

# Analysis and Prediction of Turbidity Level of Water Based on Ammonia Substance Using Random Forest and K-Nearest Neighbor

Teguh Herlambang<sup>1,2\*</sup>, Mohd Sanusi Azmi<sup>3</sup>, Zuraini Binti Othman<sup>4</sup>,  
Mochammad Romli Arief<sup>5</sup>

<sup>1</sup>Department of Information System, FEBTD, Universitas Nahdlatul Ulama Surabaya, Indonesia  
Jemursari St, No. 51-57 (Komplek RSI Jemursari), Surabaya, 60237, East Java, Indonesia

<sup>2</sup>Research Group of Intelligence System and Financial Technology, Universitas Nahdlatul Ulama  
Surabaya, Indonesia

Jemursari St, No. 51-57 (Komplek RSI Jemursari), Surabaya, 60237, East Java, Indonesia

<sup>3</sup>Department of Software Engineering, Fakulti Teknologi Maklumat dan Komunikasi, Universiti  
Teknikal Malaysia Melaka, Malaysia

Hang Tuah Jaya St, 76100 Durian Tunggal, Melaka, Malaysia

<sup>4</sup>Department of Diploma Studies, Fakulti Teknologi Maklumat dan Komunikasi, Universiti  
Teknikal Malaysia Melaka, Malaysia

Hang Tuah Jaya St, 76100 Durian Tunggal, Melaka, Malaysia.

<sup>5</sup>Undergraduate of Program Information System, FEBTD, Universitas Nahdlatul Ulama  
Surabaya, Indonesia

Jemursari St, No. 51-57 (Komplek RSI Jemursari), Surabaya, 60237, East Java, Indonesia

\*E-mail Correspondence Author: [teguh@unusa.ac.id](mailto:teguh@unusa.ac.id)

## Abstract

Water is the primary need for human survival. The need for clean water is very important, both in a household scale and in an industrial scale. The clean water used and consumed by the community comes from river water processed and distributed by the Regional Drinking Water Company (PDAM). Water conditions before being treated and distributed contain various harmful substances if not purified, one of which is ammonia. Currently, with the progression of information technology, particularly in the field of machine learning and data analysis, the process of predicting the content of ammonia substances in water is becoming increasingly facilitated. Machine learning has provided a number of scientific prediction methods that can be used. In this research, the methods used were Random Forest and K-Nearest Neighbor (KNN) methods are used to predict turbidity level of water in PDAM Surya Sembada Surabaya. This research aims to compare robustness and precision of prediction methods. The Random Forest method produced the best prediction error value of 0.0934, while the K-Nearest Neighbor (KNN) method produced the best prediction error value of 0.0918.

**Keywords:** KNN, Machine Learning, Prediction, Random Forest, Turbidity, Water.

 <https://doi.org/10.30598/parameterv4i1pp55-62>



This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution-ShareAlike 4.0 International License](#).

## 1. INTRODUCTION

Support activities for both home needs and those on a wider scale, such as for industrial needs. Before water can be used by human, a clean water treatment process is needed to ensure about safety for daily uses [1]. The availability of clean water is also inseparable from its main source, the river. Rivers has a vital role in maintaining the availability of clean water raw materials needed by the community. However, with the growth of population accompanied by an increase in activities, the quality of clean water raw materials is decreasing due to water pollution. Water pollution is the condition of changing to the natural characteristic of water in the human environment [2]. Pollution in the water caused by mixing with hazardous substances, such as ammonia. Ammonia is an inorganic substance that originating from nitrogen-hydrogen compounds from nature and its formed by humans [3].

PDAM Surya Sembada is one of the business units owned by Surabaya City Government (Pemkot) whose core activity is providing clean water to the customers such as citizens, business unit, and industry, starting from processes to the distribution. PDAM Surya Sembada always improves and notices towards quality of water through several processing [4]. This water treatment is very important because it must be safe before consumed by customers and to maintain their satisfaction. In short, customer satisfaction is defined as what consumers desire to fulfill their needs, such as goods or service [5].

With technological advance in the field of machine learning now, these condition can be handled by application of scientifically predictive based-system. Machine Learning is a subset of Artificial Intelligence (AI) [6]. Machine Learning (ML) is the study of computer algorithms that can recognize pattern in data, with purposed to be support of decision making process with less human interaction [7]. This application is important with the aim of getting accurate decision based on historical data. In this research, Random Forest and K-Nearest Neighbor (KNN) methods were used to compare the performance of both. In this research, also tested two methods that use different approaches to make predictions. Random Forest uses an ensemble tree approach, and KNN uses a distance approach. These two differences are characterized by the Random Forest and KNN methods compared to other methods such as SVM or Neural Network.

