Application of AI to Filter Anomalous Data from Sensors in an Online Water Quality Monitoring System

Heru Dwi Wahyono[1,*] Satmoko Yudo[1]

[1] Research Center For Environmental and Clean Technology, National Research and Innovation Agency
[*]Corresponding author. Email: heru009@brin.go.id

ABSTRACT
The multiprobe sensor technology used in online water quality monitoring systems can produce water quality measurement data from several parameters at once quickly and in large quantities. The accuracy of the data is highly dependent on the quality of the river water being monitored and the performance of the probe on the sensor used. The worse the water quality and the decreased performance of the sensor probe, cause the reading of data by the sensor can produce inappropriate anomalous data. Anomalous data can cause water quality analysis to be invalid, therefore we need an Artificial Intelligence (AI) application to filter anomalous data by developing computational algorithms that can provide learning for computers to identify data generated and sent to data center servers. The algorithm method was developed using several mathematical logic models using real conditions and existing needs both in terms of regulations and historical data on water quality monitoring. By applying this algorithm, the computer can have artificial intelligence in analyzing data more accurately with monitoring data that matches the range of sensor measurement capabilities, so this method can guarantee the quality of online monitoring data.

Keywords: *Online water quality monitoring, Multiprobe water quality sensors, Artificial intelligence on water quality monitoring, Anomalous data filtering algorithms, Data quality assurance*

1. INTRODUCTION

The application of technology *online monitoring* water quality can support surface water quality monitoring activities such as rivers, lakes, swamps, beaches, or wastewater industrial continuously. The quality measurement data can be obtained at any time to support decision-making on water resources management policies in an area [1]. Some data must be analyzed using several methods to make it easier. understood by decision-makers as well as by the community.

Some of the water quality analysis methods that are widely used include the method pollution index (USA), the method storet (USA), and the CCME method (Canada), namely the water quality index (IKA) method for determining water quality status [2]. The method of calculating water analysis using a pollution index and storet method is widely used in Indonesia because it has been referred to by the Decree of the State Minister for the Environment No. 115 of 2003 concerning Guidelines for Determining Water Quality Status [3], in which Article 2 states that determining the status of water quality can use the Storetic Method or the Pollution Index Method.

The storet method has also been chosen as a method for analyzing automatic water quality which is used in the system *online monitoring* (Onlimo) of river water quality to provide information on the status of river water quality which is monitored using a multiprobe sensor [4]. This system has been implemented to monitor river water quality *online* and in *real-time* in several locations in Indonesia and has become a reference for several Regulations of the Minister of Environment and Forestry of the Republic of Indonesia related to monitoring water quality and wastewater in the network [5,6,7].

The use storet method for analyzing water quality data generated by system water quality monitoring *online* may be inaccurate if there are some anomalous data on the results of the measurement sensor. Anomaly

A. D. Saputro et al. (Eds.): ICOSEAT 2022, ABSR 26, pp. 623–631, 2023.
https://doi.org/10.2991/978-94-6463-086-2_83

data itself is defined as data from sensor readings that are far different from previous readings or are not in the range of the sensor reading capability [8]. Anomalous data can cause the computer to misinterpret the data referred to as normal, maximum, or minimum measurement numbers.

For example, the pH parameter has a measurement range of 0-14. The pH measurement data by the sensor should be in this data range. So that the determination of the maximum and minimum values of pH measurement results is still in this data range. If there is a pH measurement of -7 or 100, then this number is an anomalous number that will confuse the results of the analysis with the storet method, because this number will be considered the minimum pH measurement value is -7 and the maximum is 100.

Apart from pH, several sensors measure quality parameters water such as DO, Ammonia, and Nitrate also often produce anomalous data, especially if the sensor's performance has started to decline due to water conditions that are contaminated by mud and other impurities. To solve this problem, an algorithm is needed to learn for the computer to recognize whether the data is anomalous or not, so that more accurate analysis results can be produced.

Several techniques for detecting data anomalies have been proposed in the literature. For example, Zhang J. et al. [9] proposed an anomaly detection algorithm for water quality data, which can identify faulty data from historical patterns in real-time. The algorithm is based on a linear combination of autoregressive models, the prediction interval with a moving time window (moving time window), and verification strategy trace back (backtracking) [9].

Research conducted by Leigh C. et al. [10] develop methods framework/frameworks ten-step detecting anomaly data to implement and compare a range of methods anomaly data for the high frequency of water quality data measured using the in-situ sensor [10].

Research by Zhang D et al. [11], States that to detect and classify environmental data, abnormal sensor readings (also called anomalies) need to be extracted from the incoming data set. These abnormal sensor measurements were then grouped into events based on time. A set of features is extracted that are characteristic of different anomalies and are used to define individual events. Each feature set of detected anomalies is matched with a predefined codebook and the most suitable password is used to represent the feature [11].

Muharemi F. et al. [12], In their research, took an approach to detect events in water quality data series. This case study aims to find the best method for detecting anomalies in the water quality data system. Time series were analyzed using statistical algorithms for a long time. Today methods of machine learning algorithms*(MachineLearning)* are popular because they can work well on a series of data series [12].

This study aims to apply the technology of artificial intelligence (artificial intelligence) through the development of machine learning algorithms (machine learning) regarding data

anomalies generated by system online monitoring water quality (Onlimo) that has been applied in many locations in the rivers in Indonesia. Through this algorithm, the computer must be able to distinguish between real data from sensor measurements and anomalous data generated by noise electronic from the characteristics of sensors electrochemical. So that data analysis performed by the computer using the storet method can be more accurate.

