Water Quality Classification Using K-Nearest Neighbor Algorithm

Gabriella Puteri Ayu Handiaz Mawarti[1], Eko Hari Rachmawanto[2], Noorayisahbe Bt Mohd Yaacob[3]
[1,2]Study Program in Informatics Engineering, Faculty of Computer Science, University of Dian Nuswantoro, Indonesia
[3]University Kebangsaan Malaysia, Malaysia

Article Info:

Keywords:

Accuracy
Classification
K-Nearest Neighbor
Water quality

ABSTRACT

Water quality is defined as how well water meets certain standards and criteria for various purposes, such as human consumption, agriculture, industry, and ecosystem preservation. Water quality parameters include various physical, chemical, and biological elements that can affect water properties and safety such as bacteria, mercury, nitrate, nitrite, and aluminum. To classify current and future water quality, machine learning-based classification models such as the K-Nearest Neighbors (KNN) algorithm, were calibrated using historical data and tested on independent datasets to evaluate classification accuracy. The result of this study is to use 20 attributes which are water quality parameters. Calculation of model evaluation using confusion matrix and k= 3,5,7 with split data 70% training data and 30% testing data resulted in an accuracy of 87.83% with a "safe" annotation.

Author Correspondence:

Gabriella Puteri Ayu Handiaz Mawarti,
Study Program in Informatics Engineering
Faculty of Computer Science
University of Computer Science, 50131,
Email: 111201912246@mhs.dinus.ac.id

1. INTRODUCTION

The basic basis of life for living creatures found in all parts of the world is air. Air covers almost 71% of the earth's surface, most of which is found in the sea. However, not all of the air on the surface of the earth can be used for living things. Humans need air as much as other living creatures. Humans alone need to drink at least 2 liters of clean water a day [1], [2]. The air will influence and be influenced by the surrounding conditions. The conditions in question are for washing clothes, washing eating utensils, cooking, and bathing. This causes the air that was originally clean to become polluted. The availability of clean water is gradually decreasing due to increasing public demand [3].

Clean water is water for purposes that meets the requirements and is suitable for use. The content in the air contains substances and minerals. Bacteria and minerals can contaminate clean air, which can harm the human body. Air quality is the quality that fulfills certain uses. Requirements placed on air quality vary depending on the intended use of the air consumption [4]. The use of air consumption in question is for households, industry, agriculture.

One data mining technique that is suitable for use in this research is K-Nearest Neighbor. What is meant by K-Nearest Neighbor is an algorithm used to determine groups based on the maximum k nearest neighbors [5], [6]. In this research, the K-Nearest Neighbor algorithm was used based on previous research which has higher accuracy.

.

2. METHOD

2.1. Research Dataset

The water quality data obtained came from the website www.kaggle.com which was uploaded by MsSmartyPants in 2021 with the title Water Quality (Dataset for water quality classification). The variables used in this research for water quality classification are 21 variables which include 20 attributes and include classes/labels as in https://www.kaggle.com/datasets/mssmartypants/water-quality, for an example table can be seen in Table 1 below.

Table 1. The Dataset Attributes

No	Attribute	Information
1.	Aluminium	dangerous > 2.8
2.	Ammonia	dangerous > 32.5
3.	Arsenic	dangerous > 0.01
4.	Barium	dangerous > 2
5.	Cadmium	dangerous > 0.005
6.	chloramine	dangerous > 4
7.	Chromium	dangerous > 0.1
8.	Copper	dangerous > 1.3
9.	Fluoride	dangerous > 1.5
10.	Bacteria	dangerous > 0
11.	Viruses	dangerous > 0
12.	lead	dangerous > 0.015
13.	Nitrates	Dangerous > 10
14.	nitrites	dangerous > 1
15.	Mercury	dangerous > 0.002
16.	perchlorate	dangerous > 56
17.	Radium	dangerous > 5
18.	Selenium	dangerous > 0.5
19.	Silver	dangerous > 0.1
20.	Uranium	dangerous > 0.3

2.2. Water Quality

Water quality is the value required to check whether the water is usable or not. The water needed by the community includes clean water for cooking, washing, bathing, as well as water used for consumption. Getting good quality water with appropriate standards is currently relatively expensive because a lot of water is contaminated with waste from various human activities [7], [8], [9]. Therefore, water resources experience a decline in quality. Likewise in terms of quantity, which is currently no longer able to cover the increasing demand [10].

