Research Article

Low-Cost Domestic Wastewater Pollution Monitoring System in Residential Areas using IoT: Case Studies in Bandung Indonesia

Helmy Faisal Muttaqin[1], Ucu Nugraha[2]
[1]Engineering Faculty, Widyatama University, Jalan Cikutra No 204a Bandung, Indonesia
[2]Engineering Faculty, Widyatama University, Jalan Cikutra No 204a Bandung, Indonesia
[1]Helmy.faisal@widyatama.ac.id

Article History: Received: 10 January 2021; Revised: 12 February 2021; Accepted: 27 March 2021; Published online: 20 April 2021

Abstract: Monitoring and controlling river pollution from domestic waste is currently a crucial part of urban community development. Nowadays, it is difficult to determine the source of pollutants in the river because data on the number of domestics that dispose of wastewater into rivers is still very limited. Surface water samples are tested by environmental laboratories only twice a year, or based on requests from interested parties such as companies, factories, local government, and local police departement based on suspicion or reports of suspected waste disposal in specific areas. However, during our sampling, it often occurs that the river water to be tested has returned to normal because the wastewater discharged by the domestic waste has already flowed downstream, especially in the rainy season. Therefore it is necessary to monitor in a situation that can measure the quality of river water, estimate the source of pollutants, and monitor the time of taking river water properly so that monitoring and control of river pollution can be carried out. The Internet of Things technology enables local governments to record river or ditch water quality faster through sensors of water temperature, turbidity, total dissolved solids, pH, air temperature, water level, and rain. With the Decision Tree C4.5 algorithm, successful data recorded can be analyzed to present the highest contamination time. This time of highest contamination can be used as a recommendation for sampling time for laboratory test results. The combination of Internet of Things technology, Rest API, and Decision Tree C4.5 can produce a monitoring System for river and ditch water content used for monitoring and controlling water pollution.

Keywords: Decision Tree C4.5, Internet of Things, Monitoring System, Smart Environment, Storet Method, Water Quality Standards, Low-Cost

1. Introduction

In 2013, the Citarum River in Indonesia became one of the most polluted and most toxic rivers in The World [1] [2]. The Citarum watershed becomes a dumping ground for industries located along the Citarum creek and river. Cimahi City is one of the cities located in West Java - Indonesia, with five rivers flowing into the Citarum River, has a population growth rate of 1.12%, and has an economic characteristic dominated by the industrial sector [3]. Manufacture industries dominate 14 percent of The Area and still a visible industrial area development to contribute to economic development. The implication is an increased production of waste, both from industrial waste and domestic waste. The increasing population and industry contribute to environmental pollution problems, especially pollution in river water caused by liquid and solid waste from domestics and industrial processing waste. The industry that dominates in the processing industry, and almost half of it is the textile industry [3]. There is still liquid waste disposed of by the processing industry that does not comply with the industrial waste disposal procedures. Previous research found that 41% of companies were still violating waste management regulations [4]. Of all rivers, around 54.46% of rivers are in critical condition, 16.07% are in very polluted conditions, and 17.49% are in heavily polluted conditions [5].

To fix this problem, the Local Government has made various efforts. Among these efforts is establishing a Pollution and Damage Control Task Force (PPK) for the Citarum Watershed at the provincial level in West Java. To monitor and control the river pollution level by testing river water samples routinely two times a year, socializing and coaching the community, Micro, Small, and Medium Enterprises (MSMEs), and socializing technical waste management to all industry players [5]. However, pollution control has not shown the expected results. Among them can be seen from the still finding of companies that violate waste management regulations and the decline in water quality from year to year from the laboratory test reports of river water samples from farious conditions. Also, companies that dispose of hazardous waste directly into rivers at certain times in a short time, and farious of domestic household waste from community settlements making it difficult for the Local Government to monitor it.

Until now, online monitoring has not been carried out by the Local Government. There is because the required equipment is still too expensive. In 2010-2012, they used a device called telemetering, a very expensive multiprobe water sensor, which cost $30.000. However, the device damaged by mud flood and could not be repaired due to limited funds. Therefore, we need tools that is more economical but can still be used to monitor water quality in ditch and rivers.

The use of technology for environmental pollution detection has been widely used, including Internet of Things (IoT) technology. Water sensors that are already available can be used on IoT devices for monitoring river water at a low price. Systems developed using IoT can provide services that make it a more comfortable life for communities, supported by adequate infrastructure, build comfortable and viable environmental solutions for smart cities. The use of IoT is the best choice for creating a Smart City [6]. Researchers from various countries use IoT sensors to monitor water quality with any requirement customization and ease of maintenance [7] and in the additional package, using IoT to monitor water quality combined with machine learning to obtain information on the feasibility of drinking water. The sensors used in this study include sensors for pH, water temperature, turbidity, and conductivity [8]. The use of the GSM module as an additional device for sending notification of the results of water quality monitoring has also begun to be implemented [9].

For water quality modeling, there have been studies using the Decision Tree for classification and prediction of water pollutant classes such as Nitrogen (NH3_N, NO3_N), pH, Temp _C, BOD, COD, and other parameters to obtain water quality values [10]. In other water quality studies, it also knows the feasibility of drinking water based on pH, alkalinity, conductivity, and watercolor parameters [11]. Meanwhile, environmental predictions research widely use Decision Tree C4.5, including rainfall prediction from parameters of air temperature, rain, solar radiation, humidity, wind speed, and evaporation [12].

This research is focusing on monitoring environmental pollution, especially ditch and river water pollution from domestics wastewater. Combining IoT technology with Machine Learning (C4,5 algorithm) will obtain a classification and prediction of ditch and river water pollution to help the Local Government make decisions for controlling and preventing river water pollution going worst. The output from this research will later produce an online ditch and river quality monitoring System.