Random Forest is is one of the regression models based on supervised learning techniques using features from historical data. This method proposed by Breiman [8]. Random Forest has characteristic which organizing the features and target using tree-based model [9]. KNN basically used for classification, but it can also used for regression, which regression result is obtained from average of k closest neighbors [10]. This method is categorized as nonparametric classification method and well-known as easiest algorithm [11]. These two methods are widely used for in many previous predictions and classification-based studies because they are reliable and resistant to outlier data.

## 2. RESEARCH METHOD

### 2.1 Literature Review

Below in Table 1 are references to prior research that are viewed as relevant to current research.

**Table 1. References of Prior Research**

Year	Title of Research	Description
2024	Prediction of River Water Quality Using KNN optimized with PSO [12].	KNN optimized by PSO method reached best/highest accuracy of 95.8% better than KNN without optimization that only reached 77.9%
2023	Classification Quality and Prediction Soil Water in DKI Jakarta Using Naive Bayes Algorithm [13].	Naive Bayes algorithm reached best/highest accuracy of 84.36% with splitting percentage of 80:20
2023	Prediction of River Water Quality using Machine Learning Methods : Ciliwung River Study Case [14].	Artificial Neural Network method reached best/highest accuracy of 94.6% compares with SVM, Random Forest, and Naive Bayes.
2023	Water Quality Prediction Using Machine Learning Models Based on Grid Search Method [15].	KNN method managed to achieve value of MSE of 0.0002 and R-Square value of 98.2%.
2022	Analysis of Gradient Boosting, Adaboost, Catboost Algorithms in Water Quality Classification [16].	Catboost method achieved best accuracy of 68% than Gradient Boosting and Adaboost method
2020	Water Quality System Prediction in PDAM Tirta Raharja Using K-Nearest Neighbors (KNN) [17].	KNN method reached highest accuracy above 90% when classifying potable and not potable water.

From a number of prior researches above that used as reference, these two methods are widely used to solve case studies on classification and some were done with additional approaches using other methods. Both methods exhibit satisfying results in classification with an accuracy rate above 80%, therefore, in this research, both methods are tested for their reliability in making predictions by maximizing the available parameters.

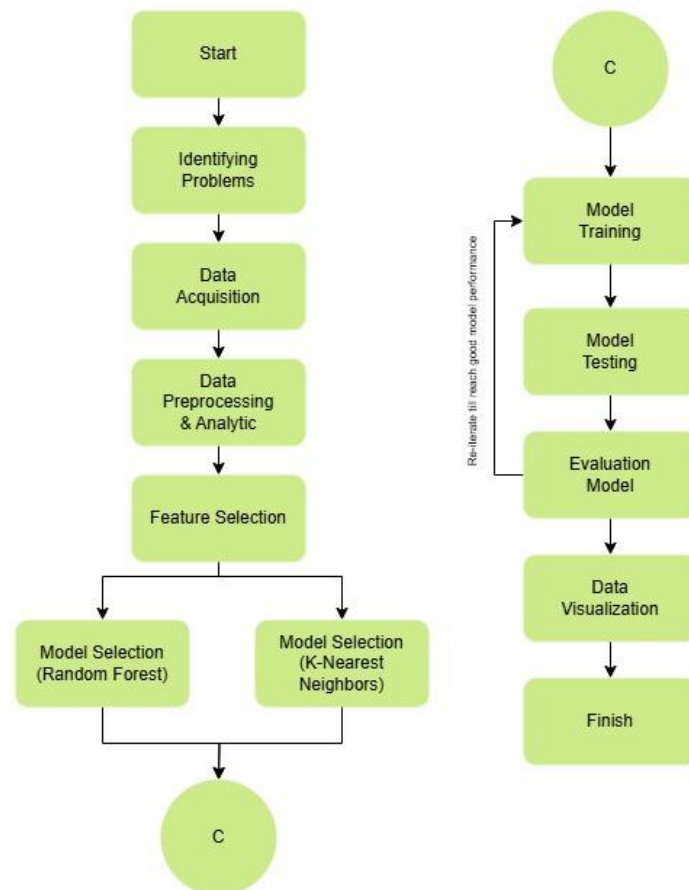
## 2.2 Type of Research

The data used in this research obtained from primary data of PDAM Surya Sembada Surabaya. The dataset consists of 7 columns and 2066 rows with a data range of 01/02/2019 to 01/11/2024. The dataset consists of measurements taken daily by relevant personnel. The data were analyzed and processed using Python to find the types, patterns, and possibility of correlation between variables. All substances have an effect on the suitability of water for consumption. Below is **Table 2** containing the historical data from the measurement of substance levels in water at PDAM Surya Sembada Surabaya.