The increasing demand is what causes the supply of clean water to decrease from time to time, this is because the population increases from the previous year to the next year. Polluted water has a characteristic pungent odor, cloudy color and sour taste [11]. Meanwhile, clean water has the opposite characteristics of polluted water, namely that it has no smell, taste and is clear in color. Examples of contaminated water are water used for washing clothes, factory waste water and river water that has been polluted by rubbish.

Dangerous substances found in water are [12], [13], [14], [15], [16]:

- Aluminum in water can be in the form of small particles dissolved in it. low aluminum concentrations are not harmful to human health. However, aluminum concentrations that are too high can cause health problems such as Alzheimer's.
- Ammonia (NH_3) is a chemical compound that can be dissolved in water and has a great influence on water ecosystems. Although ammonia is beneficial for some aquatic organisms such as algae at low concentrations. Ammonia is toxic to fish because it can damage their gills and disrupt their metabolism, even at high concentrations.
- Arsenic is a contaminant found in water and has a negative impact on human health if the concentration of arsenic in water exceeds safe limits, especially if consumed over a long period of time. Exposure to arsenic can lead to the risk of skin cancer, lung cancer, skin problems, neurological and cardiovascular problems.
- Barium is an alkaline earth metal that occurs naturally in the environment and can be found in water in the form of ions in various compounds. Some health effects associated with excessive exposure to barium include heart problems, neurological disorders, and digestive problems.
- Cadmium is a heavy metal that can be found in water due to natural processes such as the dissolution of rocks and soil deposits, or due to human activities such as industrial waste, agriculture and burning of fossil fuels. Excessive cadmium in water can cause kidney damage, liver damage, disruption of mineral balance in the body, and increased risk of cancer, especially lung cancer.

- Chloramine is a chemical compound formed from the reaction between ammonia (NH3) and chlorine (Cl2) in water. Long-term exposure to high concentrations of chloramine in drinking water can be dangerous for the human body, especially the kidneys and respiratory system.
- In water, chromium can be found in various forms, such as hexavalent chromium (Cr(VI)) and trivalent chromium (Cr(III). Exposure to hexavalent chromium can cause several health problems, such as cancer, liver and kidney damage, respiratory system disorders, and skin problems.
- Copper is a metal that is important in everyday life and can be found in water in low concentrations naturally. However, excessive exposure to copper can cause problems such as digestive problems, liver damage, and nervous system problems.
- Fluoride ions are commonly found in water and usually come from the dissolution of natural minerals. Excessive exposure to fluoride can be harmful to health, especially if it exceeds recommended limits. Excessive exposure to fluoride during growth can cause fluorosis of the bones and teeth.
- Pathogenic bacteria in water come from various sources, such as natural waste, human and animal waste, and industrial activities. Consuming water contaminated with pathogenic bacteria can cause gastroenteritis, urinary tract infections, and other illnesses.
- Viruses in water are small microorganisms that can cause disease if they are present in high concentrations and if the viruses are pathogenic to humans. Some viruses that can be found in water and cause disease in humans are norovirus, rotavirus, hepatitis A virus, and adenovirus.
- Lead is a heavy metal found in water naturally or due to human activities. Exposure to small amounts of lead can cause damage to the nervous system, kidney problems, developmental disorders in children.
- Nitrate is a chemical compound consisting of nitrogen and oxygen and is often found in water in the form of nitrate ion (NO3-). The most common effect of nitrates is methemoglobinemia, or blue baby syndrome.
- nitrite (NO2) is a chemical compound consisting of nitrogen and oxygen. In the case of humans, methemoglobinemia, a condition that interferes with the blood's ability to transport oxygen, can be caused by high exposure to nitrites.
- In natural activities and human activities, mercury, or mercury, is a heavy metal found in water and damages the kidneys, reproductive system and human nervous system.
- Perchlorate also known as perchlorate, is a chemical compound often used in industry and defense, especially as rocket fuel, rocket propellant, and explosives. Perchlorate can interfere with thyroid gland function by stopping the thyroid gland from absorbing iodine.
- Radium is a radioactive element that is naturally available in soil and water. The ionizing radiation emitted by radium increases the risk of developing cancer, especially bladder and colon cancer. This ionizing radiation can damage DNA in body cells.
- Selenium is a mineral that is essential for human health in the right amounts, excessive concentrations of selenium in water or food can cause selenosis, a toxic condition that impacts the liver, kidneys, and nervous system.
- Silver is a valuable metal found in water that can cause irritation to the skin, eyes and respiratory system, as well as damage to the nervous system.
- Uranium is a heavy metal that can dissolve in water, and can be dangerous for human health. Radioactive uranium can release alpha radiation, which can damage body tissue and increase the risk of cancer, especially bladder cancer and colon cancer.