2. MATERIALS AND METHODS

To build algorithms artificial intelligence in this study used some materials consisting of:

a. Database Onlimo on a computer server data center online water quality monitoring at *http://onlimo.bppt.go.id*
b. Data Online monitoring water quality with measurement time intervals of 60 minutes (1 hour) consisting of several days of monitoring and sourced from several stations online monitoring.
c. Data range sensor measurement and water quality standard data according to monitoring parameters using multi-probe sensors.
d. Annual average water quality data from monitoring station locations are used for analysis.
e. Personal computer for making utility programs to analyze water quality data using the storet method.
f. PHP, HTML, and programming languages Javascript to translate thinking algorithms from mathematical languages into computer programming languages.

The methods used to implement it in the online monitoring water quality system are as follows:

a. Review of literature on the use of the storet method to analyze water quality status.
b. Literature review of quality status and results of water quality monitoring in several locations on the river that is the target of algorithm testing.
c. Literature review of water quality standards according to water class and class stipulated by the government.
d. Identify several data determination methods for correcting anomalous data according to the characteristics of the monitored river location.
e. Preparation of artificial intelligence algorithms to distinguish real data and anomalous data.
f. Making computer application programs using a web-based programming language.
g. Testing the algorithm on several data series from

several online monitoring stations.

The following Figure 1 is a workflow schematic for building an AI Algorithm.

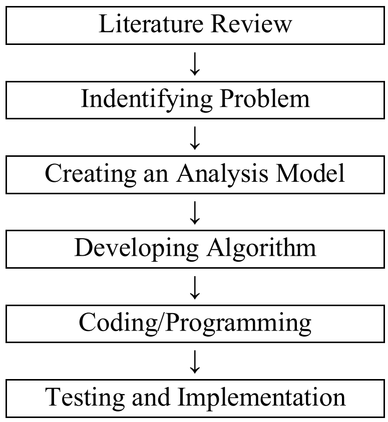


Figure 1 Schematic of Workflow to Build AI Algorithm

3. RESULTS AND DISCUSSION

3.1. Analysis of Water Quality Status Using the Storet Method

The storet method is a method that compares water quality data with water quality standards by the designation class stipulated in Government Regulation No. 82 of 2001 concerning Water Quality Management and Water Pollution Control [13].

The comparison with water class follows the method stipulated by the Ministry of Environment and Forestry in Minister of Environment Decree no. 115 2003 [14] about system ratings of US-EPA (Environmental Protection Agency).

This method gives a score for each water class, as follows:

1. Class A (Very Good): score = 0 is meets standard quality.
2. Class B (Good): score = -1 to -10 is lightly polluted.
3. Class C (Moderate): score = -11 to -30 is moderate.
4. Class D (Poor): score $\leq$ -31 is severely polluted.

The following Figure 2 is an example of the calculation results of the storet method to analyze the status of the water quality of the Citarum River at Onlimo Dayeuh Kolot Station in Bandung in October 2017.

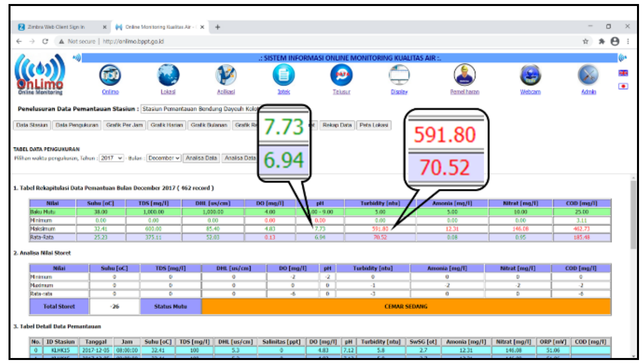


Figure 2 The status of Citarum River water quality at Onlimo Dayeuh Kolot Station in Bandung

Quality status is "Moderate" analyzed from approximately 462 monitoring data for 9 parameters (temperature, TDS, DHL, DO, pH, Turbidity, Ammonia, Nitrate & COD) during December 2017 [15]. Measurement results that are still within the standard quality range are displayed in font green and vice versa in font red. All the data obtained are still within range and the sensor measurement no anomalous data is found.

3.2. Algorithms as a Method of Implementing AI

According to the field of computer science, problem-solving is part of artificial intelligence which includes several techniques such as algorithms, heuristics to solve problems [16]. In this research, algorithm development becomes a method in applying artificial intelligence to analyze water quality status based on the only correct water quality data without anomalous data that is usually generated from sensor systems in the application of technology online monitoring.

In this study, the algorithm method used at the same time can be a data quality assurance method for data online monitoring water quality produced by the Onlimo system. The assessment and assurance of data quality are important requirements for more accurate data analysis because the quality of data analysis results directly depends on the quality of the underlying data [17].

Data quality assurance is the process of profiling data to find in consistencies and other anomalies in data, as well as carrying out data cleaning activities to improve data quality, for example removing foreign data, interpolating missing data [18]. The application of procedures data quality assurance can lead to improving data quality effectively [19].

The following Figure 3 is architecture data quality assurance that is used to develop an artificial intelligence algorithm in analyzing the water quality status of the water quality data obtained using an online water quality monitoring system. The learning method for a computer can be derived from the data series that is in the database and use the input oak Onlimo of from the existing regulation algorithms, programming, testing, and implementation.

The problem that will be solved by building this algorithm is replacing anomalous data found from the data series in the Onlimo database when compiling an analysis of water quality status. To identify data as an anomaly, it is necessary to analyze the data range of the sensor's ability to measure each parameter of water quality. According to Florback J. [21] states that anomalies are points or patterns in a data set that are different from expected “normal” behavior. Where the point anomaly is a data point that is outside the boundary of the normal area (outside the range).