The IoT devices used in this monitoring System are cheaper than the telemetering devices used by the Local Government before. The price per IoT device at one monitoring point is around $200 because the only sensors used are water temperature, TDS, pH, turbidity, air temperature, ultrasonic to detect river water level, and rain sensors. Unlike other monitoring Systems, the combination of water sensors and environmental sensors such as rain sensors, air temperature sensors, and water level sensors allows this System to detect the causes of changes in river water parameter values. Based on location coordinate data and ditch and river water parameter data detected by IoT devices, it can be estimated the location of ditch river water pollutant sources. IoT devices on this System are also portable, its mean they can be placed and moved easily anytime, so they are safe from damage. However, even though it is portable, the recording of water parameters can be adjusted as needed or continuously using a power source either from solar panels or batteries or 12v ACCU. The continuous recording will produce continuous data and form a pattern that can be a data source for predicting the highest contamination time. This prediction of the highest pollution time can be used as a reference for ditch and river water sampling. Laboratory test results on this ditch and river water samples can show results that are more appropriate to conditions in the field.

2. Background

A. HOW IS WATER POLLUTION DEFINED

The definition of water pollution is the entry or inclusion of living things, substances, energy, and or other components into the water by human activities so that the quality of water drops to a certain level, which causes water to not function according to its purpose [13].

Water content, which is used as a parameter to determine pollution, is divided into 4 [13] including:

1) Physical parameters, for example, are water temperature, dissolved residue (TDS / Total Dissolved Solids), suspended residue (TSS / Total suspended solid), turbidity (Turbidity), etc.
2) Chemical Parameters, for example, pH, DO (Dissolved Oxygen), COD (Chemical Oxygen Demand), BOD (Biological Oxygen Demand), oils and fats, etc.
3) Microbiological Parameters, for example, Fecal Coliform and Total Coliform.
4) Radio Activities, for example, Gross-A and Gross-B.

Waste generated by human activities can cause changes in water content, both in the industrial and household (domestic) scope or the environment, such as changes in weather and air temperature. Flood events that affect ditch and river water level and rainfall are also significant factors affecting river water quality. The results of the analysis in a study indicate that high rainfall and surface runoff have increased the pollutant load into ditch and river, thus impacting on water quality [14]. Likewise, rainfall events after a long-term dry season accelerate the decline in water quality because pollutants can accumulate in surface areas during the dry season and are carried by rainwater into rivers [15].

The effect of rain on the pH content in water can cause pH levels to tend to be high because the accumulation of carbonates and bicarbonates can cause the pH of river water to tend to be alkaline [16] [17]. The textile industry contributes to waste output with an acidity level of pH 8-10, while the typical pH values that are normal and according to quality standards are in the range of 6-9 [18]. As for the effect of air temperature on water temperature, every one ℃ increase in air temperature, there is an increase in water temperature by 0.6-0.8 ℃. Trends in air temperature and water temperature appear to be linear 1: 1. So for river water, it is likely that the water temperature will increase two ℃ to three ℃ because the air temperature increases three ℃ to five ℃ [19].

B. WHY WATER POLLUTION MUST BE MAPPED

Mapping of ditch and river water pollution is carried out to determine the location where the highest pollution occurs. Supervision can be carried out on companies, and community settlements in the area around that location. The location of the pollution is obtained from the distribution of ditch and river water sampling points for laboratory tests. The coordinates of the sampling location will be used as the placement point for the IoT device.

Each IoT device that is placed at a predetermined coordinate point will detect the content of each water parameter using a water parameter sensor. Any data recorded by the sensor will be stored and sent to a database in the Cloud server using Rest API.

C. HOW TO CALCULATE THE QUALITY OF WATER

The methods commonly used to determine water quality are the Pollution Index (IP) and CCME (Canadian Council of Ministers of The Environment). The Pollution Index method is superior to use for single data. Although IP has advantages in terms of cost and time, the results only present the status of water quality at one time and not within a certain period. The Storet and CCME methods use time-series data, which can present the status of water quality in a certain period [20].

In terms of sensitivity in responding to the dynamics of water quality at each monitoring point, the CCME is the most sensitive method because it can calculate with few or many parameters and with or without bacteriological parameters [21].

D. HOW TO PREDICT WATER POLLUTION

Water pollution predictions were obtained using the Decision Tree C4.5 on the data that was successfully recorded by the IoT device at each location. A modeling tree will be built so that the results of the confidence “Yes” and “No” to determine the time and hour of the highest pollution.

The training data that will be built by the tree model is training data which is sourced from:

1) Data from sensing results of IoT devices at 15 ditch and river points (limited number sampling)
2) Environmental Laboratory Data
3) BMKG Website data (trends in air temperature and rainfalls)
4) Simulation data from ditch and river water samples taken from several polluted ditch and river points and a combination of the data patterns above.

The resulting training data are as follows:

1) The number of Raw Data in table tmp_wqc is 328,320 records
2) Total data after ETL 54,720 records
3) Training data of 54,720 records (Data for January June 2020
4) Test data of 6,840 records 456 per 20-day data stream point in January 2021

After the tree modeling has been successfully constructed, a test will be conducted using test data so that the values of confidence, accuracy, precision, recall, and classification error will be generated. The highest confidence value will be taken as the day and hour predicted to have the highest level of pollution.

3. Method

A. STORET METHOD

The Storet method is one of the methods used to determine the status of water quality by comparing the parameters of the water content with the water quality standard that has been determined according to its designation. The Storet method uses the US-EPA (Environmental Protection Agency) value system by classifying water quality into four classes [22]. Data grouping was based on water parameters related to physical factors (sight, touch, taste, smell), chemistry (dissolved substances), and biology (presence of microorganisms) [23].