2.3. K-Nearest Neighbor (K-NN)

The K-Nearest Neighbor (KNN) algorithm is a data classification based on the closest distance to an object or what is usually called a neighbor in training data and testing data [9], [13]. The following is the formula for K-Nearest Neighbor:

$$\sqrt{\sum_{i=1}^{k}(xi - yi)^2}$$

Where :
K = determines the data attribute
Xi = determines the training data

Yi = determines testing data or test data
K-Nearest Neighbor has calculation steps, namely:
1. Determine the set/parameter k = the closest number of data
2. Calculate the amount of new data with training data
3. Sort the closest data based on the minimum distance of the K value
4. Check the class data from the closest data
5. Determine the results from the closest data as the predicted value of the new data.

Calculating the distance between new data and old data can be calculated using several methods, one of which uses the Euclidean distance formula. The formula for calculating Euclidean distance can be seen as follows:

$$\sqrt{(a1 - b1)^2 + \cdots + (an - bn)^2}$$

Where :
a = a1, a2, a3, …., an up to the nth value
b = b1, b2, b3, …., bn up to the nth value

2.4. Proposed Method

Figure 1 is the flow that will be implemented using the K-Nearest Neighbor Algorithm. Preparations before calculating are as below:

- Cleaning data. Data cleaning is a step that needs to be carried out to deal with missing values and inconsistent data.
- Data Selection. Data Selection is the stage of reducing the quantity of data used in the mining stage while still presenting the original data. In this data selection step, only data with important characteristics that are usually used in good water quality prediction classes are selected. In this study, all available attributes influenced the results. There are 21 attributes used including class/tag.
- Data Transformation. Data transformation is carried out to correct the shape and format of the data.