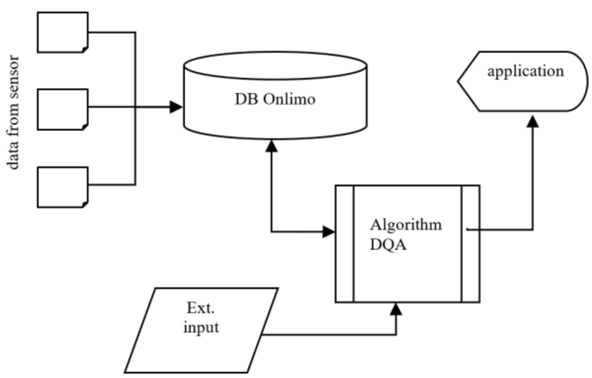


Figure 3 Architecture Data Quality Assurance (DQA)

3.3. Analysis and Compilation of Algorithms Data Quality Assurance in the Onlimo System

According to Sebastian Combefis [20], et al there are 4 main steps in designing the algorithm, namely 1). The understanding problem, 2). Kernel idea, 3). Execution and correctness analysis, and 4). Evaluation and efficiency analysis. These four steps can be developed according to system requirements. To compile an algorithm for data quality assurance on the Onlimo system, stages have been carried out, namely defining the problem, creating an analysis model, compiling design algorithms, programming, testing, and implementation.

The problem that will be solved by building this algorithm is replacing anomalous data found from the data series in the Onlimo database when compiling an analysis of water quality status. To identify data as an anomaly, it is necessary to analyze the data range *of* the sensor's ability to measure each parameter of water quality. According to Florback J. [21] states that anomalies are points or patterns in a data set that are different from expected “normal” behavior. Where the point anomaly is a data point that is outside the boundary of the normal area (outside the range).

In the onlimo system, there are several vendor's sensors that are used and have a range measurement as shown in the table 1 below.

Table 1. Sensor Measurement Range based on Anonym [22]

Para-meters	Sensor Measurement Range				*Range Filter*
	Aqua Read	*WQC-24*	*YSI 6820*	*OPM 1610*	
Temp	-5~50	-5~55	-5~50	-	-5~55
EC	0~200	0~100	0~100	-	0~200
TDS	0~100	0~100	-	-	0~100
Sal.	0~70	0~40	0~70	-	0~70
DO	0~50	0~20	0~50	-	0~50
pH	0~14	0~14	0~14	-	0~14
Turb.	0~3000	0~800	0~1000	-	0~3000
SwSG	0~50	0~50	-	-	0~50
Depth	0~100	0~100	0~61	-	0~100
Nitrate	0~30	0~62	0~0,2	-	0~62
NH_4	0~9	0~1,8	0~0,2	-	0~9
ORP	-2~2	-2~2	-0,9~0,9	-	-2~2
COD	-	-	-	0~500	0~500

Note: Temp (°C), EC (mS/cm), TDS (g/L), Sal. (ppt), DO (mg/L), pH, Turb. (NTU), SwSG (σt), Depth (m), Nitrate (g/L), NH_4 (g/L), ORP (V), COD (mg/L) (recap from sensor specifications)

To determine the minimum and maximum values that can cover all specifications of all vendors, sensor the value is used range measurement in the column Range Filter. This data value will then be used to filter the Onlimo data series from anomalous data. The algorithm design can be compiled from the mathematical sentence of the value range measurement of each parameter, namely "If the parameter value is less than the minimum value or more than the maximum value, then the parameter value is an anomaly". This sentence can be written as:

Param <Min or Param> Max => Param = Anomaly

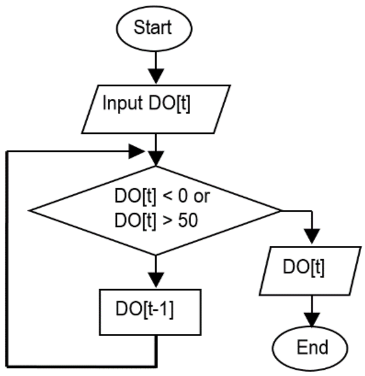
An example for the DO (parameter Dissolved Oxygen) is if the DO value is less than 0 or DO value is more than 50, so DO anomaly can be written as:

$0 < DO$ or $DO > 50$ then DO = Anomaly

Param=Anomaly (DO=Anomaly) in the math sentence above indicates that the parameter being processed is data that is outside the range measurement and must be corrected. There are two methods that can be used, namely:

- Set the value is the same as the previous value which is not anomaly data. So it becomes Param [t] = Param [t-1] (DO [t] = DO [t-1]).
- Specified by the value specified in the variable which is the annual average of that parameter. So it becomes Param [t] = DataParam (DO [t] = DataDO).

The design of the design flow diagram algorithm above is like a diagram with a program subroutine as shown in Figure 4 of the pseudocode below.




```
Pseudocode
Start
Input DO[t}
If DO [t] < 0 |
   DO [t] > 50 then
   DO [t]= DO [t-1]
else
   DO [t]= DO [t]
end if
End
```

Figure 4 Flow Diagram Algorithm Data Quality Assurance for DO Parameters

The design above traces the DO anomaly data and set the new value with the previous DO data which is not an anomaly in a closed iteration process. Iteration closed (Closed Loop Iteration) is one of the simplest methods in engineering drafting program algorithm to solve a problem/process [23]. DO [t-1] can be replaced by the DO value in the DO data variable. The algorithm design for the other parameters has the same process flow as the picture above.