**Table 2. Dataset**

Date	Ammonia	Turbidity	Ph	Residual Chlor	Temp	Organic Substances
2019-02-01	0.2783	0.73	6.81	0.45	27.5	7.27
2019-02-04	0.1449	0.58	7.32	1	28	8.9
2019-02-06	0.1449	0.81	6.96	1	27	7.06
2019-02-07	0.1449	0.28	6.85	1	26	9.82
2019-02-08	0.1449	0.55	7.04	1	27	7.52
2019-02-11	0.0992	0.4	7.13	1	27	6.77
2019-02-12	0.0992	0.43	7.15	0.71	26.8	6.02
2019-02-13	0.0992	0.46	6.95	1.03	26	5.52
2019-02-14	0.0992	0.66	7.12	0.8	27	6.75
2019-02-15	0.0992	0,36	7.04	0.84	29	7.13
2019-02-18	0.1602	0.67	7.05	0.83	29	8.98
2019-02-19	0.1602	0.41	6.96	0.46	28.5	7.44
2019-02-20	0.1602	0.65	7.05	0.45	28	6.26
...	...	...	...	...	...	...
2024-11-01	0.19	1.16	6.83	0.11	29	6.34

After acquiring and exploring the data, the next sequence is executing the stages of the research methodology as presented in [Figure 1](#) below.



**Figure 1. Research Methodology**

From the above research methodology flow, the detailed steps at each stage can be explained. The following is an explanation of each stage in the research methodology diagram above.

### 2.3 Problem Identification

This research raises case study aim to obtain prediction result of ammonia substance on water turbidity water managed by PDAM Surya Sembada Surabaya and also to assist stakeholders in case of decision-making.

### 2.4 Data Preprocessing and Analytics

The dataset used in this research is obtained from PDAM Surya Sembada Surabaya with a time span of 01/02/2019 to 01/11/2024. Data preprocessing and analytics are basically used to identify initial condition of dataset that obtained from source. On this phase, the dataset is identified starting from data type, correlation, and missing value that possible occurred. Also, statistical analysis needs to be done to avoid poor prediction result.

### 2.5 Feature Selection

In many prediction cases with numerical data, the way to determine the features used as independent and dependent variables is to use analysis techniques based on the Pearson Product rule [18]. The Pearson Product rule or function produces a relationship value between the independent variable and the dependent variable with a value in the range of 0-1. If the value is close to 1, it can be concluded that the variable has a positive correlation and vice versa. Below in Equation (1) is function of Pearson Product.

$$r_{xy} = \frac{n \sum XY - (X)(Y)}{\sqrt{n \sum X^2 - \sum X^2} \sqrt{n \sum Y^2 - \sum Y^2}} \quad (1)$$

with:

- $r_{xy}$  : relationship coefficient
- $n$  : number of samples used
- $X$  : total score of question
- $Y$  : sum of total scores

### 2.6 Model Selection

The Random Forest method is development of Decision Tree. Decision Tree is ensemble tree based method that has root node [19]. Below on Equation (2-3) are the steps to implement the Random Forest. The working principle Random Forest method begins with create bootstrap sampling which randomly selected from data. This process applied to construct different training dataset for each decision tree.

$$D = \{(x_1, y_1), (x_2, y_2), (x_3, y_3), \dots (x_N, y_N)\} \quad (2)$$

$$D^* = \{(x_1^*, y_1^*), (x_2^*, y_2^*), (x_3^*, y_3^*), \dots (x_N^*, y_N^*)\} \quad (3)$$

In both equations above,  $D$  represents the training dataset and  $x_1^*, y_1^*$  values represent the data points selected from bootstrap sampling. Then, the best split points of each branch of decision tree is selected among  $m$  randomly selected variables. The split point represented by function on Equation (4) below.

$$\operatorname{argmin}_{j,s} = \sum_{i:x_i^j \leq s} (y_i - \hat{y}_{\text{left}})^2 + \sum_{i:x_i^j > s} (y_i - \hat{y}_{\text{right}})^2 \quad (4)$$