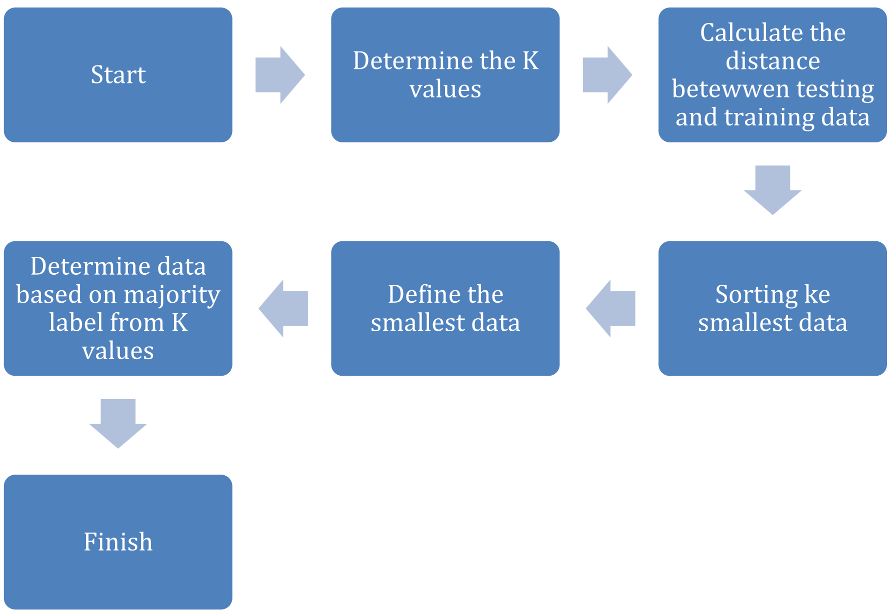

Figure 1. Classification of water quality based on K-NN

The model evaluation phase determines the performance of the algorithm. To understand the accuracy of the model, a confusion matrix must be calculated. Confusion matrix is a technique that can be used to test the working performance of a model, especially in terms of classification (supervised learning) in machine learning as in Table 2.

Table 2. Confussion Matrix

	Actual value

Confusion Matrix		Positive	Negative
Prediction value	Positive	True Positives	False Positives
	Negative	False Negatives	True Negatives

Description of the 4 values in Table 2, namely:

a. True Positives (TP): data that has a positive value and is requested to be positive.
b. False Positives (FP): data that has a negative value and is requested to be positive.
c. False Negatives (FN): data that has a positive value and is requested to be negative.
d. True Negatives (TN): data that has a negative value and is requested to have a negative value

3. RESULTS AND DISCUSSION

Table 1 and Table 2 below are visualizations of training test data and manual calculation test data using the KNN algorithm. The first step in calculating the KNN algorithm using the Euclidean Distance method is to determine the value of K. With this manual calculation, the author decided to apply the value k = 3. Next, we will calculate the training data using Euclidean Distance as sample:

$D_{(1,11)} =$

$\sqrt{(1{,}65-0{,}61)^2+(9{,}08-2{,}41)^2+(0.04-0{,}03)^2+(2{,}85-0{,}59)^2}+$

$\sqrt{(0{,}07-0{,}002)^2+(0{,}35-1{,}94)^2+(0{,}83-0{,}77)^2+(0{,}17-1{,}54)^2}+$

$\sqrt{(0{,}05-0{,}62)^2+(0{,}2-0{,}23)^2+(0-0{,}001)^2+(0{,}54-0{,}017)^2}+$

$\sqrt{(16{,}08-1{,}99)^2+(1{,}13-1{,}08)^2+(0{,}007-0{,}007)^2+(37{,}75-11{,}16)^2}$

$\sqrt{+(6{,}78-0{,}98)^2+(0{,}08-0{,}001)^2+(0{,}34-0{,}47)^2+(0{,}02-0{,}03)^2}$

$= \sqrt{1{,}0816+44{,}4889+0{,}0001+5{,}1076+0{,}004624}$

$\sqrt{2{,}5281+0{,}0036+1{,}8769+0{,}3249+0{,}0009}$

$\sqrt{0{,}001+0{,}273529+198{,}5281+0{,}0025+0+707{,}0281+33{,}64}$

$\sqrt{0{,}0049+0{,}0169+0{,}0001}$

$= \sqrt{994{,}912353}$

$= 31{,}5422309$

$D_{(2,11)} =$

$\sqrt{(2{,}32-0{,}61)^2+(21{,}16-2{,}41)^2+0{,}01-0{,}03)^2+(3{,}31-0{,}59)^2}+$