To test the data quality assurance algorithm for all water quality parameters that have been designed, an application program is made by preparing several program functions, including:

- The image below is a s$param = [Temp, DHL, TDS, Sal, DO, pH, Turb, SwSG, Depth, Ammonia, Nitrate, ORP, COD]
- setting_bakumutu($param)
- setting_rangeparamter($param)
- $datasrc = (select*from tabel_pantau where idstasiun like id and tgl like datetime)
- $dataset = datasrc(idstasiun, tgl, $param)
- filter_anomali($dataset)

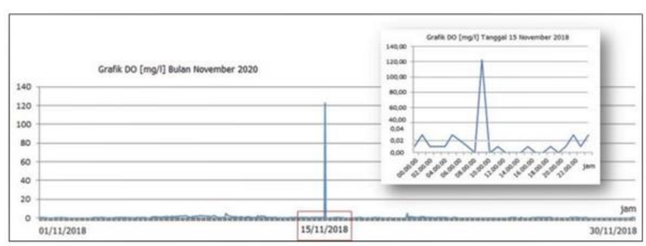
The Figure 5 is a snippet of the source code algorithm programming according to the algorithm design above to find DO anomaly data and fix it. At the initial stage, the param data in this example is DO filtering whether the data on the current record is an anomaly or not using predetermined variables, namely the minimum and maximum DO values. If the DO data is outside the range of the sensor's measurement capability, the data is anomalous and the initial DO value is not an anomaly. If the DO data is not an anomaly, the DO data will be the initial DO data for DO data in the process filter next.

```
                                 :
 // Filter Data if not Anomaly
 if ($record['DO'] >= $batasminimalDO &&
     $record['DO'] <= $batasmaksimalDO) {
     $datasebelumanomaliDO = $record['DO'];
 } else {
 // Filter Start Data for Data Anomaly
 if($firstDO) {
    if($record['DO'] < $batasminimalDO){
       $record['DO'] = $batasminimalDO;
       $datasebelumanomaliDO = $record['DO'];
    } else if($record['DO'] > $batasmaksimalDO){
           $record['DO'] = $batasmaksimalDO;
           $datasebelumanomaliDO = $record['DO'];
           } // Filter Data for next raw
    } else { $record['DO'] = $datasebelumanomaliDO; }
}
$jumlahDO += $record['DO'];
if($firstDO){
   $minimalDO = $record['DO'];
   $maksimalDO = $record['DO'];
   $firstDO = false;
} else {
  if($record['DO'] < $minimalDO){
     $minimalDO = $record['DO'];
  }
  if($record['DO'] > $maksimalDO){
     $maksimalDO = $record['DO'];
  }
  $firstDO = false;
}
                               :
```

Figure 5 Source Code Programming and Data Quality Assurance Algorithms [24]

3.4. Test Results of Algorithms Data Quality Assurance for Analysis of Water Quality Status Program

Testing is carried out on data from the monitoring results of the Onlimo system at the Ciliwung river monitoring station in Manggarai, South Jakarta, which has been operating since 2015. The following Figure 6 is a DO monitoring chart for November 2018 which shows anomalous data on 11/15/2018.



No.	ID Stasiun	Tanggal	Jam	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	Salinitas [ppt]	DO [mg/l]	pH	Turbidity [ntu]
330	KLHK2	2018-11-14	18:00:00	32.3	0	3.9	0	0.01	4.25	15.2
331	KLHK2	2018-11-14	19:00:00	32.1	0	4.2	0	0.43	3.93	14
332	KLHK2	2018-11-14	20:00:00	32.4	0	4.1	0	0.01	4.05	14.1
333	KLHK2	2018-11-14	21:00:00	32	0	4.4	0	0.15	3.93	22.7
334	KLHK2	2018-11-14	22:00:00	32.3	0	4.2	0	0.01	4.25	22.7
335	KLHK2	2018-11-14	23:00:00	31.8	0	4.6	0	0.13	4.01	22.8
336	KLHK2	2018-11-15	00:00:00	31.9	0	4.4	0	0.01	4.32	21.8
337	KLHK2	2018-11-15	01:00:00	31.5	0	4.6	0	0.03	4.18	20.2
338	KLHK2	2018-11-15	02:00:00	31.8	0	4.4	0	0.01	4.37	19.5
339	KLHK2	2018-11-15	03:00:00	31.4	0	4.5	0	0.01	4.17	19.8
340	KLHK2	2018-11-15	04:00:00	31.5	0	4.9	0	0.01	4.46	48.2
341	KLHK2	2018-11-15	05:00:00	31	0	4.8	0	0.03	4.07	20.3
342	KLHK2	2018-11-15	06:00:00	31.1	0	4.8	0	0.02	4.37	29.4
343	KLHK2	2018-11-15	07:00:00	30.6	0	5	0	0.01	4.29	23.8
344	KLHK2	2018-11-15	08:00:00	30.7	0	5	0	0	4.48	22.7
345	KLHK2	2018-11-15	09:00:00	30.4	0	5	0	122.88	4.24	22.2
346	KLHK2	2018-11-15	10:00:00	31	0	5.2	0	0	4.42	20.7
347	KLHK2	2018-11-15	11:00:00	31.5	0	5.3	0	0.01	4.27	19.2

Figure 6 Monitoring Graph of DO in November 2018 at

STO Manggarai [25]

The results of the quality status analysis of the Ciliwung River water quality data in November 2018 using the storet method are shown in Figure 7. The analysis was carried out automatically on 729 records of measurement data during November 2018 and it was known that the water quality status was **CEMAR SEDANG** (moderate polluted) with a maximum DO = 122.88 mg/l which occurred on 2018/11/15 09:00:00 and is anomalous data (range normal: 0 - 50 mg/l).

