainfall has resulted in a variety of natural disasters, including floods, soil erosion, plant death, and house collapse. The socio-economic impact of rain ought to be examined because its effects range from disruptions in displacement networks to the destruction of flood-affected infrastructure [1]. Rainfall forecasts were also required in another study to mitigate the effects of falling agricultural production rates [2]. Therefore, increasing community preparedness, particularly in the Semarang City, Indonesia, region, necessitates calculating the amount of rainfall.

Cloudy, light rain, medium rain, heavy rain, very heavy rain, and excessive rain are the six types of rain, according to the meteorological, climatological, and geophysical agency (BMKG). agency. Rainfall amounts range from 0 to 0.5 millimeters per day in overcast conditions to 0.5 to 20 millimeters per day in light rain, 20 to 50 millimeters per day in medium rain, 50 to 100 millimeters per day in heavy rain, 100 to 150 millimeters per day in very heavy rain, and more than 150 millimeters per day in extreme rain. As a result, the

kind of rain that will fall can also be predicted if one day the severity of the rainfall is known. The minimum temperature, maximum temperature, general temperature, humidity, length of sunlight, maximum wind speed, general wind speed, and wind direction are some of the variables that can be used to assess the severity of rainfall. Residents may find it easier to predict the weather as a result, minimizing losses caused by severe rain.

Machine learning algorithms are used to test rainfall predictions. Machine teaching and learning activities investigate how computers can use information to practice. The objective is for the computer program to learn to automatically recognize patterns in the environment and make intelligent decisions using information [3]. There are several algorithms owned by machine learning, but in general machine learning has two basic algorithms, that are supervised learning and unsupervised learning [4]. Rainfall data can be processed with a supervised learning algorithm, because it has a target variable, namely rainfall. While the prediction is done by doing regression.

Previous research on rainfall prediction is discussed in [5]-[9]. Liyew and Melese [5] using the multivariate linear regression, random forest, and extreme gradient boost algorithms, not only predicting whether or not it will rain, but also estimating how much rain will fall each day. However, this research only focuses on discussing the comparison of algorithm performance rather than discussing the output of each method. Suparta and Samah [6], Rachmawati *et al.* [7], prediction models with machine learning have been developed, namely Bayesian quantile regression and adaptive network based fuzzy inference system (ANFIS) to obtain accurate predictions. However, performance comparisons with other machine learning algorithms were not carried out. While studies on [8] and [9] uses a powerful deep learning architecture for threshold applications and a simplified form of estimated rainfall to present a comparative analysis. It is based on conventional machine learning algorithms.

We will use the regression method in machine learning, namely multiple linear regression, random forest regression, and neural networks, to examine daily rainfall estimates for the City of Semarang. These methods include information gathering methods, information preprocessing methods, machine learning methods, ability measurement methods, and reviews of previously conducted research.

2. METHOD

2.1. Data collection

The BMKG provided the daily rainfall data for Semarang City from January 1, 2018, to March 20, 2021 for the purposes of this study. There are 8 features in the data that have an effect on the amount of rainfall the minimum temperature, the maximum temperature, the general temperature, the humidity, the length of the day, the maximum wind speed, the general wind speed, and the direction of the wind, and has 1 target, namely rainfall. A Microsoft Excel file chart is used to record the rainfall information in the file.

2.2. Data preprocessing

A method of data mining to replace anomalous data is preprocessing. This approach eliminates the issue of insufficient, inconsistent, and imprecise raw data. This can result in the collection of low-quality data, which in turn can produce low-quality data. Cleaning data, reducing data, scaling data, transforming data, and partitioning data are the five primary roles that preprocessing data typically performs [10]. In this research, data cleaning was carried out by considering missing values, and data transformation was carried out.

There are two ways to handle missing values in building operational data. The first is to remove data samples with missing values, because most data mining algorithms cannot handle data with missing values. This method is selected when the class label value does not exist, and is used when the tuple has multiple attributes with empty values [11]. Only if the proportion of missing values is not significant can this method be used [10]. The second step is to practice using the missing number imputation technique to substitute inferred numbers for the missing data. Filling in missing values with the average number, median, or feature form is the most common method. The goal of data conversion is to turn data into a form that can be used in data mining techniques [12]. The square root, logarithm, arcsin, inverse, and others are examples of data transformations.