$\sqrt{(0{,}02-0{,}002)^2+(5{,}28-1{,}94)^2+(0{,}68-0{,}77)^2+(0{,}66-1{,}54)^2}+$

$\sqrt{(0{,}9-0{,}62)^2+(0{,}65-0{,}23)^2+(0{,}65-0{,}001)^2+(0{,}1-0{,}017)^2}+$

$\sqrt{(2{,}01-1{,}99)^2+(1{,}93-1{,}08)^2+(0{,}003-0{,}007)^2+(32{,}26-11{,}16)^2}$

$\sqrt{+(3{,}21-0{,}98)^2+(0{,}08-0{,}01)^2+(0{,}27-0{,}47)^2+(0{,}05-0{,}03)^2}+$

$= \sqrt{2{,}9241+351{,}5625+0{,}0004+7{,}3984+0{,}000324+11{,}1556}$

$= \sqrt{0{,}0081+0{,}7744+0{,}0784+0{,}1764+0{,}421201}$

$= \sqrt{0{,}006889+0{,}0004+0{,}7225+0{,}00016+445{,}21+4{,}9729}$

$= \sqrt{0{,}0049+0{,}04+0{,}0004}$

$= \sqrt{825{,}530877}$

$= 28{,}7320531$

Table 3. Training data

		1	2	3	4	5	6	7	8	9	10
Attribute	aluminium	1.65	2.32	1.01	1.36	0.92	0.94	2.36	3.93	0.6	0.22
	ammonia	9.08	2116	14.02	11.33	24.33	14.47	5.6	19.87	24.58	16.76
	arsenic	0.04	0.01	0.04	0.04	0.03	0.03	0.01	0.04	0.01	0.02
	barium	2.85	3.31	0.58	2.96	0.2	2.88	1.35	0.66	0.71	1.37
	cadmium	0.007	0.002	0.008	0.001	0.006	0.003	0.004	0.001	0.005	0.007
	chloramine	0.35	5.28	4.24	7.23	2.67	0.8	1.28	6.22	3.14	6.4
	Chromium	0.83	0.68	0.53	0.03	0.69	0.43	0.62	0.1	0.77	0.49
	Copper	0.17	0.66	0.02	1.66	0.57	1.38	1.88	1.86	1.45	0.82
	Flouride	0.05	0.9	0.99	1.08	0.61	0.11	0.33	0.86	0.98	1.24
	Bacteria	0.2	0.65	0.05	0.71	0.13	0.67	0.13	0.16	0.35	0.83
	Viruses	0	0.65	0.003	0.71	0.001	0.67	0.007	0.005	0.002	0.83
	Lead	0.054	0.1	0.078	0.016	0.117	0.135	0.021	0.197	0.167	0.109
	Nitrates	16.08	2.01	14.16	1.41	6.74	9.75	18.6	13.65	14.66	4.79
	Nitrites	1.13	1.93	1.11	1.29	1.11	1.89	1.78	1.81	1.84	1.46
	Mercury	0.007	0.003	0.006	0.004	0.003	0.006	0.007	0.001	0.004	0.1
	Perchlorate	37.75	32.26	50.28	9.12	16.9	27.17	45.34	53.35	23.43	30.42
	Radium	6.78	3.21	7.07	1.72	2.41	5.42	2.84	7.24	4.99	0.08
	Selenium	0.08	0.08	0.07	0.02	0.02	0.08	0.1	0.08	0.08	0.03
	Silver	0.34	0.27	0.44	0.45	0.06	0.19	0.24	0.08	0.25	0.31
	Uranium	0.02	0.05	0.01	0.05	0.02	0.02	0.08	0.07	0.08	0.01
class	Is_safe	safe	safe	unsafe	safe	safe	safe	unsafe	unsafe	safe	safe