1. Tabel Rekapitulasi Data Pemantuan Bulan November 2018 (720 record)

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Baku Mutu	38.00	1,000.00	1,000.00	4.00	6.00 - 9.00	5.00	5.00	10.00	25.00
Minimum	27.30	0.00	1.50	0.00	0.03	2.90	1.00	0.64	0.00
Maksimum	34.60	0.00	12.40	122.88	7.01	170.40	3.05	20,000.00	0.00
Rata-Rata	30.49	0.00	6.64	0.46	5.40	34.27	2.09	1,545.27	0.00

2. Analisa Nilai Storet

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Minimum	0	0	0	-2	-2	0	0	0	0
Maximum	0	0	0	0	0	-1	0	-2	0
Rata-rata	0	0	0	-6	-6	-3	0	-6	0
Total Storet	-28	Status Mutu	CEMAR SEDANG						

Figure 7 Analysis of Ciliwung River Water Quality Status in November 2018 at STO Manggarai [25]

The algorithm implementation on the data series in November 2018 was able to correct the results of the analysis of the water quality status to be like the Figure 8 below. The maximum DO value to 5.26 mg/l is still below the quality standard of 10 mg/l. Correction of the measurement data to 0 mg/l is the same as the previous data on 11/15/2018 09:00:00.

1. Tabel Rekapitulasi Data Pemantuan Bulan November 2018 (720 record)

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Baku Mutu	38.00	1,000.00	1,000.00	4.00	6.00 - 9.00	5.00	5.00	10.00	25.00
Minimum	27.30	0.00	1.50	0.00	0.03	2.90	1.00	0.64	0.00
Maksimum	34.60	0.00	12.40	5.16	7.01	170.40	3.05	200.00	0.00
Rata-Rata	30.49	0.00	6.64	0.29	5.40	34.27	2.09	125.24	0.00

2. Analisa Nilai Storet

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Minimum	0	0	0	-2	-2	0	0	0	0
Maximum	0	0	0	0	0	-1	0	-2	0
Rata-rata	0	0	0	-6	-6	-3	0	-6	0
Total Storet	-28	Status Mutu	CEMAR SEDANG						

No.	ID Stasiun	Tanggal	Jam	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	Salinitas [ppt]	DO [mg/l]	pH	Turbidity [ntu]
330	KLHK2	2018-11-14	18:00:00	32.3	0	3.9	0	0.01	4.25	15.2
331	KLHK2	2018-11-14	19:00:00	32.1	0	4.2	0	0.43	3.93	14
332	KLHK2	2018-11-14	20:00:00	32.4	0	4.1	0	0.01	4.05	14.1
333	KLHK2	2018-11-14	21:00:00	32	0	4.4	0	0.15	3.93	22.7
334	KLHK2	2018-11-14	22:00:00	32.3	0	4.2	0	0.01	4.25	22.7
335	KLHK2	2018-11-14	23:00:00	31.8	0	4.6	0	0.13	4.01	22.8
336	KLHK2	2018-11-15	00:00:00	31.9	0	4.4	0	0.01	4.32	21.8
337	KLHK2	2018-11-15	01:00:00	31.5	0	4.6	0	0.03	4.18	20.2
338	KLHK2	2018-11-15	02:00:00	31.8	0	4.4	0	0.01	4.37	19.5
339	KLHK2	2018-11-15	03:00:00	31.4	0	4.5	0	0.01	4.17	19.8
340	KLHK2	2018-11-15	04:00:00	31.5	0	4.9	0	0.01	4.46	48.2
341	KLHK2	2018-11-15	05:00:00	31	0	4.8	0	0.03	4.07	20.3
342	KLHK2	2018-11-15	06:00:00	31.1	0	4.8	0	0.02	4.37	29.4
343	KLHK2	2018-11-15	07:00:00	30.6	0	5	0	0.01	4.29	23.8
344	KLHK2	2018-11-15	08:00:00	30.7	0	5	0	0	4.48	22.7
345	KLHK2	2018-11-15	09:00:00	30.4	0	5	0	0	4.24	22.2
346	KLHK2	2018-11-15	10:00:00	31	0	5.2	0	0	4.42	20.7
347	KLHK2	2018-11-15	11:00:00	31.5	0	5.3	0	0.01	4.27	19.2

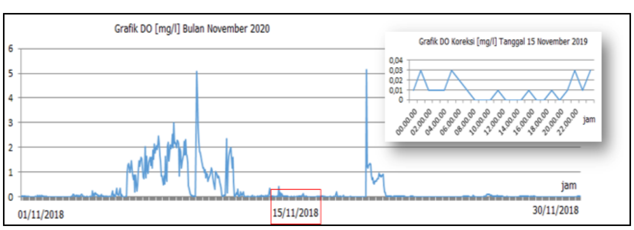
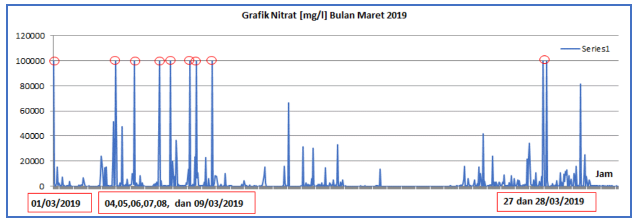


Figure 8 Correction of Anomaly Data in Analysis of Ciliwung River Water Quality Status in November 2018 at STO Manggarai [25]

Another program test was carried out on data from the monitoring results of the Onlimo system at the Serayu river monitoring station at Bendung Gerak Serayu Banyumas which has been operating since 2017. The following Figure 9 is a Nitrate monitoring chart for March 2019 which shows anomalous data on 1,4,5, 6,7,8,9,27, and March 28, 2019.