2.3. Machine learning

Predictions are made using a machine learning activity algorithm in this study. For research on rainfall estimation using regression problems, a review of some machine learning algorithms has been done. Multiple linear regression, random forest regression, and neural networks are the algorithms that will be compared to determine which one is the most effective for estimating rainfall.

2.3.1. Multiple linear regression (MLR)

A statistical technique known as MLR is used to determine relations between independent with dependent variable which is linear [13]. There are more than two independent variables and one dependent variable in multiple regression [14]. A particular MLR model is shown in (1) [15]:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon \quad (1)$$

where $\beta_0, \beta_1, \dots, \beta_m$ are model parameters. This is a constant whose true value is unknown, and is estimated from the data using least-squares estimation. While represents the error value.

Next, the following error assumptions are given [4]:

- Zero-mean assumption. Error ε is a random variable with mean or expected value $E(\varepsilon) = 0$.
- Assuming constant variance. The variance of ε denoted by σ^2 is constant, regardless of the value of x_1, x_2, \dots, x_m .
- Assumption of independence. The value of ε is independent.
- Normality assumption. Error ε is a normally distributed random variable.

In other words, the values of the error term ε_i are independent normal random variables with mean 0 and variance σ^2 .

Then given four implications for the behavior of the response variable y , as follows:

- Based on the zero-mean, we have as (2),

$$\begin{aligned} E(y) &= E(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon) \\ &= E(\beta_0) + E(\beta_1 x_1) + \dots + E(\beta_m x_m) + E(\varepsilon) \\ &= \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m \end{aligned} \quad (2)$$

that is, for each set of values for x_1, x_2, \dots, x_m , the mean of the y 's lies on the regression line.

- Based on the constant-variance assumption, we have the variance of y , $Var(y)$, given as (3):

$$Var(y) = Var(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \varepsilon) = Var(\varepsilon) = \sigma^2 \quad (3)$$

that is, regardless of which value is taken by the predictors x_1, x_2, \dots, x_m , the variance of the y 's is always constant.

- Based on the independence assumption, it follows that for any particular set of values for x_1, x_2, \dots, x_m , the values of y are independent as well.
- Based on the normality assumption, it follows that y is also a normally distributed random variable.

In other words, the values of the response variable y_i are independent normal random variables with variance σ^2 and mean $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m$.

2.3.2. Random forest regression (RFR)

The best ways to improve the performance of decision tree-based algorithms is to use ensembles of trees, such as random forest [16]. The random forest model usually performs well on a variety of problems, including features with non-linear relationships [5]. RFR is a machine learning based regression method. Its foundation lies in the bagging and random subspace methods [17]. During the training phase, it creates multiple decision trees and uses the class average as a prediction for all of them. The following is how the RF algorithm works:

- Take a data point p at random from the training set
- Construct a decision tree related to these p data points
- Take N number of trees to build and repeat steps a and b
- For new data points, create each N trees predicting y values for data points and assign the new data points to the average of all predicted y values.

During the nursery training period, a number of decision plants are selected and categories that are typical of individual plant regression or estimation are eliminated in random forest regression. Based upon [18] the RF algorithm works well with large data files, and using large data files with most of the numbers missing yields good experimental results. However, there are some drawbacks to this method as well. One of these is that the interpretation is difficult to understand when compared to individual tree, indicating that the feature is not very real or significant [19].

2.3.3. Artificial neural network (ANN)

ANNs are computational algorithms that were initially designed to simulate "neuron"-based biological systems. Based on the shape of human neurons, ANN can perform machine learning and learning

activities as well as pattern identification [20]. Based on the premise that biological neurons or human knowledge can be represented mathematically in the form of replicative neural networks [21]:

- a) The parts that process data are called neurons.
- b) The connecting link transmits a signal to the neuron as it moves.
- c) A number is multiplied by the sent signal for each connecting link.
- d) The output signal is divided by each neuron in its input network using an activation function that is typically non-linear and represents the sum of the weights of the input signals. There are three factors that determine a neural network:
 - Neuronal pattern that is associated with the network (network architecture).
 - How to predict the weight of a link (training/learning/algorithm method).
 - The amount of output from a neuron is determined by the activation function.