Table 4. Testing data

		1	2	3	4	5
Attribute	aluminium	0.61	3.47	2.11	4.88	4.12
	ammonia	2.41	15.84	17.03	26.94	17.99
	arsenic	0.03	0.02	0.02	0.2	0.2
	barium	0.59	0.06	0.88	0.36	3.43
	cadmium	0.002	0.001	0.009	0.001	0.001
	chloramine	1.94	5.29	7.78	1.21	0.01
	Chromium	0.77	0.47	0.88	0.68	0.41
	Copper	1.54	1.08	1.15	0.71	1.82
	Flouride	0.62	1.43	0.34	0.99	0.22
	Bacteria	0.23	0.89	0.85	0.75	0.99
	Viruses	0.001	0.89	0.85	0.75	0.99
	Lead	0.017	0.08	0.065	0.071	0.108
	Nitrates	1.99	1.91	17.86	0.31	8.06
	Nitrites	1.08	1.2	1.53	1.22	1.76
	Mercury	0.007	0.008	0.003	0.002	0.005
	Perchlorate	11.16	0.18	19.4	56.7	24.29
	Radium	0,98	6.89	1.14	1	0.88
	Selenium	0.01	0.06	0.1	0	0.1
	Silver	0.47	0.12	0.4	0.41	0.1
	Uranium	0.03	0.08	0.01	0.05	0.07
Class	Is_safe	safe	Safe	Safe	Unsafe	safe