In [Equation \(4\)](#),  $\hat{y}_{left}$  and  $\hat{y}_{right}$  are indicating the average target values of left and right subnodes, respectively. Then, to create final prediction, all of the results obtained by decision tree method will be calculated to find average value from all trees. Below in [Equation \(5\)](#) is the function for final prediction.

$$\hat{f}(x) = \frac{1}{B} \sum_{b=1}^B \hat{f}_b(x) \quad (5)$$

KNN method is a method that works using closest distance between training data and testing data to find k objects [\[20\]](#). This method uses the function of Euclidean Distance. Below on [Equation \(6\)](#) is the function of the KNN method:

$$D = \sqrt{(x_2 - x_1)^2 - (y_2 - y_1)^2} \quad (6)$$

with:

$D$  : distance  
 $x$  : data Sample  
 $y$  : testing data

## 2.7 Evaluation Metrics

At evaluation phase, the model trained and tested is calculated for accuracy based on the resulting error value. This research uses the Root Mean Square Error (RMSE) method to calculate the error value generated by the model. The function of the Root Mean Square Error (RMSE) is on [Equation \(7\)](#) below.

$$RMSE = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n}} \quad (7)$$

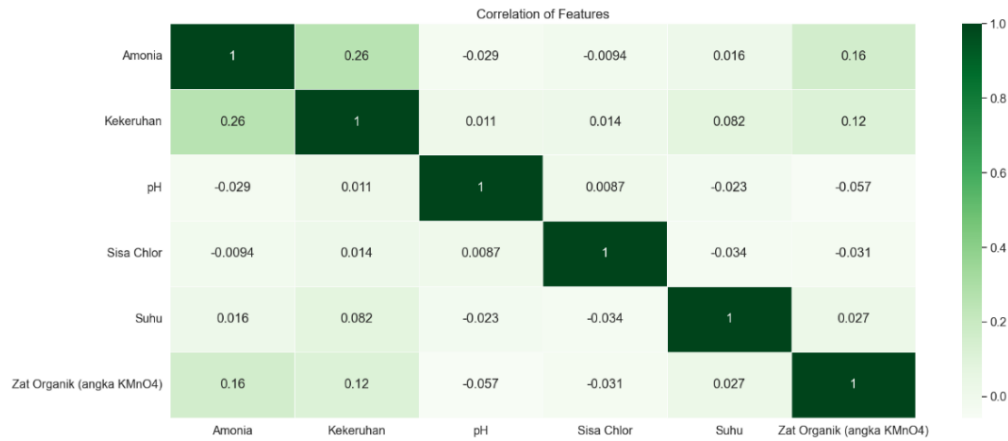
with:

$n$  : quantity of data  
 $y_i$  : actual value at the i-th data  
 $\hat{y}_i$  : predicted value at the i-th data

## 3 RESULT AND DISCUSSION

### 3.2 Descriptive of Independent and Dependent Variables

Based on result of Pearson correlation, then obtained result value to determine variable independent and dependent. Below in [Figure 2](#) is plot of correlation based on Pearson.



**Figure 2. Pearson Correlation**

It can be seen in [Figure 2](#), variables ammonia and turbidity have correlation value of 0.26. Both of them have positive relationship but at weak level, but not weaker than the correlation values of other variables. If concerned, the correlation value of other independent variables does not exceed 0.2. So, the conclusion is variable ammonia as a independent variable and turbidity as dependent variable.

### 3.3 Hyperparameters Tuning and Data Splitting

In this phase, static parameter adjustment are implemented for Random Forest and K-Nearest Neighbor (KNN) before being used as prediction model. Below in [Table 3](#) and [Table 4](#) can be seen details of hyperparameters tuning and data splitting.

**Table 3. Hyperparameters Tuning**

Random Forest	
Name of Hyperparameter	Value
n_estimator	250
Criterion	squared error
max_depth	70
min_samples_split	25
min_samples_leaf	20
K-Nearest Neighbor (KNN)	
n_neighbor	5