No.	ID Stasiun	Tanggal	Jam	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]
0	KLHK7	2019-03-01	00:00:00	27.4	100	13.9	4.71	7.99	30.9	0.38	99999
1	KLHK7	2019-03-01	01:00:00	27.3	100	14.2	4.77	8.42	23.8	0.32	2340
2	KLHK7	2019-03-01	02:00:00	27.3	100	14.3	4.52	7.88	26	0.3	29.9
3	KLHK7	2019-03-01	03:00:00	27.3	100	14.2	4.5	7.53	17.6	0.39	87.2
81	KLHK7	2019-03-04	09:00:00	26.5	100	14.5	4.67	7.81	39.2	0.26	1269
82	KLHK7	2019-03-04	10:00:00	26.5	100	14.6	4.88	10.02	46.3	0.3	99999
83	KLHK7	2019-03-04	11:00:00	26.6	100	14.6	4.4	7.72	46.4	0.25	13.66
153	KLHK7	2019-03-07	09:00:00	25.8	100	14.1	5.62	8.53	117.3	0.25	124.7
154	KLHK7	2019-03-07	10:00:00	25.6	100	14.4	5.68	7.7	106.1	0.24	99999
155	KLHK7	2019-03-07	11:00:00	25.6	100	14.7	5.43	8.18	104.8	0.2	625
187	KLHK7	2019-03-08	19:00:00	26.8	100	16.8	4.45	7.71	33.3	0.32	895
188	KLHK7	2019-03-08	20:00:00	26.8	100	16.5	4.77	8.08	31.5	0.56	99999
189	KLHK7	2019-03-08	21:00:00	26.9	100	16.3	4.77	9.3	31.1	0.52	6790

Figure 9 Graph of Nitrate Monitoring in March 2019 at STO Bendung Gerak Serayu [25]

The results of the quality status analysis in March 2019 which were calculated using the storet method are shown in Figure 10 below. The analysis was carried out automatically on 744 records measurement data during March 2019 and it was known that the water quality status was **CEMAR SEDANG** (moderate polluted) with a maximum value of Nitrate = 99.999 mg/l which occurred on several dates in March 2019 and is anomalous data (range normal: 0 - 62,000 mg/l).

1. Tabel Rekapitulasi Data Pemantuan Bulan March 2019 (744 record)

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Baku Mutu	38.00	1,000.00	1,000.00	4.00	6.00 - 9.00	5.00	5.00	10.00	25.00
Minimum	23.00	0.00	0.00	1.36	6.96	0.00	0.00	0.22	0.00
Maksimum	28.90	100.00	17.60	8.35	10.02	258.10	0.86	99,999.00	0.00
Rata-Rata	26.79	79.84	14.26	4.17	7.83	46.63	0.29	3,289.46	0.00

2. Analisa Nilai Storet

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Minimum	0	0	0	-2	0	0	0	0	0
Maximum	0	0	0	0	-2	-1	0	-2	0
Rata-rata	0	0	0	0	0	-3	0	-6	0
Total Storet	-16	Status Mutu	CEMAR SEDANG						

Figure 10 Analysis of Serayu River Water Quality Status in March 2019 at STO Bendung Gerak Serayu [25]

The results of applying the algorithm to the data series in March 2019 above have been able to correct the results of the analysis of the water quality status to be like the Figure 11. The maximum value of Nitrate to 200 mg/l is still above the quality standard of 10 mg/l. Correct the measurement data to:

- Date 01/03/2019 ◊ Nitrate 19.85 mg/l (same as data 28/02/2019 23:00:00) ·
- Date 04/03/2019 ◊ Nitrate 171.9 mg / l (same as data 04/03/2019 08:00:00) ·

- Date 05/03/2019 ◊ Nitrate 188,7 mg/l (same as data 05/03/2019 03:00:00) ·
- Date 06/03 /2019 ◊ Nitrate 45,6 mg/l (same as data 06/03/2019 18:00:00)
- Date 07/03/2019 ◊ Nitrate 124,7 mg/l (same as data 07/03/2019 09 :00:00) ·
- Date 08/03/2019 ◊ Nitrate 181,7 mg/l (same as data 08/03/2019 06:00:00) ·
- Date 09/03/2019 ◊ Nitrate 49,1 mg/l (same as data 09/03/2019 19:00:00)
- Date 27/03/2019 ◊ Nitrate 81,3 mg/l (same as data 27/03/2019 20:00:00)
- Date 28/03 /2019 ◊ Nitrate 134,2 mg/l (same as data 28/03/2019 00:00:00)

1. Tabel Rekapitulasi Data Pemantuan Bulan March 2019 (744 record)

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Baku Mutu	38.00	1,000.00	1,000.00	4.00	6.00 - 9.00	5.00	5.00	10.00	25.00
Minimum	23.00	0.00	0.00	1.36	6.96	0.00	0.00	0.22	0.00
Maksimum	28.90	100.00	17.60	6.84	10.02	258.10	0.86	200.00	0.00
Rata-Rata	26.79	79.84	14.26	4.16	7.83	46.63	0.29	61.84	0.00