Layers and neurons are two of the many processing components that make up ANN. A weight-specific link connects each neuron to other neurons. There are three types of arrangements, which are [22]:

- a) The neuron units that serve as the neural network's information processing inputs make up the input layer.
- b) The hidden layer is made up of neuron units that are similar to the hidden layer and serve as a layer that keeps the reaction from the input going.
- c) One neuron in the output layer is responsible for sharing the solution to the input information.

The sigmoid function is one of many possibilities for the activation function of a neural network's hidden layers [23].

2.4. Measuring performance

The root mean squared error (RMSE) and mean absolute error (MAE) were used to compare and contrast the prediction performance of each machine learning algorithm utilized in this study. For measuring continuous elastic accuracy, the two most commonly used tools are RMSE and MAE. The RMSE has a ratio of measurements that is comparable to the information being evaluated, whereas the MAE provides the mean error in a manner that is more impulsive. As a result, this precision measure was chosen [24]. MAE measures the magnitude of the error in general across a series of predictions and observations that match regardless of direction as shown in (4) [5].

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j| \quad (4)$$

The RMSE is a quadratic scoring rule that measures the mean magnitude of the error. This is the square root of the average of the squared differences between predictions and actual observations as (5).

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2} \quad (5)$$

RMSE assigns a relatively large weight to big mistakes. This means that RMSE is very useful when big mistakes are not expected. MAE and RMSE can be used together to diagnose error variations in a set of estimates. The RMSE will always be greater than or equal to the MAE, the greater the comparison between the two, the greater the version of human error in the illustration. If $RMSE = MAE$, then all errors have the same magnitude [5].

3. RESULTS AND DISCUSSION

Because the data associated with the target is not balanced, the obtained daily rainfall dataset is not balanced. Regression estimates are better than classification for this reason. Cross-validation will be problematic if classification is attempted. The 1096-data-strong dataset was split in half, with 70% going to training and 30% to testing. Data pre-processing are used to deal with missing numbers in data pre-processing, and logarithmic conversion is used to transform the data. Additionally, the outcomes of studies utilizing random forest regression, multiple linear regression, and artificial neural network are as follows:

3.1. Results

Data processing performed by the multiple linear regression algorithm produces the (6):

$$y = -31.8589 - 204.12x_1 - 4.78861x_2 + 66.7904x_3 + 123.115x_4 - 0.416294x_5 - 4.29463x_6 + 0.549371x_7 + 1.54933x_8 \quad (6)$$

where y is the variable of the amount of rainfall, while $x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8$ are the variables that affect the rainfall as mentioned in section 2.1. While the random forest regression algorithm with 18 Trees, the depth limit for planted trees is 3, and the split subset limit is less than 5, producing the following output.

In Figure 1, all of the decision tree forms studied in the random forest are depicted in the Pythagorean Forest. It is depicted as a Pythagorean tree in the Pythagorean forest, with each illustration representing a randomly formed tree. Trees with the shortest branches and brightest colors are the best. This indicates that several variables of dividing branches effectively. The random forest model in Figure 1 is the result of 18 decision trees. A tree with 13 vertices and 7 leaves is selected displayed in the Figure 2. After that, we tried using the neural network algorithm with one output variable, eight hidden layers, and eight input variables to work with the data. Figure 2 depicts the neural network's architecture as Figure 3.