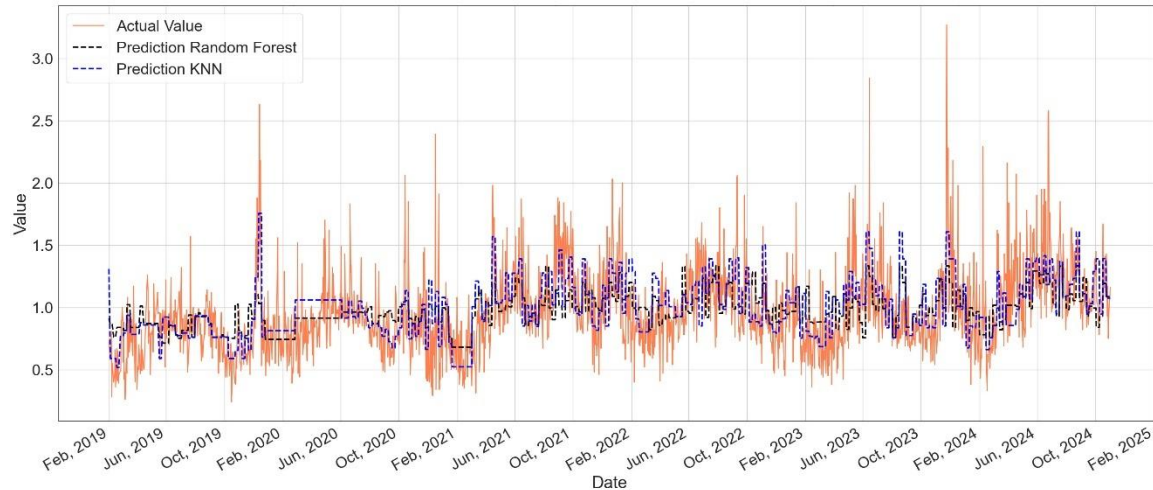
**Table 4. Data Splitting**

Percentage	Number of Training Data	Number of Testing Data
65% : 35%	1343	723
75% : 25%	1549	517
85% : 15%	1756	310

In the [Table 3](#), it can be seen that the value of each hyperparameter of both methods. In this research, the simulations of prediction are conducted by using static hyperparameter. In [Table 4](#) is the details of data splitting percentage which applied by three different splits. It aims to compare the result obtained by each method in the case of prediction which influenced by the number of training data.

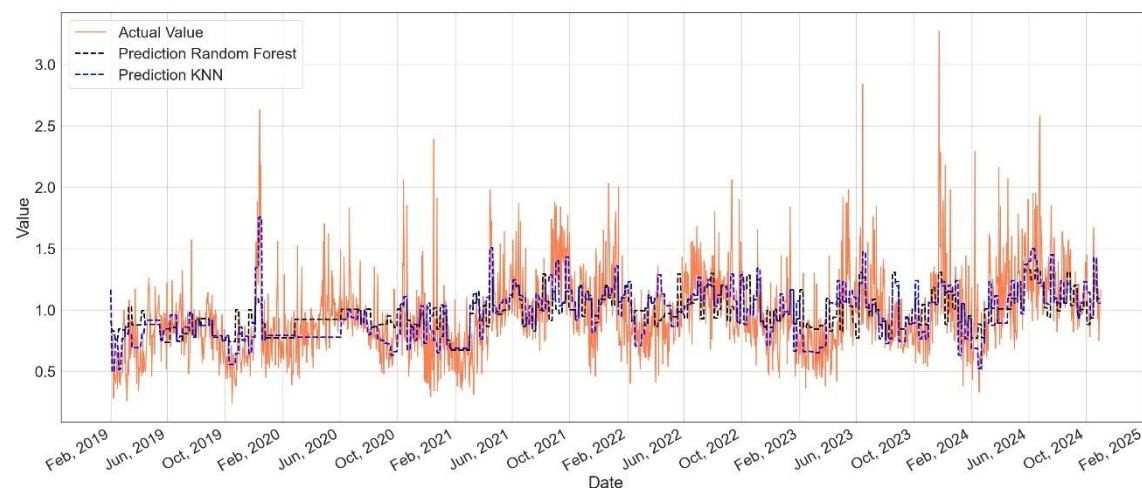
### 3.3 Simulation Result

This research implement two methods of machine learning, namely Random Forest and KNN to predict the level of water turbidity based on ammonia content. The prediction results based on the method and difference in the composition of training data and testing data with the actual values as shown in [Figure 3](#) till [Figure 5](#). [Figure 3](#) is the simulation results of the Random Forest and KNN methods with 65% train data and 35% test data.