2. Analisa Nilai Storet

Nilai	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]	COD [mg/l]
Minimum	0	0	0	-2	0	0	0	0	0
Maximum	0	0	0	0	-2	-1	0	-2	0
Rata-rata	0	0	0	0	0	-3	0	-6	0
Total Storet	-16	Status Mutu	CEMAR SEDANG						

No.	ID Stasiun	Tanggal	Jam	Suhu [oC]	TDS [mg/l]	DHL [us/cm]	DO [mg/l]	pH	Turbidity [ntu]	Amonia [mg/l]	Nitrat [mg/l]
0	KLHK7	2019-03-01	00:00:00	27.4	100	13.9	4.71	7.99	30.9	0.38	**19.85**
1	KLHK7	2019-03-01	01:00:00	27.3	100	14.2	4.77	8.42	23.8	0.32	2340
2	KLHK7	2019-03-01	02:00:00	27.3	100	14.3	4.52	7.88	26	0.3	29.9
3	KLHK7	2019-03-01	03:00:00	27.3	100	14.2	4.5	7.53	17.6	0.39	87.2
81	KLHK7	2019-03-04	09:00:00	26.5	100	14.5	4.67	7.81	39.2	0.26	1269
82	KLHK7	2019-03-04	10:00:00	26.5	100	14.6	4.88	10.02	46.3	0.3	**171.9**
83	KLHK7	2019-03-04	11:00:00	26.6	100	14.6	4.4	7.72	46.4	0.25	13.66
153	KLHK7	2019-03-07	09:00:00	25.8	100	14.1	5.62	8.53	117.3	0.25	124.7
154	KLHK7	2019-03-07	10:00:00	25.6	100	14.4	5.68	7.7	106.1	0.24	**124.7**
155	KLHK7	2019-03-07	11:00:00	25.6	100	14.7	5.43	8.18	104.8	0.2	625
187	KLHK7	2019-03-08	19:00:00	26.8	100	16.8	4.45	7.71	33.3	0.32	895
188	KLHK7	2019-03-08	20:00:00	26.8	100	16.5	4.77	8.08	31.5	0.56	**58.2**
189	KLHK7	2019-03-08	21:00:00	26.9	100	16.3	4.77	9.3	31.1	0.52	6790

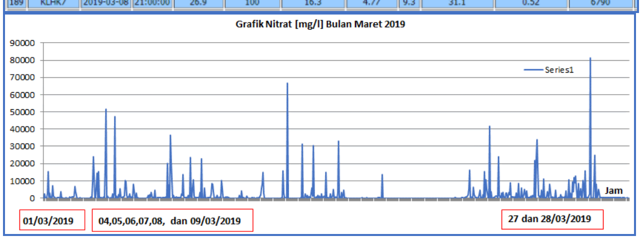


Figure 11 Correction of Anomalous Data in the Analysis of Serayu River Water Quality Status in March 2018 at STO Bendung Gerak Serayu [20,25]

4. CONCLUSION

Some things that can be concluded from the development of data quality assurance algorithms for analysis of water quality status in the application of online river water quality monitoring technology are as follows:

1. Measuring water quality using sensors in the long term can produce useless anomaly data, which can make the analysis of the water quality status inaccurate.
2. To ensure the quality of the data generated by this sensor, an artificial intelligence technique was developed by building a data quality assurance algorithm with several inputs related to the characteristics of the sensor and the characteristics of each water quality parameter.
3. The algorithm that has been developed in this study processes anomalous data into good data based on input criteria given by the user. In this study, the algorithm developed has been able to overcome the anomalous data that is present in one monthly measurement period to analyze the water quality status.

AUTHORS' CONTRIBUTIONS

HDW: coordinating research and development activities for online water quality monitoring systems by applying artificial intelligence algorithms to filter anomalous data, reviewing literature related to water quality status analysis according to national or international regulations, processing data management information based on expert system models into algorithm formulations, coordinating module development analysis of water quality status which has been equipped with anomalous data filtration algorithms, compiling the contents of the discussion of the material in the article.

SY: reviewing literature related to artificial intelligence using expert system algorithm development methods, preparing and processing online water quality monitoring data for testing anomaly data filter algorithms, analyzing the results of anomaly data filters and comparing with manual processes, making analysis charts and comparing data before and after being filtered with the algorithm created, compiling the literature used in the article.

ACKNOWLEDGEMENTS

The author expresses his gratitude to the Directorate of Water Pollution Control (PPA), Directorate General of PPKL, Ministry of Environment and Forestry (KLHK) who have collaborated with the Center for Environmental Technology (PTL), TPSA, BPPT since 2014 to date in supporting technology application activities. online monitoring that has been developed by BPPT.

We also thank the colleagues of the Onlimo Team who have assisted in carrying out R & D activities and implementing online and real-time water quality monitoring technology in polluted rivers in Priority Watershed in Indonesia.

REFERENCES

[1] Anonym, (2013). Technical Report Series No. 3.

Planning of Water-Quality Monitoring Systems. World Meteorological Organization.

[2] Sri Puji S., Sunyoto, Bambang AK, Suwarno H. "Kajian Kajian Bentuk Dan Sensitivitas Rumus Indeks Pi, Storet, Ccme Untuk Penentuan Status Mutu Perairan Sungai Tropis Di Indonesia", Jurnal Manusia dan Lingkungan 21(2) (2014) 129-142. DOI: 10.22146/jml.18536

[3] Keputusan Menteri Negara Lingkungan Hidup Nomor 115 Tahun 2003 Tentang Pedoman Penentuan Status Mutu air.