The steps for the Storet calculation are as follows:

1) Collect water quality and water discharge data periodically to form data from time to time (time series data).
2) Compare the measurement data of each water parameter with the standard quality value according to the water class.
3) If the measurement results meet the water quality standard (measurement results <quality standard), then a score of 0 is given.

4) If the measurement results do not meet the water quality standard (measurement results> quality standard), then a score is given in table II.2.

5) The number of negatives of all parameters is calculated, and the quality status is determined from the total score obtained by using a value system.

TABLE II
WATER QUALITY CLASS US-EPA

Class	Score	Explanation
A: Very Good	0	Meet quality standards
B: Good	-1 to -10	Lightly polluted
C: Moderate	-11 to -30	Moderately polluted
D: Bad	≥ -30	Heavy polluted

TABLE III
DETERMINATION OF VALUE SYSTEMS TO DETERMINE WATER QUALITY STATUS

Number of Examples (Canter, 1977)	Grade	Parameter		
		Physics	Chemical	Biology
< 10	Maximum	-1	-2	-3
	Minimum	-1	-2	-3
	Average	-3	-6	-9
≥ 10	Maximum	-2	-4	-6
	Minimum	-2	-4	-6
	Average	-6	-12	-18

The method of calculating the status of water quality currently used is the Storet method [24]. Another study using the Storet Method for measuring ditch and river water quality was also carried out in the Kalibaru river [25].

B. DECISION TREE C4.5

In 1993, Ross Quinlan proposed a decision tree C4.5, which is a development of an ID3 decision tree that overcomes the limitations of the ID3 algorithm [26]. In the ID3 algorithm, the best Attribute is selected based on the concept of entropy and gain information to build a tree [26]. Whereas in the C4.5 algorithm, there are several improved ID3 behaviors, including:

1) Possibility to use continuous data.
2) It can use unknown or missing values.
3) It`s ability to use attributes with different weights.
4) It`s ability to prun the tree.
5) Pessimistic prediction errors.
6) Increase the Sub Tree.

Like ID3, which uses entropy and information gain, C4.5 continues its calculations with the information split and gain ratio [26].

1) ENTROPY

This entropy states the impurity of a collection of objects [27], which is used as a parameter to measure the heterogeneity of the data sample—the greater the entropy, the more heterogeneous.

$$Entropy\ (S) = \sum_{j=1}^{k} - P_j\ log_2\ P_j \qquad (1)$$

S is the sample dataset used for training data, k is the number of partitions S, and Pj is the probability obtained from the total number (Yes) divided by the total cases.

2) INFORMATION GAIN

Information Gain is an impurity-based criterion that uses the entropy measure (originating from information theory) as the impurity measure [28]. Information gain is the most popular criterion for attribute selection [27].

$$Gain\ (a) = Entropy\ (S) - \sum_{i=1}^{k} \frac{|S_i|}{|S|}\ x\ Entropy\ (S_i) \qquad (2)$$

S is the sample dataset used for training data, a is Attribute, | Si | is the number of samples for the value I, and | S | is the sum of all data samples.

3) SPLIT INFORMATION

SplitInfo or Information Split states entropy or potential information [27].

$$SplitInfo\ (S,\ a) = -\sum_{i=1}^{n} \frac{S_i}{S}\ log_2 \frac{S_i}{S} \qquad (3)$$

S is the sample dataset used for training data, a is the Attribute, and Si is the number of samples for Attribute i.

3) CONTINOUS ATTRIBUTE

Continuous Attribute is an attribute that is not separate. Continuous attributes are usually represented as floating-point variables [29]. C4.5 can solve cases that have continuous attribute values. The way to get the right split for Continous Attributes is by sorting the data from smallest to largest. Then partitioned and calculated the gain of each partition to get the most significant gain [12].

4) GAIN RATIO

The gain ratio is the division of gain by splitting information [26].

$$Gain\ Ratio\ (a) = \frac{Gain\ (a)}{Split\ (a)} \qquad (4)$$

5) PRUNING

In 1984, Breiman et al. Developed a pruning method that pruned branches that did not contribute to overall accuracy [30]. Pruning attempted to identify and remove tree branches to increase the classification accuracy of unseen data [29]. So pruning is part of the decision tree process where the use of pruning will reduce data outliers or noise in the initial decision tree to increase the accuracy of the data classification.

6) CONFUSION MATRIX

Confusion Matrix is a tool to analyze how well the classifier can recognize tuples of different classes [31]. The Confusion Matrix is used to determine the level of accuracy of the classification model created and consists of columns representing the predictive data class and rows representing the original data class or other.

TABLE IV
CONFUSION MATRIX

		Predicted Class		
		Yes	***No***	***Total***
Actual Class	***Yes***	True Positive (TP)	False Negative (FN)	P
	No	False Positive (FP)	True Negative (TN)	N
	Total	P’	N’	P +N

The accuracy of a classification model can be calculated as in formula (5). Meanwhile, the error rate or classification error can be calculated using formula (6). Precision can be calculated using formula (7), and formula (8) is the formula for recall.

$$accuracy = \frac{FP + FN}{P + N} \qquad (5)$$

$$error\ rate = \frac{FP + FN}{P + N} \qquad (6)$$

$$precision = \frac{TP}{TP + FN} \qquad (7)$$

$$recall = \frac{TP}{TP + FN} = \frac{TP}{P} \qquad (8)$$

C. REST API

RestAPI was chosen in this study because it has several advantages [32], namely:

1) Scalability. Because there is a separation of servers and clients on the System to be built in this study, where the server System is on the Cloud and the IoT device as a client, the System can be scaled without much difficulty when there is further development.

2) Portability and flexibility. In this study, one of the processes on the System is the transfer of data from the IoT device to the database server in the Cloud. With the help of RestAPI, it is enough to get the required data in one request. It will also provide access to modifications to the database at any time.