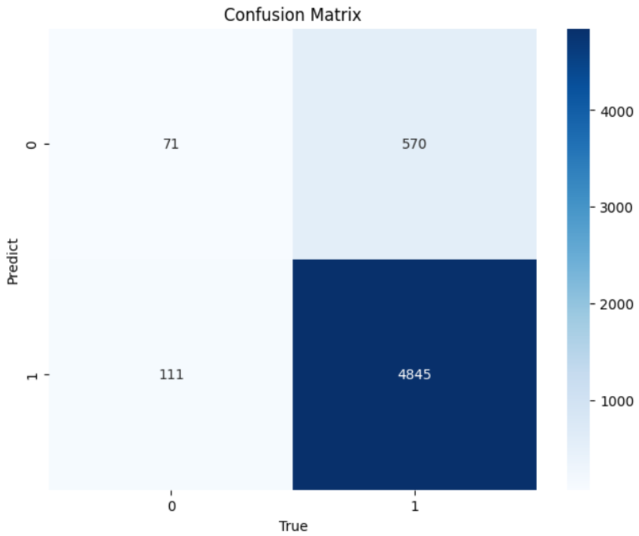

Figure 2. Confusion Matrix using training 70%, test 30%

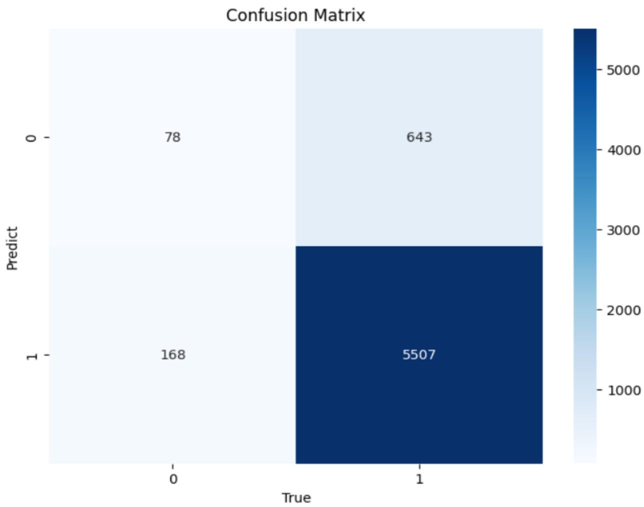

Figure 3. Confusion Matrix using training 80%, test 20%

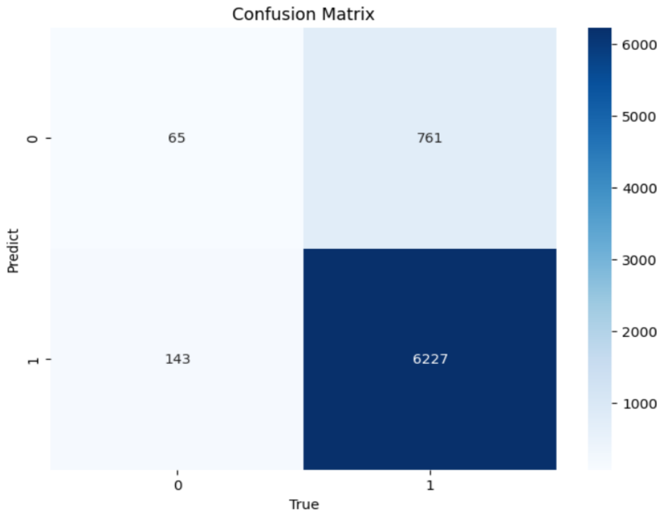

Figure 4. Confusion Matrix using training 90%, test 10%

Table 5. Calculation result of testing data

No	Data	Distance	Sequence	Class
1	$D_{(1,11)}$	31,5422309	7	safe
2	$D_{(2,11)}$	28,7320531	6	safe
3	$D_{(3,11)}$	43,2717001	9	unsafe
4	$D_{(4,11)}$	10,9464453	1	safe
5	$D_{(5,11)}$	23,2338449	3	safe
6	$D_{(6,11)}$	22,1392021	2	safe
7	$D_{(7,11)}$	38,2480196	8	unsafe
8	$D_{(8,11)}$	47,8660592	10	Unsafe
9	$D_{(9,11)}$	28,6588224	5	safe
10	$D_{(10,11)}$	24,6587983	4	safe

Table 6. Evaluation of K-NN Performance

Split Data	K Value	Accuracy
70 : 30	3	87,83%
	5	87,83%
	7	87,83%
80 : 20	3	87,32%
	5	87,32%
	7	87,32%
90 : 10	3	87,43%
	5	87,43%
	7	87,43%

From the three data in Table 5, the value k = 3 shows that the data in the order 1 to 3 have the description "safe". Of all the comparative accuracy evaluation results, the split data results were 70:30, giving the most accurate results of 87.835%. Other data split differences have lower results. From Table 6 it can be seen that the accuracy value has a very large level, namely 87.83%.

4. CONCLUSION

Based on the calculations that have been carried out, it can be concluded that the K-Nearest Neighbor algorithm is capable of predicting water quality by getting a fairly high accuracy of 87.83% using a value of k = 3.5.7 where the value of K3 = 87.83%, K5 = 87 .83%, K7 = 87.83% and split training data 70% and testing data 30% with the statement "safe". Future research can use and develop this research using other algorithmic methods so that calculations are more accurate. In future research, there needs to be a comparison between several other algorithms. It is hoped that in future research we can use k value tests which the author did not use in testing the K-NN algorithm, as well as test training data and testing data with varying percentages.

REFERENCES

[1] M. G. Uddin, S. Nash, A. Rahman, and A. I. Olbert, “Performance analysis of the water quality index model for predicting water state using machine learning techniques,” *Process Safety and Environmental Protection*, vol. 169, pp. 808–828, Jan. 2023, doi: 10.1016/j.psep.2022.11.073.

[2] M. G. Uddin, S. Nash, M. T. Mahammad Diganta, A. Rahman, and A. I. Olbert, “Robust machine learning algorithms for predicting coastal water quality index,” *J Environ Manage*, vol. 321, Nov. 2022, doi: 10.1016/j.jenvman.2022.115923.

[3] N. Nasir *et al.*, “Water quality classification using machine learning algorithms,” *Journal of Water Process Engineering*, vol. 48, p. 102920, Aug. 2022, doi: 10.1016/j.jwpe.2022.102920.