**Figure 3.** First Simulation Plot of Random Forest dan KNN (using percentage split of 65%:35%)

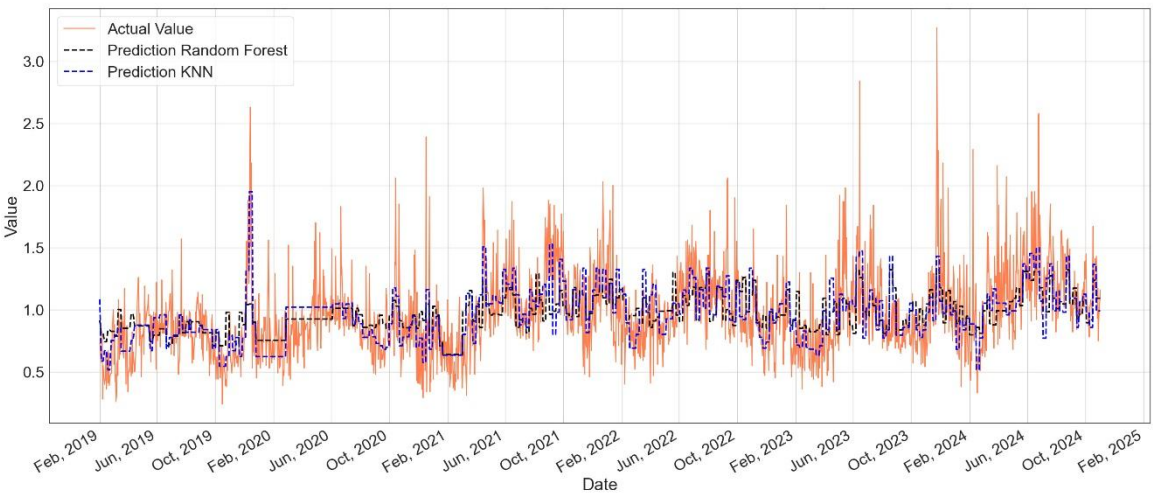
In the first simulation, it can be seen on [Figure 3](#) that the results of the Random Forest method, manage to achieved prediction output close to the actual value (black line). The resulting error value (RMSE) is 0.0967. Then, the simulation results of the KNN method successfully produced prediction output close to the actual value (blue line). The resulting error value (RMSE) of KNN is 0.0987. The output of the error value (RMSE) produced by the two methods is not too wide, that is 0.0020.



**Figure 4.** Second Simulation Plot of Random Forest dan KNN (using percentage split of 75%:25%)

In the second simulation, it can be seen on [Figure 4](#) that the results of the Random Forest method manage to achieved prediction output close to the actual value (black line). The resulting error value (RMSE) is 0.0934. Then, the simulation results of the KNN method successfully produced prediction output close to the actual value (blue line). The

resulting error value (RMSE) is 0.0918. The difference output of the error value (RMSE) produced by the two methods is 0.0016.



**Figure 5.** Third Simulation Plot of Random Forest dan KNN (using percentage split of 85%:15%)

In the third simulation, it can be seen on [Figure 5](#) that the results of the Random Forest method successfully show prediction outputs close to the actual value (black line). The resulting error value (RMSE) is 0.0964. Then, the simulation results of the KNN method show promising prediction also with outputs close to the actual value (blue line). Both methods also have the same prediction line pattern. The resulting error value (RMSE) is 0.0952. The output of the error value (RMSE) produced by the two methods shows a slight difference, that is, 0.0012. The following is a recapitulation table of the simulation results of the Random Forest method and KNN methods as in the [Table 5](#) below.

**Table 5.** Comparison RMSE Value

Percentage of Train Data and Test Data	RMSE Value of Random Forest	RMSE Value of KNN
65% : 35%	0.0967	0.0987
75% : 25%	0.0934	0.0918
85% : 15%	0.0964	0.0952

The [Table 5](#) above shows the simulation results generated by the Random Forest method and KNN methods from the first simulation to the third simulation. In the second simulation results with 75% training data and 25% testing data, Random Forest method produced the lowest RMSE value of the entire simulation. This method successfully produced the best RMSE value of 0.0934, while the KNN method also produced the best error value in the second simulation with a value of 0.0918. This proves that both methods are still capable of predicting data that has significant up and down conditions. Then, based on prior research, it was proven that these two methods are quite reliable not only for classification but also for prediction. Then, also shows in this research, both Random Forest and KNN methods show good performance, although at each simulation stage it can be seen that both methods produce dynamic RMSE values. Based on the selection of only ammonia as the independent variable, it has been proven to minimize prediction error values. This is because it has a significant effect in prediction models. However, the resulting error value (RMSE) is still good, which is below 1. By promising results, it can

be summarized that both methods have satisfied the research objectives and develop a renewal of prior research.