3) Independence. The separation between server and client, which is done exclusively, makes it easy for the protocol to work on various development projects and be carried out independently. RestAPI can also adapt to different work Systems and syntax. This brings many opportunities to work on the System building in this research.

4. System development

A. BUILD AN IOT DEVICE

The selection of sensors for IoT devices is based on existing research and other considerations based on the effects on the ups and downs of ditch and river water quality. Water temperature, TDS, turbidity, and pH sensors are commonly chosen sensors to monitor water quality [33, 37]. The four sensors also include parameters that have been set in the State Environmental Metering Decree as part of the parameters for determining water quality [22]. Other sensors that are added are air temperature, distance, and rain sensors.

An air sensor is added to this IoT device because air temperature dramatically affects the temperature of the water in the area to detect when there is liquid waste passing through the sensor. Expected water temperature is usually below the air temperature, with water temperature increasing by about 0.6–0.8 ° C for every 1 ° C increase in air temperature [19]. If the water temperature is close to air temperature and exceeds the air temperature value, there is undoubtedly liquid waste passing through the device. This is due to the nature of the waste, which has a higher temperature than clean water temperature [34].

The selection of the rain sensor is strongly influenced by the increase in pH levels during the rainy season. This is due to the accumulation of carbonate and bicarbonate compounds, so that ditch and river water is more alkaline [16]. One of the characteristics of textile industrial waste is that it has an acidity of pH 8-10 [18]. With the rain sensor, it is hoped that the System can distinguish between changes in pH by waste and pH increases during the rainy season.

In addition to sensors for air temperature and rain status, there are additional proximity sensors that are selected to determine the water level of the ditch and river. This proximity sensor can be combined with rain status to help detect river level rise so the System can distinguish between rain level rise and industrial waste disposal. The proximity sensor device used is a waterproof ultrasonic distance sensor to withstand a field that can be exposed to water at any time.

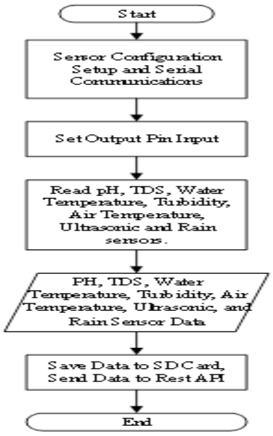


FIGURE 1. Internet of Things device reading flowchart

The selection of location points for placing IoT devices is based on the location of the water sampling point. This is done because this System aims to provide a recommended time for water sampling at the time of the highest pollution at the sampling point. During implementation, the infrastructure surrounding the device must also be considered, such as the availability of milestones for storing IoT devices, the GPRS signal strength of the selected GSM vendor, and the power source. Apart from infrastructure security, the security of tools installed from theft and vandalism must also be a significant concern.

B. DEVELOPING PREDICTIVE MODELS

The database used to store data from IoT devices, Storet calculation results, and ETL results is a wqms database with a data structure, as shown in Fig. 2.

The data collected is data from 5 (five) rivers consisting of 15 monitoring points. The data recorded by the sensor has various conditions, as follows:

1) There are still dirty data caused by the adaptation of the tool when it is turned on, and the unstable output location.

2) There is null data due to the failure of the recording by the sensor.

The sensor will record data once every 20 seconds and stored in a file with a txt extension. The data in the file is still dirty as recorded by the sensor. The dirty data will then be cleaned before being saved to the database in the tmp_wqc table. The results of the data cleaning process are stored in the tmp_wqc table totaling 328,320 records recorded from January to June 2020, with an average recording of each record being 10 minutes. After the data is stored in the tmp_wqc table, the average data per hour will be calculated and stored in the td_formatwqms table.

In this td_formatwqms table, the time attribute shows the time per hour, and data that has decimal values have been rounded. This is the first step in data simplification before data formatting is carried out, which is the transform stage of the ETL process.

