

Prediksi Kondisi Pencemaran Air Sungai Citarum Berbasis Internet of Things dan Klasifikasi Naive Bayes

Annisa Marwa Nursantoso ^{#1}, Novian Anggis Suwastika ^{#2}, Rahmat Yasirandi ^{#3}

[#]*Prodi Informatika, Fakultas Informatika, Universitas Telkom*

Jl. Telekomunikasi No. 1 Terusan Buah Batu, Bandung, Indonesia

¹annisamn@student.telkomuniversity.ac.id

²anggis@telkomuniversity.ac.id

³batanganhitam@telkomuniversity.ac.id

Abstract

Water quality condition of Citarum river can be predict by compute and process of pH, temperature, and total dissolved solid (TDS) data of river's water. The tools and research related to predictions and predictions of air pollution are still very few. At the Citarum river area, a device has been implemented for reading pollution data, but this device has not been able to regulate and predict river water pollution. In this study a system was developed that was able to read Internet of Things (IoT) based river water readings and pollution prediction systems using the Naive Bayes classification. IoT implementation besides being able to carry out continuous data reading is also capable of sending data between devices via the internet network. Naif Bayes was able to classify river water pollution levels. The results of the implementation and testing of the system built using the Naive Bayes classification method in the river water pollution classification process have an accuracy rate of 96%. Taking from the accuracy result, this system can be predicted for river's water quality condition.

Keywords: Citarum river, Water quality, classification, Prediction, Naive Bayes, Internet of Things (IoT).

Abstrak

Kondisi pencemaran air sungai Citarum dapat ditentukan dan diprediksi dengan melakukan pembacaan dan pengolahan data pH, suhu, dan *Total Dissolved Solid* (TDS) air sungai. Perangkat dan penelitian yang berkaitan dengan penentuan dan prediksi pencemaran air masih sangat sedikit. Pada area sungai Citarum saat ini sudah diimplementasikan perangkat untuk pembacaan data pencemaran, tetapi pada perangkat ini belum dapat melakukan penentuan dan prediksi pencemaran air sungai. Pada penelitian ini dibangun sistem yang mampu melakukan pembacaan data air sungai secara terus-menerus berbasis *Internet of Things* (IoT) dan sistem prediksi pencemaran menggunakan klasifikasi *Naive Bayes*. Implementasi IoT selain mampu melakukan pembacaan data secara terus-menerus juga mampu mengirimkan data antar perangkat melalui jaringan internet. *Naive Bayes* mampu melakukan klasifikasi tingkat pencemaran air sungai. Hasil dari implementasi dan pengujian dari sistem yang dibangun menunjukkan penggunaan metode klasifikasi *Naive Bayes* pada proses klasifikasi pencemaran air sungai memiliki tingkat akurasi sebesar 96%. Dengan akurasi tersebut sistem ini dapat digunakan untuk memprediksi kondisi pencemaran air sungai.

Kata Kunci: Sungai Citarum, kualitas air, klasifikasi, Prediksi, *Naive Bayes*, *Internet of Things* (IoT).

I. PENDAHULUAN

MENURUT [23], sungai Citarum yang berada di Indonesia merupakan satu dari sepuluh sungai tercemar di Dunia. Sumber utama dari pencemaran sungai Citarum yaitu lebih dari 500 industri, sekitar 5 juta penduduk [18], dan tidak kurang dari 200ton limbah yang sebagian besar membuang zat sisa produksi langsung ke sungai, tanpa melalui proses Instalasi Pengolahan Air Limbah (IPAL) [13]. Hal itu mengakibatkan dampak yang sangat buruk dan meluas, terutama bagi kesehatan masyarakat yang berdekatan, maupun seluruh lingkungan. Serta terjadinya penurunan terhadap kualitas air [16] [4]. Upaya pemerintah terhadap sungai Citarum telah dilakukan selama 5 tahun dalam berbagai bentuk program, namun hasilnya belum signifikan [2]. Salah satu kendala dalam program untuk menanggulangi pencemaran air di sungai Citarum adalah tidak ada data dan informasi yang menunjukkan tingkat pencemaran dan lokasi pencemaran, serta penentuan dan prediksi pencemaran air sungai masih sedikit. Kondisi pencemaran air sungai Citarum dapat ditentukan dan di prediksi dengan melakukan pembacaan dan pengolahan data pH, suhu, dan TDS air sungai [5].