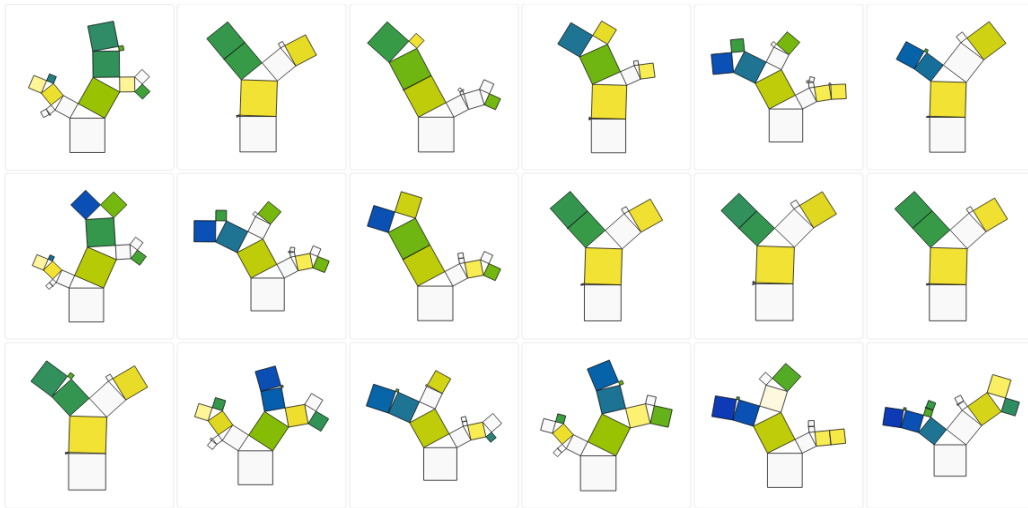


Figure 1. Visualization of trees in pythagorean forest

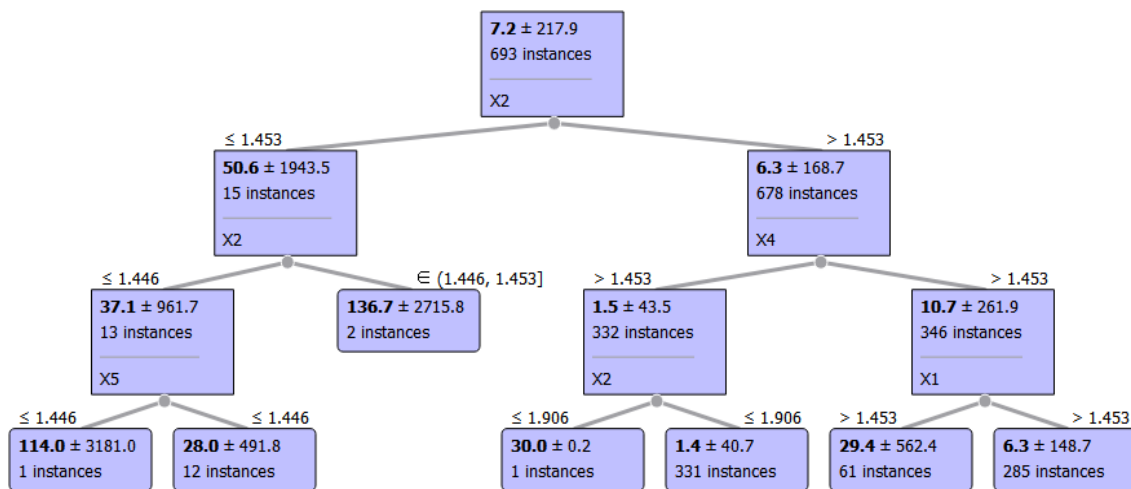


Figure 2. Decision tree regression figure

In Figure 3, the input contains 8 variables that affect rainfall. The output in the figure is a prediction of the amount of rainfall. In this research, the sigmoid activation function with the stochastic gradient descent (SGD) algorithm was used. The activation function is a function that defines how input is processed to produce output [25]. A neural network structure-training optimization algorithm known as SGD is available. In order to determine the variable's current value, the SGD algorithm requires that the gradient be calculated for each variable in the model.

The neural network, random forest regression, and multiple linear regression algorithms are tested with data testing. The following table depicts the approximate outcomes of the three algorithms in Table 1. In Table 1, columns 2, 3, and 4 are the predictions of the model, while column 1 is the target variable, that is rainfall. The calculation of the RMSE and MAE values was carried out, and the following values were obtained in Table 2.