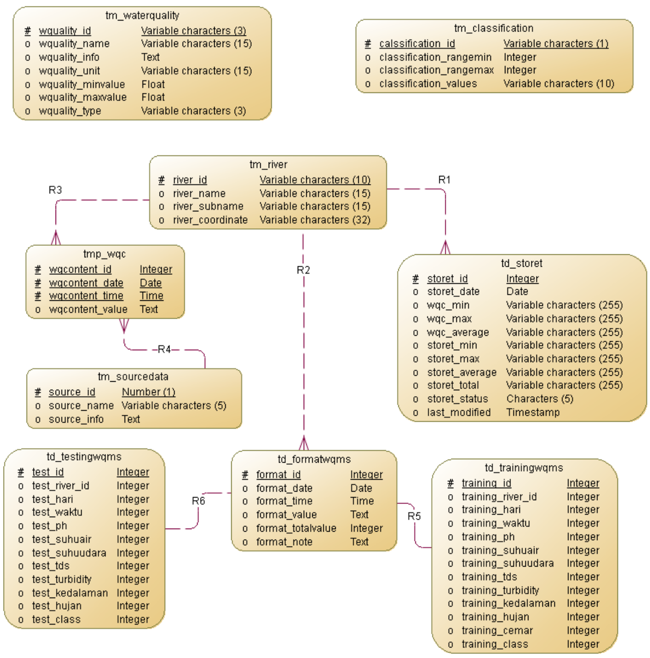


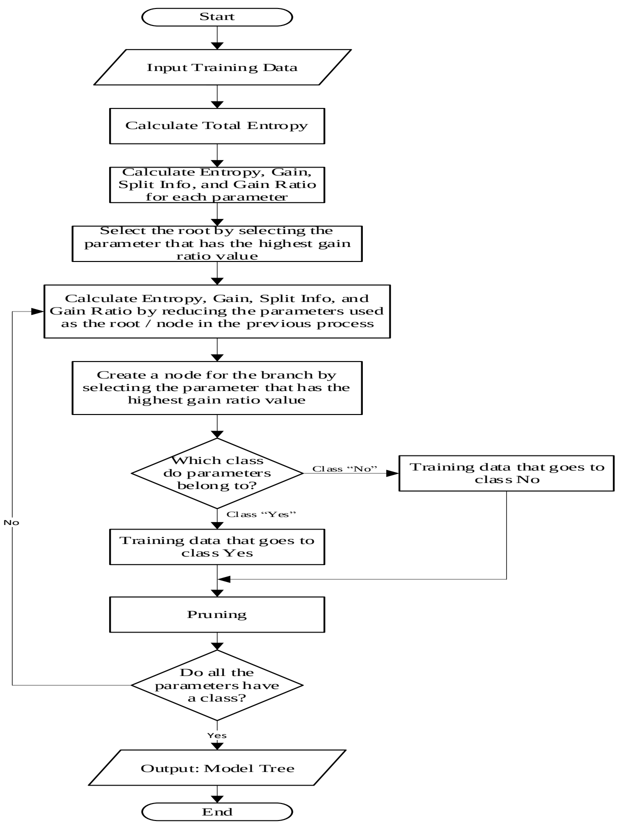
FIGURE 2. Database Monitoring System

The data in the td_formatwqms table will be used to calculate the Storet Method and training data. For the calculation of the Storet Method has been discussed in the previous sub-chapter. While the use of training data, the data formatting process will be carried out before it is stored in the td_trainingwqms table. One of the data formattings is changing the date attribute, which initially had the yyyy-mm-dd format to days with formats 1, 2, 3, up to 7, as in table V regarding the data simplification rules on this System.

TABLE V
RULES FOR SIMPLIFYING DATA

Condition Parameters	Value of the Sensor	Simplification	Value	Information
Day	2021-01-01	Friday	4	Monday = 1, Tuesday= 2, etc.
Time	00:00:00 01:00:00	1 AM 2 AM	0 1	Using 0 - 23 time format
Rain Status	Rain Dry	True False	1 0	-

ETL process flow for preparation of training data and test data can be seen in Figures IV.14 and IV.15. The total number of raw data that was successfully stored in the tmp_wqc table was 328,320 records, and the total number of data after the ETL process was complete was 54,720 records.



The data stored in the td_ trainingwqms table is ready to be used for C4.5 decision tree modeling. To model the C4.5 tree, in general, the steps taken are [35]:

1) Determine the root by counting
2) Create a branch for each value
3) Share cases within branches
4) Repeat the process for each branch until all cases on each branch and attributes have all classes.

Because the data generated by the sensor is continuous, so to determine the entropy of each Attribute, the Attribute must be discrete first. As in the ph parameter, The ph value is of type continuous, from 2 to 12. Then it must be counted how many pH records, which has a pH value of 2, 3, etc. so that the split value will be found at each pH value. It can be seen in table IV.8 about the number of pH parameter records and their respective values and searched for each value for the criteria less equal to (<=) and greater (>). Then it must be counted how many pH records, which has a pH value of 2, 3, etc. so that the split value will be found at each pH value. It can be seen in table IV.8 about the number of pH parameter records and their respective values and searched for each value for the criteria less equal to (<=) and greater (>).

```
tree
=====
turbidity > 1479: 1 {No=0, Yes=2}
turbidity ≤ 1479
|   suhuair > 30.500
|   |   tds > 1003.500: 1 {No=76, Yes=739}
|   |   tds ≤ 1003.500: 0 {No=1408, Yes=0}
|   suhuair ≤ 30.500
|   |   turbidity > 1438
|   |   |   tds > 1224: 1 {No=0, Yes=3}
|   |   |   tds ≤ 1224: 0 {No=5, Yes=0}
|   |   turbidity ≤ 1438
|   |   |   turbidity > 1427.500: 0 {No=3, Yes=1}
|   |   |   turbidity ≤ 1427.500
|   |   |   |   turbidity > 1194.500
|   |   |   |   |   suhuair > 29.500
|   |   |   |   |   |   ph > 9.500
|   |   |   |   |   |   |   turbidity > 1349.500
|   |   |   |   |   |   |   |   hujan = false: 0 {No=1, Yes=1}
|   |   |   |   |   |   |   |   hujan = true: 1 {No=0, Yes=2}
|   |   |   |   |   |   |   turbidity ≤ 1349.500: 1 {No=0, Yes=27}
|   |   |   |   |   |   ph ≤ 9.500: 0 {No=9, Yes=1}
|   |   |   |   |   suhuair ≤ 29.500
|   |   |   |   |   |   suhuair > 28.500
|   |   |   |   |   |   |   ph > 9.500
|   |   |   |   |   |   |   |   suhuudara > 29.500: 0 {No=14, Yes=0}
|   |   |   |   |   |   |   |   suhuudara ≤ 29.500: 1 {No=0, Yes=57}
|   |   |   |   |   |   |   ph ≤ 9.500: 0 {No=32, Yes=4}
|   |   |   |   |   |   suhuair ≤ 28.500
|   |   |   |   |   |   |   suhuair > 25.500
|   |   |   |   |   |   |   |   suhuudara > 26.500
|   |   |   |   |   |   |   |   |   suhuudara > 28.500: 0 {No=432, Yes=0}
|   |   |   |   |   |   |   |   |   suhuudara ≤ 28.500
|   |   |   |   |   |   |   |   |   |   suhuair > 27.500
|   |   |   |   |   |   |   |   |   |   |   ph > 9.500: 1 {No=0, Yes=39}
|   |   |   |   |   |   |   |   |   |   |   ph ≤ 9.500: 0 {No=10, Yes=1}
|   |   |   |   |   |   |   |   |   |   suhuair ≤ 27.500
|   |   |   |   |   |   |   |   |   |   |   suhuair > 26.500
|   |   |   |   |   |   |   |   |   |   |   |   suhuudara > 27.500: 0 {No=42, Yes=0}
|   |   |   |   |   |   |   |   |   |   |   |   suhuudara ≤ 27.500
|   |   |   |   |   |   |   |   |   |   |   |   |   ph > 9.500: 1 {No=3, Yes=17}
|   |   |   |   |   |   |   |   |   |   |   |   |   ph ≤ 9.500: 0 {No=5, Yes=0}
|   |   |   |   |   |   |   |   |   |   |   suhuair ≤ 26.500: 0 {No=182, Yes=0}
|   |   |   |   |   |   |   |   suhuudara ≤ 26.500
|   |   |   |   |   |   |   |   |   ph > 9.500: 1 {No=2, Yes=142}
|   |   |   |   |   |   |   |   |   ph ≤ 9.500: 0 {No=36, Yes=0}
|   |   |   |   |   |   |   suhuair ≤ 25.500: 0 {No=1670, Yes=4}
|   |   |   |   turbidity ≤ 1194.500: 0 {No=44147, Yes=131}
```