#### 4. CONCLUSION

Based on the results of the simulations conducted, it can be concluded that the results of the first to third simulations using the Random Forest method managed to get the best prediction error value (RMSE) in the second simulation with a splitting of 75% training data and 25% testing data, that is, 0.0934. While the KNN method managed to get the best prediction error value (RMSE) of 0.0918 also in the second simulation with only ammonia as independent variable and turbidity as dependent variable. These results prove that the Random Forest and KNN methods provide good and consistent prediction results also has fulfilled the objective of this research. The result obtained by this research also can be used as a reference by PDAM stakeholders making decision. In addition, both methods can be recommended for use in further research with different prediction cases or using other optimizer methods such as PSO, GA, etc.

#### Acknowledgments

The authors would like thank to Centre for Research and Innovation Management of Universiti Teknikal Malaysia Melaka (UTeM) for sponsoring this work under the Tabung Penerbitan CRIM UTeM.

#### Funding Information

This article is sponsored by Centre for Research and Innovation Management of Universiti Teknikal Malaysia Melaka (UTeM) under the Tabung Penerbitan CRIM UTeM.

#### Conflict Of Interest Statement

The authors declare that there are no conflicts of interest in the writing of this article.

#### Data Availability

The dataset used in this research is not open access.

#### REFERENCES

- [1] N. Yahyaoui, M. Neji, M. Kallel, A. Wali, and S. Hajji, "A comparative approach of ML algorithms to rank irrigation water quality: case of Oriental-Coast shallow aquifer in Cap-Bon, northeastern of Tunisia," *Modeling Earth Systems and Environment*, vol. 9, no. 3, pp. 3733–3746, Jan. 2023, doi: 10.1007/s40808-023-01694-6
- [2] M. Akhlaq, A. Ellahi, R. Niaz, M. Khan, S. Sh. Sammen, and M. Scholz, "Comparative analysis of machine learning algorithms for water quality prediction," *Tellus a Dynamic Meteorology and Oceanography*, vol. 76, no. 1, pp. 177–192, Jan. 2024, doi: 10.16993/tellusa.4069.
- [3] F. Alkindi, R. Budiono, and F. Al-Islami, "PENGUJIAN ANALISIS KADAR AMONIA DALAM AIR SUNGAI DI DAERAH INDUSTRI SIER SURABAYA MENGGUNAKAN METODE FENAT SECARA SPEKTROFOTOMETRI VISIBLE", *medfarm*, vol. 12, no. 2, pp. 181–189, Dec. 2023. <https://doi.org/10.48191/medfarm.v12i2.234>
- [4] A. A. Nofitri and A. D. Radityaningrum, "Evaluasi Kinerja Pengolahan Air Bersih IPAM Karangpilang III PDAM Kota Surabaya," *Prosiding Seminar Nasional Sains dan Teknologi Terapan*, Nov. 2023.