[4] Heru Dwi Wahjono. Pengembangan Sistem Sampling Air untuk Mengatasi Gangguan Lumpur pada Sistem Online Monitoring Kualitas Air Sungai. Jurnal Teknologi Lingkungan 20(1) (2019) 113-122. DOI: https://doi.org/10.29122/jtl.v20i1.3078

[5] Peraturan Menteri Lingkungan Hidup Dan Kehutanan Republik Indonesia Nomor P.93/Menlhk/Setjen/Kum.1/8/2018, Tahun 2018 Tentang Pemantauan Kualitas Air Limbah Secara Terus Menerus Dan Dalam Jaringan Bagi Usaha Dan/Atau Kegiatan.

[6] Peraturan Menteri Lingkungan Hidup Dan Kehutanan Republik Indonesia Nomor P.80/Menlhk/Setjen/ Kum.1/10/2019, Tahun 2019 Tentang Perubahan Atas Peraturan Menteri Lingkungan Hidup Dan Kehutanan Nomor P.93/Menlhk/Setjen/ Kum.1/8/2018 Tentang Pemantauan Kualitas Air Limbah Secara Terus Menerus Dan Dalam Jaringan Bagi Usaha Dan/Atau Kegiatan.

[7] Peraturan Menteri Lingkungan Hidup Dan Kehutanan Republik Indonesia Nomor P.7/Menlhk/Setjen/ Kum.1/1/2020, Tahun 2020 Tentang Penggunaan Dana Alokasi Khusus Fisik Penugasan Bidang Lingkungan Hidup Dan Kehutanan Tahun Anggaran 2020

[8] Zhang, D., Sullivan, T., Burghina, CB, Murphy, K., McGuinness, K., O'Connor, NE, Smeaton, A., and Regan, F. "Detection and Classification of Anomalous Events in Water Quality Datasets Within a Smart City - Smart Bay Project", International Journal on Advances in Intelligent Systems 7(1&2) (2014) 167-178.

[9] Zhang J., Chen X., Yue Y., and Wong, PW. A Real-time Anomaly Detection Algorithm for Water Quality Data. Computer Science 2017 Seventh International Conference on Innovative Computing Technology (INTECH) (2017) 36-41.

[10] C. Leigh, O. Alsibai, RJ Hyndman, S. Kandanaarachchi, OC King, JM McGree, C. Neelamraju, J. Strauss, P. Dilini Talagala, RDR Turner, K. Mengersen, EE Peterson. "A framework for automated anomaly detection in high-frequency water-quality data from in situ sensors. Science of The Total Environment,". Volume 664 (2019) 885-898.

[11] Zhang D., Sullivany T., Burghinay CB, Murphyy K., McGuinness K., O'Connor NE, Smeaton A., and Regan F., "Detection and Classification of Anomalous Events in Water Quality Datasets Within a Smart City Smart Bay Project", International Journal On Advance in Intelligent System, 7(1) (2014) 167-178.

[12] Muharem F., Logofatu D., Leon F., "Machine learning approaches for anomaly detection of water quality on a real-world data set". Journal of Information and Telecommunication. 3(3) (2019) 294-307.

[13] Anonym, "Peraturan Pemerintah No. 82 Tahun 2001 Tentang Pengelolaan Kualitas Air Dan Pengendalian Pencemaran Air", (2001).

[14] Anonym, "Keputusan Menteri Lingkungan Hidup No. 115 Tahun 2003 Tentang Pedoman Penentuan Status Mutu Air". Menteri Lingkungan Hidup (2003).

[15] Anonym, "Sistem Informasi Online Monitoring Kualitas Air, Pusat Teknologi Lingkungan BPPT". (2020). Available from: http://onlimo.bppt.go.id/ Access: 5th August 2020.

[16] Shrotriya S., Pandey A. "Imitating Humans A Technical Approach | Using AI Techniques". Lulu Press, Raleigh, NC, 2013, pp 36, (*www.lulu.com*).

[17] Lisa Ehrlinger, Woflram W. (2017). *Automated Data Quality Monitoring. 22nd MIT International Conference on Information Quality*, UA Little Rock. pp 1-19

[18] Anonym, (2005). *Data Quality*. Available from: *https://en.wikipedia.org/wiki/Data_quality#Data_quality_assurance* (Diakses tanggal 24 Desember 2020, jam 10:30)

[19] Arts DGT, De Keizer NF, Scheffer GJ, "Defining and Improving Data Quality in Medical Registries: A Literature Review, Case Study, and Generic Framework", Journal of the American Medical Informatics Association, 9(6) (2002) 600-611.

[20] Combéfis S., Barry SA, Crappe M., David M., "Learning and Teaching Algorithm Design and Optimisation Using Contests Tasks, Olympiads in Informatics", Vol. 11, Vilnius University, 2017, pp 19-28

[21] Florback J., Anomaly Detection in Logged Sensor Data. Master Thesis, Department of Applied Mechanics Chalmers University of Technology, Goteborg, Sweden, 2015.

[22] Anonym, "Aquaread, TOA-DKK, YESI, Horiba". Water Quality Meter Specification (2018).

[23] F. Panahifar, E. Murphy, and A. Ledwith, "A new framework for modelling schedules I complex and uncertain NPD projects", International Journal of Product Lifecycle Management, 11(1) (2018) 66-84.

[24] Heru DW, Aulia DS, dkk (2017). Penjaminan

Kualitas Data Online Monitoring Kualitas Air. Source code modul program filter anomali aplikasi situs web Onlimo. Availble from: http://onlimo.bppt.go.id/

[25] Pusat Teknologi Lingkungan, (2018). Sistem Informasi *Online Monitoring* Kualitas Air. BPPT. (http://onlimo.bppt.go.id).