5. Using system

The use of IoT devices to record ditch and river water content is carried out at the points specified in Table I. An overview of the IoT devices that are built is as shown in Fig. 3.

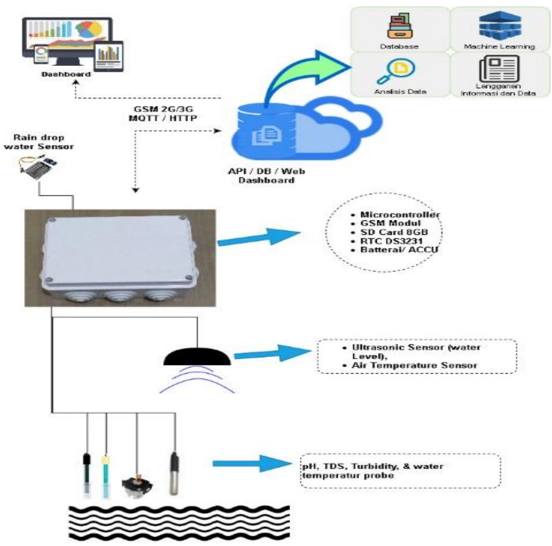


FIGURE 3. IoT-based ditch and river Water Quality Monitoring Syste

The data understanding process on this system uses the Cloud at the Infrastructure as a Service (IaaS) layer. Use this layer is because the Infrastructure-as-a-Service model provides resources in the form of computing power (CPU hour units), storage (gigabyte units), network bandwidth (Internet data transfer, GB per unit day), and electrical power (kWh). This IaaS model is based on application and resource characteristics so that it can reduce idle resources. [36].

In Figure 4, the Tree C4.5 modeling flowchart for the prediction of the highest pollution time. The tree modeling produced in this study can be seen in Fig. 4.

river_id	storet_date	wqc_min	wqc_max	wqc_average	storet_min	storet_max	storet_average	storet_total	storet_status
3	2020-03-16	1.2#18#87#212#	16.6#39#1420#1627#	8#26#618#667#	-2#0#0#0#	-2#-1#-1#-1#	0#-3#0#0#	-10	B
3	2020-03-17	2#16#217#218#	15.7#39#1414#1539#	9#26#701#730#	-2#0#0#0#	-2#-1#-1#-1#	0#-3#0#0#	-10	B
3	2020-03-18	1.9#16#214#214#	15.4#38#1345#1512#	9#25#771#793#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-19	1.1#16#82#225#	15.3#39#1439#1613#	9#25#658#736#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-20	2#16#82#223#	16#39#1353#1595#	8#24#671#696#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-21	0.8#16#107#204#	15.1#31#1413#1541#	8#24#664#740#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-22	1.6#16#95#232#	15.3#38#1421#1538#	9#25#627#680#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-23	1.6#16#179#224#	15.4#38#1421#1615#	8#25#708#746#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-24	1.2#16#126#230#	15.8#35#1380#1511#	9#24#828#917#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#-3#	-10	B
3	2020-03-25	2#16#78#228#	15.1#40#1356#1542#	9#25#718#753#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-26	1.3#16#118#234#	16.1#38#1394#1594#	9#25#726#790#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-27	2.3#16#104#229#	16.2#38#1409#1537#	10#25#782#832#	-2#0#0#0#	-2#-1#-1#-1#	-6#0#0#-3#	-16	C
3	2020-03-28	2.3#16#239#229#	15.2#41#1428#1546#	9#26#723#783#	-2#0#0#0#	-2#-1#-1#-1#	0#-3#0#0#	-10	B
3	2020-03-29	1.4#16#73#215#	14#37#1406#1531#	8#24#710#760#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-30	2.1#16#82#210#	16.7#41#1346#1541#	9#25#704#770#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-03-31	2.6#16#115#211#	15.8#38#1427#1501#	9#24#717#782#	-2#0#0#0#	-2#-1#-1#-1#	0#0#0#0#	-7	B
3	2020-04-01	1.5#16#265#236#	15.3#39#1097#1025#	8#26#446#548#	-2#0#0#0#	-2#-1#-1#-1#	0#-3#0#0#	-10	B
3	2020-04-02	1#16#161#239#	15.6#37#686#823#	8#24#431#522#	-2#0#0#0#	-2#-1#0#-1#	0#0#0#0#	-6	B

FIGURE 4. IoT-based ditch and river Water Quality Monitoring System

6. Discussion

The research carried out succeeded in building an Low-Cost IoT-based Ditch and River Water Quality Monitoring System.