[4] M. Hmoud Al-Adhaileh and F. Waselallah Alsaade, “Modelling and Prediction of Water Quality by Using Artificial Intelligence,” *Sustainability*, vol. 13, no. 8, p. 4259, Apr. 2021, doi: 10.3390/su13084259.

[5] H. Tahraoui *et al.*, “Advancing Water Quality Research: K-Nearest Neighbor Coupled with the Improved Grey Wolf Optimizer Algorithm Model Unveils New Possibilities for Dry Residue Prediction,” *Water (Basel)*, vol. 15, no. 14, p. 2631, Jul. 2023, doi: 10.3390/w15142631.

[6] A. Danades, D. Pratama, D. Anggraini, and D. Anggriani, “Comparison of accuracy level K-Nearest Neighbor algorithm and Support Vector Machine algorithm in classification water quality status,” in

2016 6th International Conference on System Engineering and Technology (ICSET), IEEE, Oct. 2016, pp. 137–141. doi: 10.1109/ICSEngT.2016.7849638.

[7] B. Aslam, A. Maqsoom, A. H. Cheema, F. Ullah, A. Alharbi, and M. Imran, "Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach," *IEEE Access*, vol. 10, pp. 119692–119705, 2022, doi: 10.1109/ACCESS.2022.3221430.

[8] A. Juna *et al.*, "Water Quality Prediction Using KNN Imputer and Multilayer Perceptron," *Water (Basel)*, vol. 14, no. 17, p. 2592, Aug. 2022, doi: 10.3390/w14172592.

[9] X. Jia, "Detecting Water Quality Using KNN, Bayesian and Decision Tree," in *2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML)*, IEEE, Mar. 2022, pp. 323–327. doi: 10.1109/CACML55074.2022.00061.

[10] 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," *Multimed Tools Appl*, vol. 83, no. 12, pp. 35307–35334, Sep. 2023, doi: 10.1007/s11042-023-16737-4.

[11] Y. Li *et al.*, "CART and PSO+KNN algorithms to estimate the impact of water level change on water quality in Poyang Lake, China," *Arabian Journal of Geosciences*, vol. 12, no. 9, May 2019, doi: 10.1007/s12517-019-4350-z.

[12] L. Dong, X. Zuo, and Y. Xiong, "Prediction of hydrological and water quality data based on granular-ball rough set and k-nearest neighbor analysis," *PLoS One*, vol. 19, no. 2 February, Feb. 2024, doi: 10.1371/journal.pone.0298664.

[13] X. Guoqiang, Z. Yi, X. Shiyi, H. Miaofen, and Z. Ying, "Multi-Classification Method for Determining Coastal Water Quality based on SVM with Grid Search and KNN," *International Journal of Performability Engineering*, vol. 15, no. 10, p. 2618, 2019, doi: 10.23940/ijpe.19.10.p7.26182627.

[14] M. Hamzaoui, M. O.-E. Aoueileyine, and R. Bouallegue, "A Hybrid Method of K-Nearest Neighbors with Decision Tree for Water Quality Classification in Aquaculture," in *Communications in Computer and Information Science*, vol. 1864 CCIS, Springer Science and Business Media Deutschland GmbH, 2023, pp. 287–299. doi: 10.1007/978-3-031-41774-0_23.

[15] E. Dritsas and M. Trigka, "Efficient Data-Driven Machine Learning Models for Water Quality Prediction," *Computation*, vol. 11, no. 2, p. 16, Jan. 2023, doi: 10.3390/computation11020016.

[16] A. Jha, M. Chowdhury, and A. N. Satpute, "Surface Water Quality Forecasting Using Machine Learning Approach," in *Surface and Groundwater Resources Development and Management in Semi-arid Region: Strategies and Solutions for Sustainable Water Management*, 2023, pp. 293–315. doi: 10.1007/978-3-031-29394-8_16.