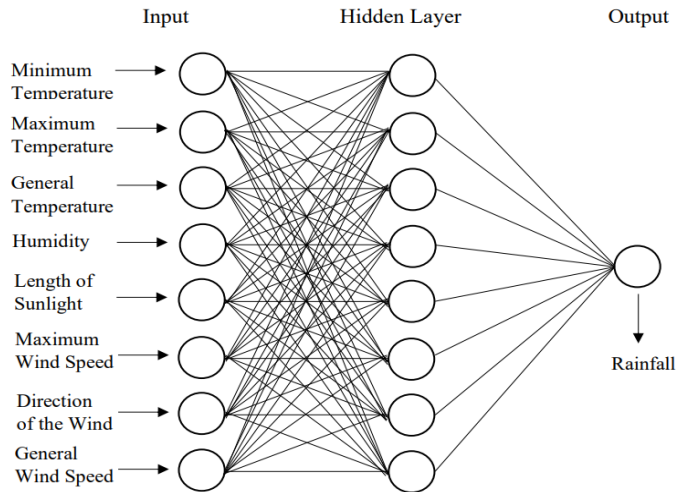


Figure 3. Neural network architecture for rainfall data

Table 1. Rainfall prediction table using MLR, RFR, and ANN methods

| Rainfall | MLR | RFR | ANN |
|----------|----------|----------|-----------|
| 46.0 | 16.9639 | 19.0807 | 21.482 |
| 24.0 | 7.68347 | 4.16978 | 4.61182 |
| 16.0 | 15.3333 | 8.67323 | 17.732 |
| 11.0 | 9.40366 | 7.22454 | 7.77768 |
| 10.0 | 11.2921 | 15.3348 | 11.4544 |
| 9.0 | 13.9039 | 8.67323 | 14.7805 |
| 6.0 | 13.8212 | 7.87055 | 13.449 |
| 6.0 | 8.13459 | 6.05605 | 4.10397 |
| 5.0 | 10.6928 | 7.88875 | 11.6505 |
| 5.0 | 7.13313 | 5.00855 | 4.00905 |
| 4.0 | 12.4314 | 8.67323 | 13.8802 |
| 2.0 | 13.1717 | 8.67323 | 12.9567 |
| 1.0 | 7.45517 | 7.22454 | 5.95172 |
| 0.0 | 9.62081 | 9.01207 | 11.4811 |
| 0.0 | 7.16078 | 5.67584 | 3.89938 |
| 0.0 | -8.01349 | 0.903419 | -0.30349 |
| 0.0 | -1.23285 | 1.74098 | 0.0363358 |

Table 2. The performance of machine learning model

| Model | RMSE | MAE |
|----------------------------|--------|-------|
| Artificial Neural Network | 13.055 | 6.621 |
| Random Forest Regression | 13.403 | 6.643 |
| Multiple Linear Regression | 13.475 | 7.843 |

3.2. Discussion

From the results of the multiple linear regression algorithm output, the regression line equation is obtained, namely (6). Prediction of the amount of rainfall can be searched by inputting the value of the variable x_1 to x_8 . While the output of the random forest regression algorithm in the form of a phytagorean forest in Figure 1 imply that tree number 2 is the best, because it has the shortest branch and the strongest color. Meanwhile, the neural network algorithm only has 1 output, because the prediction results obtained are numerical and continuous. In Table 2, based on RMSE and MAE calculations for the three algorithms, it is evident that the ANN algorithm has the smallest error value compared to other algorithms, that are 13,055 for RMSE and 6,621 for MAE. Table 1 is a table which is the result of prediction of rainfall from testing data. The numbers in the rainfall column determine the algorithm's error level in columns 2, 3, and 4. Since the ANN algorithm in Table 2 has better capabilities, column 4's forecast results are naturally closer to those in the rainfall column. For example, in row 1, the daily rainfall rate is 46 millimeters. According to the six types of rain that were discussed in the introduction, if the daily rainfall is 46 millimeters, medium rain occurs that day. The same thing also happened in column 4 with a total rainfall of 21.482 mm/day and the medium rain occurred. However, in columns 2 and 3 there was light rain, because it had a total rainfall of 16.9639 and 19.0807 mm/day.

4. CONCLUSION

Prediction of rainfall in the Semarang City is important to improve the preparedness of the people of the Semarang City in dealing with various types of rain, especially extreme rain that can cause natural disasters. The various machine learning algorithms used to predict rainfall are examined in this study. Using data from the Semarang City BMKG, algorithms for machine learning activities like multiple linear regression, random forest regression, and neural network regression were presented and tested. The input variables for the model of machine learning activities are selected environment features that are appropriate for rain prediction. The three algorithms' accuracy analogies have been established. With an error value of 13.055 for RMSE and 6.621 for MAE, the research demonstrates that neural network regression is a machine learning algorithm better suited to calculating daily rainfall amounts. It is anticipated that the future research will investigate machine learning algorithms, especially for predicting the amount of rainfall, can develop to increase the correlation value and minimize error values in the dataset.

ACKNOWLEDGEMENTS

This research was funded by the Development and Implementation Research Scheme, Diponegoro University in 2022.