The results of storet calculations carried out on the data that were successfully recorded at the ditch and river point using four parameters were only able to detect the maximum pollution of -25 or moderate pollution. The results of storet calculations can be seen in Fig. 5.

FIGURE 5. Result of Storet Calculation in the Downstream

While the use of machine learning, mostly the C4.5 decision tree algorithm for the prediction of the highest contamination day and time, is obtained from the highest confidence calculation results. The experimental data recording process can be seen in Fig. 7 and Fig. 8, and the results obtained can be seen in Fig. 9.

FIGURE 7. Data Recording Experiments Using IoT devices

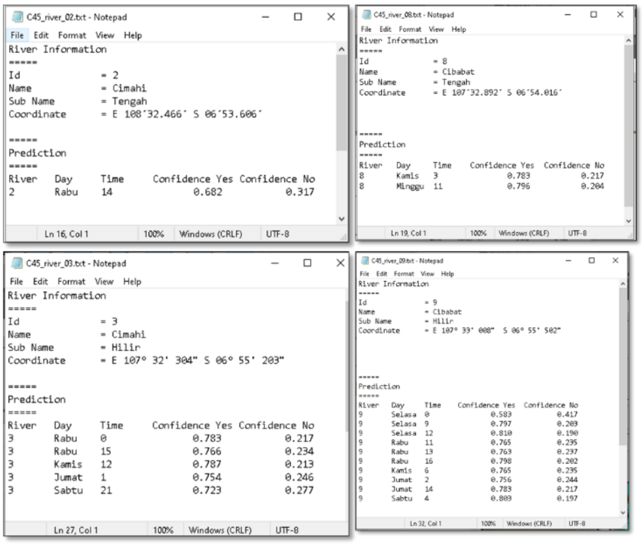


FIGURE 9. Data Recording Experiments Using IoT devices

7. conclusion

The technology that is suitable for monitoring ditch and river water quality at any time of need is the Internet of Things technology. With this technology, ditch and river water quality data can be recorded and stored for later processing using the storet calculation method to obtain ditch and river water quality levels. However, this research has only succeeded in obtaining the maximum detectable status, namely -25 or Moderate Pollution. This is due to the limited types of sensors used.

In this research, Monitoring was successfully developed. The System includes IoT tools, Rest API, and machine learning algorithm Decision Tree C4.5 according to business needs. This monitoring System can still be developed in future studies to complement its features.

To perform classification machine learning modeling in this case, the ETL process can be handled automatically with the software as a whole. This is due to the characteristics of ditch and river quality data, which have attributes, types, and formats that can be converted and standardized according to machine learning needs.

The machine learning model developed has been able to find data classes and predict the highest contamination time, which can be recommended to stakeholders to take ditch and river water samples used for water quality laboratory tests.

The implementation of this monitoring system is expected to contribute to improving the performance of the Local Government in monitoring and controlling ditch and river water pollution. It is hoped that it can improve ditch and river water quality for irrigation needs for agriculture, livestock (fish, poultry, and others) without worrying about being exposed to toxic waste from the domestics, and industrial sector.