- [5] Muhammad Niqey Dzi Qolbin Saliem and Enny Aryanny, "Penerapan Metode Crm (Customer Relationshipmanagement) Pada Divisi Pelayanan Teknis Barat Di Pdam Surya Sembada Kota Surabaya", *JTMEI*, vol. 3, no. 1, pp. 126–135, Dec. 2023. <https://doi.org/10.55606/jtmei.v3i1.3258>
- [6] T. Madan, S. Sagar, and D. Virmani, "Air Quality Prediction using Machine Learning Algorithms –A Review," *2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN)*, Dec. 2020, doi: [10.1109/icacccn51052.2020.9362912](https://doi.org/10.1109/icacccn51052.2020.9362912).
- [7] A. Maslan and A. A. . Fajrin, "Predicting the Spread of Covid-19 in Indonesia Using Machine Learning Models", *SNISTEK*, vol. 4, pp. 441–446, Jan. 2022.
- [8] S. Kouadri, A. Elbeltagi, A. R. Md. T. Islam, and S. Kateb, "Performance of machine learning methods in predicting water quality index based on irregular data set: application on Illizi region (Algerian southeast)," *Applied Water Science*, vol. 11, no. 12, Nov. 2021, doi: [10.1007/s13201-021-01528-9](https://doi.org/10.1007/s13201-021-01528-9).
- [9] R. M. X. Wu *et al.*, "Comparative study of ten machine learning algorithms for short-term forecasting in gas warning systems," *Scientific Reports*, vol. 14, no. 1, Sep. 2024, doi: 10.1038/s41598-024-67283-4.
- [10] N. Mustakim, N. Rosdina, D. Ramadhani, M. Afdal, and M. Rahmawita, "The Classification Status of River Water Quality in Riau Province Using Modified K-Nearest Neighbor Algorithm with STORET Modeling and Water Pollution Index," *Journal of Physics Conference Series*, vol. 1783, no. 1, p. 012020, Feb. 2021, doi: 10.1088/1742-6596/1783/1/012020.
- [11] N. Zamri *et al.*, "River quality classification using different distances in k-nearest neighbors algorithm," *Procedia Computer Science*, vol. 204, pp. 180–186, Jan. 2022, doi: [10.1016/j.procs.2022.08.022](https://doi.org/10.1016/j.procs.2022.08.022).
- [12] I. N. Simbolon, "PREDIKSI KUALITAS AIR SUNGAI DI JAKARTA MENGGUNAKAN KNN YANG DIOPTIMALISASI DENGAN PSO", *JITET*, vol. 12, no. 2, Apr. 2024. <https://doi.org/10.23960/jitet.v12i2.4191>
- [13] A. Fathiarahma, N. Sulistiyowati, T. . Ridwan, and A. Voutama, "Klasifikasi Kualitas dan Prediksi Kondisi Air Tanah di DKI Jakarta Menggunakan Algoritma Naïve Bayes", *INSYST*, vol. 5, no. 2, pp. 73–82, Oct. 2023. <https://doi.org/10.52985/insyst.v5i2.325>
- [14] M. HAEKAL and W. C. WIBOWO, "Prediksi Kualitas Air Sungai Menggunakan Metode Pembelajaran Mesin: Studi Kasus Sungai Ciliwung: Prediction of River Water Quality Using Machine Learning Methods: Ciliwung River Case Study ", *j. teknologi lingkungan*, vol. 24, no. 2, pp. 273–282, Jul. 2023. <https://doi.org/10.55981/jtl.2023.795>
- [15] M. Y. Shams, A. M. Elshewey, E.-S. M. El-Kenawy, A. Ibrahim, F. M. Talaat, and Z. Tarek, "Water quality prediction using machine learning models based on grid search method," *Multimedia Tools and Applications*, vol. 83, no. 12, pp. 35307–35334, Sep. 2023, doi: [10.1007/s11042-023-16737-4](https://doi.org/10.1007/s11042-023-16737-4).
- [16] T. Z. Jasman, M. A. Fadhlullah, A. L. Pratama, and Rismayani, "Analysis of Gradient Boosting, Adaboost, Catboost Algorithms in Water Quality Classification," *Jurnal Teknik Informatika dan Sistem Informasi*, vol. 8, no. 2, Aug. 7AD, doi: <http://dx.doi.org/10.28932/jutisi.v8i2.4906>.
- [17] Rahandanu Rachmat, Yulison Herry Chrisnanto, & Fajri Rakhmat Umbara. (2020). Sistem Prediksi Mutu Air Di Perusahaan Daerah Air Minum Tirta Raharja Menggunakan K – Nearest Neighbors (K – NN). *Prosiding SISFOTEK*, 4(1), 189 - 193. Retrieved from <https://www.seminar.iaii.or.id/index.php/SISFOTEK/article/view/211>
- [18] J. Liu *et al.*, "Accurate Prediction scheme of water quality in smart mariculture with deep Bi-S-SRU learning network," *IEEE Access*, vol. 8, pp. 24784–24798, Jan. 2020, doi: [10.1109/access.2020.2971253](https://doi.org/10.1109/access.2020.2971253).
- [19] M. B. Nafouanti, J. Li, E. E. Nyakilla, G. C. Mwakipunda, and A. Mulashani, "A novel hybrid random forest linear model approach for forecasting groundwater fluoride contamination,"

*Environmental Science and Pollution Research*, vol. 30, no. 17, pp. 50661–50674, Feb. 2023, doi: 10.1007/s11356-023-25886-w.

- [20] A. Faisol, Nusyirwan, and A. Nadwah, "IMPLEMENTASI WEIGHTED K-NEAREST NEIGHBOR UNTUK PERAMALAN DATA DERET WAKTU," *Jurnal Riset dan Aplikasi Matematika*, vol. 5, no. 2, doi: <https://doi.org/10.26740/jram.v5n2.p111-117>.