REFERENCES

- [1] T. T. Le, B. T. Pham, H. B. Ly, A. Shirzadi, and L. M. Le, "Development of 48-hour precipitation forecasting model using nonlinear autoregressive neural network," *In CIGOS 2019, Innovation for Sustainable Infrastructure: Proceedings of the 5th International Conference on Geotechnics, Civil Engineering Works and Structures*, vol. 54, pp. 1191–1196, 2020, doi: 10.1007/978-981-15-0802-8_191.
- [2] Z. Guido *et al.*, "Farmer forecasts: Impacts of seasonal rainfall expectations on agricultural decision-making in Sub-Saharan Africa," *Climate Risk Management*, vol. 30, no. July, p. 100247, 2020, doi: 10.1016/j.crm.2020.100247.
- [3] J. Han, M. Kamber, and J. Pei, "Data mining: Data mining concepts and techniques," *In 2013 International Conference on Machine Intelligence and Research Advancement*, IEEE, 2014, pp. 203–207, doi: 10.1109/ICMIRA.2013.45.
- [4] D. T. Larose, "Data mining methods and models," *John Wiley and Sons*, vol. 21, no. 1, 2006, doi: 10.1037/014836.
- [5] C. M. Liyew and H. A. Melese, "Machine learning techniques to predict daily rainfall amount," *Journal Big Data*, vol. 8, no. 1, 2021, doi: 10.1186/s40537-021-00545-4.
- [6] W. Suparta and A. A. Samah, "Rainfall prediction by using ANFIS times series technique in South Tangerang, Indonesia," *Geodesy and Geodynamics*, vol. 11, no. 6, pp. 411–417, 2020, doi: 10.1016/j.geog.2020.08.001.
- [7] R. N. Rachmawati, I. Sungkawa, and A. Rahayu, "Extreme rainfall prediction using Bayesian quantile regression in statistical downscaling modeling," *Procedia Computer Science*, vol. 157, pp. 406–413, 2019, doi: 10.1016/j.procs.2019.08.232.
- [8] A. Y. Barrera-Animas, L. O. Oyedele, M. Bilal, T. D. Akinosho, J. M. D. Delgado, and L. A. Akanbi, "Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting," *Machine Learning with Applications*, vol. 7, no. October 2021, p. 100204, 2022, doi: 10.1016/j.mlwa.2021.100204.
- [9] E. G. Wahyuni, L. M. F. Fauzan, F. Abriyani, N. F. Muchlis, and M. Ulfa, "Rainfall prediction with backpropagation method," *In Journal of Physics: Conference Series*, vol. 983, no. 1, 2018, doi: 10.1088/1742-6596/983/1/012059.
- [10] C. Fan, M. Chen, X. Wang, J. Wang, and B. Huang, "A review on data preprocessing techniques toward efficient and reliable knowledge discovery from building operational data," *Frontiers in Energy Research*, vol. 9, no. March, pp. 1–17, 2021, doi: 10.3389/feenrg.2021.652801.
- [11] S. A. Alasadi and W. S. Bhaya, "Review of data preprocessing techniques.pdf," *Journal of Engineering and Applied Sciences*, vol. 12, no. 16, pp. 4102–4107, 2017.
- [12] A. P. Joshi and B. V. Patel, "Data preprocessing: the techniques for preparing clean and quality data for data analytics process," *Oriental Journal of Computer Science and Technology*, vol. 13, no. 0203, pp. 78–81, 2021, doi: 10.13005/ojst13.0203.03.
- [13] S. Mansyur, "Multiple regression analysis on influence factors of household cooking fuels in Indonesia," *Conserve: Journal of Energy and Environmental Studies*, vol. 1, no. 1, pp. 9–19, 2017, doi: 10.30588/cjees.v1i1.249.
- [14] M. Yanto, S. Sanjaya, Yulasmi, D. Guswandi, and S. Arlis, "Implementation multiple linear regression in neural network predict gold price," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 22, no. 3, pp. 1635–1642, 2021, doi: 10.11591/ijeecs.v22.i3.pp1635-1642.
- [15] G. K. Uyanik and N. Güler, "A study on multiple linear regression analysis," *Procedia-Social and Behavioral Sciences*, vol. 106, pp. 234–240, 2013, doi: 10.1016/j.sbspro.2013.12.027.
- [16] V. Svetnik, A. Liaw, C. Tong, J. C. Culberson, R. P. Sheridan, and B. P. Feuston, "Random forest: a classification and regression tool for compound classification and QSAR modeling," *Journal of chemical information and computer sciences*, vol. 43, no. 6, pp. 1947–1958, 2003, doi: 10.1021/ci034160g.
- [17] N. Ganesh, P. Jain, A. Choudhury, P. Dutta, K. Kalita, and P. Barsocchi, "Random forest regression-based machine learning model for accurate estimation of fluid flow in curved pipes," *Processes*, vol. 9, no. 11, 2021, doi: 10.3390/pr9112095.
- [18] R. Couronné, P. Probst, and A. L. Boulesteix, "Random forest versus logistic regression: A large-scale benchmark experiment," *BMC Bioinformatics*, vol. 19, no. 1, pp. 1–14, 2018, doi: 10.1186/s12859-018-2264-5.
- [19] K. Thaker, V. Charles, A. Pant, and T. Gherman, "A DEA and random forest regression approach to studying bank efficiency and corporate governance," *Journal of the Operational Research Society*, vol. 73, no. 6, pp. 1258–1277, 2022, doi: 10.1080/01605682.2021.1907239.
- [20] M. S. Khrisat, A. Alabadi, S. Khawatreh, M. O. Al-Dwairi, and Z. A. Alqadi, "Autoregressive prediction analysis using machine deep learning," *Indonesian Journal of Electrical Engineering and Computer Science (IJECS)*, vol. 27, no. 3, pp. 1509–1516, 2022, doi: 10.11591/ijeecs.v27.i3.pp1509-1516.
- [21] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad, "State-of-the-art in artificial neural network applications: A survey," *Heliyon*, vol. 4, no. 11, p. e00938, 2018, doi: 10.1016/j.heliyon.2018.e00938.