Berdasarkan permasalahan tersebut sistem ini menggunakan node sensor untuk dapat mengetahui dan memprediksi tingkat pencemaran air sungai berbasis IoT sehingga sistem mampu melakukan pembacaan data yang dibutuhkan dan mengirimkan data ke perangkat yang lain melalui jaringan internet. Selain itu, perlunya klasifikasi tingkat pencemaran air untuk mengetahui kualitas air dengan hasil tingkat akurasi yang lebih baik, karena Kualitas air juga menunjukkan ukuran kondisi air relatif terhadap kebutuhan biota air dan manusia [9]. Melalui penerapan teknologi IoT pada sistem ini, server yang ditanamkan algoritma *Naive Bayes* dapat melakukan prediksi secara otomatis dan terus-menerus. Pada penelitian kali ini metode *Naive Bayes* digunakan untuk melakukan klasifikasi tingkat pencemaran air sungai Citarum melalui data yang didapat oleh sensor node dalam jaringan IoT. *Naive Bayes* merupakan salah satu algoritma dalam teknik data *mining* yang menerapkan teori *Bayes* dalam klasifikasi yang mempunyai karakteristik asumsi yang sangat kuat (*naif*) akan independensi dari masing-masing kondisi atau kejadian [17] [26].

Selanjutnya, indikator parameter yang digunakan untuk mengetahui pencemaran kualitas air sungai yaitu *pH*, suhu dan TDS. Dengan sensor parameter yang digunakan, akan terhubung pada mikrokontroler arduino untuk mendeteksi kualitas air sungai, dan *module GPS* untuk menampilkan koordinat *latitude* dan *longitude* untuk mengetahui lokasi yang diuji. Lokasi yang diuji berbeda-beda tempat sesuai dengan titik pengujian di setiap aliran sungai Citarum berdasarkan peta topografi. Hasil akhirnya sistem yang dibangun mampu menampilkan nilai indikator setiap parameter dan hasil klasifikasi pencemaran air sungai dari deteksi lokasi sumber pencemaran. hasil klasifikasi berupa parameter yang dipakai, dengan kondisi baik, tercemar ringan, tercemar sedang dan tercemar berat.

II. STUDI TERKAIT

Teknik pengukuran parameter fisika dan kimia kualitas air pada [22] [5] telah ditentukan, parameter yang digunakan untuk mengukur kualitas air adalah suhu, derajat keasaman (*pH*) dan TDS (*Total Dissolved Solid/* total padatan terlarut) untuk mengetahui konduktivitas yang memberikan seberapa banyak kotoran dalam air.

Sistem pemantauan kualitas air dengan menggunakan teknologi *Internet of Things* (IoT) secara *realtime*, sistem menggunakan modul *Wi-Fi inbuilt* untuk memantau parameter seperti *pH*, kekeruhan dan konduktivitas. Sistem ini juga mencakup fasilitas peringatan, untuk memberi tahu pengguna tentang penyimpangan parameter kualitas air. Sistem ini juga menyediakan peringatan kepada pengguna jarak jauh, ketika ada penyimpangan parameter kualitas air [7]. Pada penelitian Shumei Wang et.al. [24], penerapan *Internet of Things* pada pemantauan kualitas air sungai perkotaan, Teluk Xinglin, Xiamen, Cina. Dengan menggunakan sistem manajemen kualitas air *online/Online Water Quality*

Management System (OWQMS), terbukti efektif dalam mengelola sungai dan menjaga kualitas air tetap stabil.