References

1. D. Biello, "The World's 10 Most Polluted Places," Scientific American, a Division of Springer Nature America, Inc., 5 Nov 2013. [Online]. Available: https://www.scientificamerican.com/article/10-most-polluted-places-in-the-world1/. [Accessed August 2020].
2. G. Hutton, "Cleaning Up One of the World's Most Polluted Places," The World Bank Group, 19 Dec 2013. [Online]. Available: https://blogs.worldbank.org/water/cleaning-one-world-s-most-polluted-places. [Accessed August 2020].
3. Badan Pusat Statistik, "Statistik Daerah Kota Cimahi 2019," Kota Cimahi, 2019.
4. R. N. Febriani, "Pikiran Rakyat," 23 Mei 2018. [Online]. Available: https://www.pikiran-rakyat.com/bandung-raya/pr-01297435/cimahi-menuju-smart-city-layanan-berbasis-aplikasi-terus-dikembangkan-424845.
5. Pemerintah Kota Cimahi, "Pelaksanaan Program Citarum Harum di Kota Cimahi," Pemerintah Kota Cimahi, 3 Oktober 2019. [Online]. Available: https://cimahikota.go.id/artikel/detail/399. [Accessed 3 Agustus 2020].
6. D. M. A. Pradhan, S. Patankar, A. Shinde, V. Shivarkar, and P. Phadatare, "Smart City: Improvising Smart Environment," in Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, India, 2017.
7. [Jabarullah, N.H. & Othman, R. (2019) Steam reforming of shale gas over Al2O3 supported Ni-Cu nano-catalysts, Petroleum Science and Technology, 37 (4), 386 – 389.
8. M. Mukta, S. Islam, S. D. Barman, A. W. Reza and M. S. H. Khan, "IoT based Smart Water Quality Monitoring System," in IEEE 4th International Conference on Computer and Communication Systems, Singapore, 2019.
9. D. N. R. Morpathi, C. Mukesh and D. P. V. Sagar, "Water Quality Monitoring System Using IOT," in 4th International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB-18), 2018.
10. S. Jaloree, A. Rajput and S. Gour, "Decision tree approach to build a model for water quality," Binary Journal of Data Mining & Networking 4, pp. 25-28, 2014.
11. C. Gakii and J. Jepkoech, "A Classification Model for Water Quality Analysis Using Decision Tree," European Journal of Computer Science and Information Technology Vol.7, No.3, pp. 1-8, 2019.
12. J. A. Suyatno, F. Nhita, and A. A. Rohmawati, "Rainfall Forecasting in Bandung Regency Using C4.5 Algorithm," in 6th International Conference on Information and Communication Technology (ICoICT), Bandung, 2018.
13. P. R. "Peraturan Pemerintah Nomor 82 Tahun 2001 tentang Pengelolaan Kualitas Air Dan Pengendalian Pencemaran Air," PEMERINTAH REPUBLIK INDONESIA, 2001.
14. Y. C. Ching, Y. H. Lee, M. E. Toriman, M. Abdullah, and B. B. Yatim, "Effect of the big flood events on the water quality of the Muar River, Malaysia," 3 Juni 2015. [Online]. Available: https://doi.org/10.1007/s40899-015-0009-4.
15. H.-K. Bae, "Changes of River's Water Quality Responded to Rainfall Events," Environment and Ecology Research 1(1), pp. 21-25, 2013.
16. V. Novotny and O. H, Water Quality: Prevention, Identification, and Management of Diffuse Pollution, New York: van Nostrand Reinhold, 1994.
17. I. K. Sundra, "Studi kualitas perairan Sungai Nyuling di Karangasem ditijnjau dari aspek fisik kima dan mikrobiologi," J. Biologi. 5(1), pp. 9-20., 2001.
18. E. Yulia, E. Ekawati and E. M. Budi, "Perancangan Sistem Aktuator Laju Aliran Masuk dan Simulasi Proses Netralisasi pH pada Plant Simulator Instalasi Pengolahan Air Limbah Industri Tekstil (IPLIT)," in Seminar Nasional Instrumentasi, Kontrol dan Otomasi (SNIKO) 2018, Bandung, 2018.
19. J. C. Morril, R. C. Bales and M. H. Conklin, "Estimating Stream Temperature from Air Temperature: Implications for Future Water Quality," Journal of Environmental Engineering-ASCE - J ENVIRON ENG-ASCE. 131. 10.1061/(ASCE)0733-9372(2005), p. 131:1(139), 2005.
20. Y. Romdania, A. Herison, G. E. Susilo and E. Novilyansa, "Kajian Penggunaan Metode Ip, Storet, Dan Ccme WQI Dalam Menentukan Status Kualitas Air," Jurnal SPATIAL Wahana Komunikasi Dan Informasi Geografi, Vol 18 No 2, pp. 133 - 144, 2018.
21. S. P. Saraswati, S. B. A. Kironoto and S. Hadisusanto, "Kajian Bentuk Dan Sensitivitas Rumus Indeks Pi, Storet, CCME untuk Penentuan Status Mutu Perairan Sungai Tropis di Indonesia," J. MANUSIA DAN LINGKUNGAN, Vol. 21, No.2, pp. 129-142, 2014.
22. M. N. L. H. "Keputusan Menteri Negara Lingkungan Hidup Nomor 115 Tahun 2003 Tentang Pedoman Penentuan Status Mutu Air," Pemerintah RI, 2003.

23. D. Purwitasari, O. P. Putri and W. N. Khotimah, "ATURAN ASOSIASI DENGAN STANDAR STORET PADA MODEL PREDIKSI PARAMETER PENDUKUNG UJI KUALITAS AIR BAKU," Journal of Information Systems Engineering and Business Intelligence Vol. 1 No. 1, pp. 1-5, 2015.
24. A. I. Haryanto, Interviewee, Metode yang Dipakai di DLH Kota Cimahi. [Interview]. 16 April 2020.
25. B. Hariono, R. Wijaya, S. S. Anwar, and N. D. Wahyono, in International Conference on Applied Science and Technology (ICAST), Manado, Indonesia, 2018.
26. J. R. Quinlan, C4.5: Programs for Machine Learning, San Mateo, California: Morgan Kaufmann Publishers, 1993.
27. B. Santosa and A. Umam, Data Mining dan Big Data Analytics : Teori dan Implementasi Menggunakan Python & Apache Spark Edisi 2, Yogyakarta: Penebar Media Pustaka, 2018.
28. L. Rokach and O. Maimon, DATA-MINING WITH DECISION TREES Theory and Applications 2nd Edition, World Scientific Publishing Co. Pte. Ltd., 2015.
29. J. Han, M. Kamber, and J. Pei, Data Mining Concepts and Techniques Third Edition, Elsevier Inc., 2012.
30. L. Breiman, J. Friedman, C. J. Stone, and R. Olshen, Classification and Regression Trees, Wadsworth Int. Group, 1984.
31. J. Han and M. Kamber, Data Mining: Concept and Techniques, Morgan Kaufmann Publishers, 2006.
32. N. Shingala, “REST API and Its Advantages,” 2020. [Online]. Available: https://www.kovair.com/blog/rest-api-and-its-advantages/.
33. A. T., B. C. R., and P. D., “IoT Based Low-Cost System for Monitoring of Water Quality in Real-Time,” International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue:05, pp. 1658-1663, 2018.
34. G. Tchobanoglous, Teknik Sumber Daya Air, Edisi Ketiga, Jakarta: Erlangga, 1991.
35. E. P. Cynthia and E. Ismanto, "Metode Decision Tree Algoritma C.45 Dalam Mengklasifikasi Data Penjualan Bisnis Gerai Makanan Cepat Saji," Jurnal Riset Sistem Informasi Dan Teknik Informatika (JURASIK) Vol.3, pp. 1-13, 2018.
36. A. Widjajarto, S. H. Supangkat, Y. S. Gondokaryono and A. S. Prihatmanto, "Cloud Computing Reference Model: The Modelling of Service Availability Based on Application Profile and Resource Allocation," in 2012 International Conference on Cloud Computing and Social Networking (ICCCSN), Bandung, 2012.
37. Muposhi, A. (2019). EMERGENCE OF GREEN MARKETING CAPITALISM IN SOUTH AFRICA: IMPLICATIONS FOR GREEN ECONOMY AGENDA. International Journal of Business and Management Studies, 11(1), 1-15.