- [22] M. Şahin and R. Erol, "A comparative study of neural networks and ANFIS for forecasting attendance rate of soccer games," *Mathematical and computational applications*, vol. 22, no. 4, p. 43, 2017, doi: 10.3390/mca22040043.
- [23] S. Han, J. Xu, M. Yan, and Z. Liu, "Using multiple linear regression and BP neural network to predict critical meteorological conditions of expressway bridge pavement icing," *PLoS One*, vol. 17, no. 2 February, pp. 1–15, 2022, doi: 10.1371/journal.pone.0263539.
- [24] I. N. Y. Saputra, S. Saadah, and P. E. Yunanto, "Analysis of random forest, multiple regression, and backpropagation methods in predicting apartment price index in Indonesia," *Jurnal Ilmiah Teknik Elektro Komputer dan Informatika (JITEK)*, vol. 7, no. 2, p. 238, 2021, doi: 10.26555/jiteki.v7i2.20997.
- [25] S. Sharma, S. Sharma, and A. Athaiya, "Activation Functions in Neural Networks," *Int. J. Eng. Appl. Sci. Technol.*, vol. 04, no. 12, pp. 310–316, 2020, doi: 10.33564/ijeast.2020.v04i12.054.

BIOGRAPHIES OF AUTHORS






Carissa Devina Usman    born in Semarang, August 17, 1998. She is student in master degree at the Diponegoro University, Indonesia, majoring in Mathematics in 2021. She can be contacted at email: carissa.devina2@gmail.com.






Aris Puji Widodo    born in Grobogan, April 01, 1974. He is student in doctoral education at the Gajah Mada University, Indonesia, majoring in Computer Sciences in 2015. Research area in the field of government interoperaility and software engineering. He can be contacted at email: arispw@gmail.com.



Kusworo Adi    born in Temanggung, March 17, 1972. He is undergraduate education at Diponegoro University, Indonesia, majoring in physics from 1991 to 1996. Then the master's education at the Bandung Institute of Technology, Indonesia, majoring in electrical engineering from 2000 to 2002. Doctoral education at the Bandung Institute of Technology, Indonesia, majoring in electrical engineering in 2005 to 2010. Research area in the field of instrumentation and image processing. Several publications received in reputable international journals, especially in the field of instrumentation and image processing. The research being carried out now is about transportation safety by utilizing the development of instrumentation and image processing. Dr. Kusworo Adi is currently a member of the IEEE. He can be contacted at email: kusworoadi@gmail.com



Rahmat Gernowo    born in Kebumen, November 23, 1965. He is student in doctoral education at the Bandung Institute Technology, Indonesia, majoring in Geo Physics in 2009. Research area in the field of geo physics and atmosfer science. He can be contacted at email: gernowo@yahoo.com.