ROV *Water Quality System* [19], ide tentang ROV *Water Quality System* ini pemantauan kualitas air menggunakan kapal selam yang dikendalikan oleh radio yang berbentuk robot. ROV digunakan dalam penelitian untuk pengamatan sungai, laut dan yang lainnya. Merancang modifikasi ROV kecil untuk pemantauan kualitas air sungai. *Portable Water Quality Monitoring System* [20], sistem untuk pemantauan kualitas air dengan sebuah alat portable dan murah, sistem ini terdiri dari unit pengontrol utama yang didasarkan pada komputer papan COTS (*Commercial off-the-shelf*) tunggal dan dilengkapi dengan layar sentuh LCD untuk menampilkan informasi data, yang hasilnya dapat langsung dikirim ke server.

Hasil *paper* Hartatik et al. [8] telah melakukan studi mengenai perbandingan klasifikasi pencemaran air sungai dengan metode *Backpropagation* dan *Naive Bayes*. Kedua algoritma tersebut dibandingkan nilai keakuratannya dalam melakukan klasifikasi pencemaran air sungai. Dalam melakukan klasifikasi pencemaran air sungai hasilnya didapatkan dengan algoritma *Naive Bayes* memiliki nilai akurasi yang lebih baik yaitu 72,79% dibandingkan dengan menggunakan algoritma *Backpropagation* yang hanya memiliki nilai akurasi 64,75%.

Hasil *paper* Putrada et al. [12], telah melakukan studi mengenai peningkatan akurasi penggolong asap menggunakan metode *Naive Bayes* di *Internet of Things* (IoT). Metode klasifikasi *Naive Bayes* dipilih karena memiliki pemodelan dan hasil akurasi yang baik dalam set pelatihan data. Hasil penelitian menunjukkan penggunaan metode klasifikasi *Naive Bayes* pada proses klasifikasi memiliki rentang tingkat akurasi 88% hingga 91%. Hasil ini dapat diterima untuk akurasi klasifikasi.

Hasil *paper* M Zakizadeh et al. [14], telah melakukan studi mengenai penilaian kinerja metode klasifikasi dalam studi kualitas air, studi kasus di Sungai Karaj, Iran. Studi ini menjelaskan bahwa dengan menggunakan metode *Naive Bayes* ini dengan akurasi yang wajar dalam studi penelitian kualitas air. Analisis sensitivitas diimplementasikan menggunakan sepuluh variabel kualitas air (pH, COD, EC, TDA, kekeruhan, nitrat, nitrit, sulfat, TC, dan FC) untuk menemukan variabel yang paling penting dalam klasifikasi Sungai Karaj yang menunjukkan kekeruhan, di sebelah COD, pH, nitrat, dan sulfat, memiliki kontribusi terbesar dalam bidang ini.

Metode klasifikasi berbasis probabilitas dengan mengasumsikan diantara atribut atau fitur objek bersifat independen yaitu menggunakan *Naive Bayes Classifier* [3] [27]. Klasifikasi probabilitas sederhana yang berdasarkan pada teorema *Bayes*, asumsi bebas yang kuat [6]. *Naive Bayes* juga merupakan algoritma klasifikasi yang utama pada data *mining* dan banyak diterapkan dalam masalah klasifikasi kehidupan sehari-hari karena memiliki performa klasifikasi yang tinggi [25]. Teknologi *Internet of Things* menggunakan klasifikasi *Naive Bayes*, membahas mengenai sistem alarm kebakaran dengan menerapkan metode *Naive Bayes* untuk meningkatkan akurasi penggolongan asap pada lingkungan *Internet of Things* (IoT) [12] [15]. Hasil penelitian menunjukkan penggunaan metode klasifikasi *Naive Bayes* pada proses klasifikasi memiliki tingkat akurasi 88% hingga 91%.

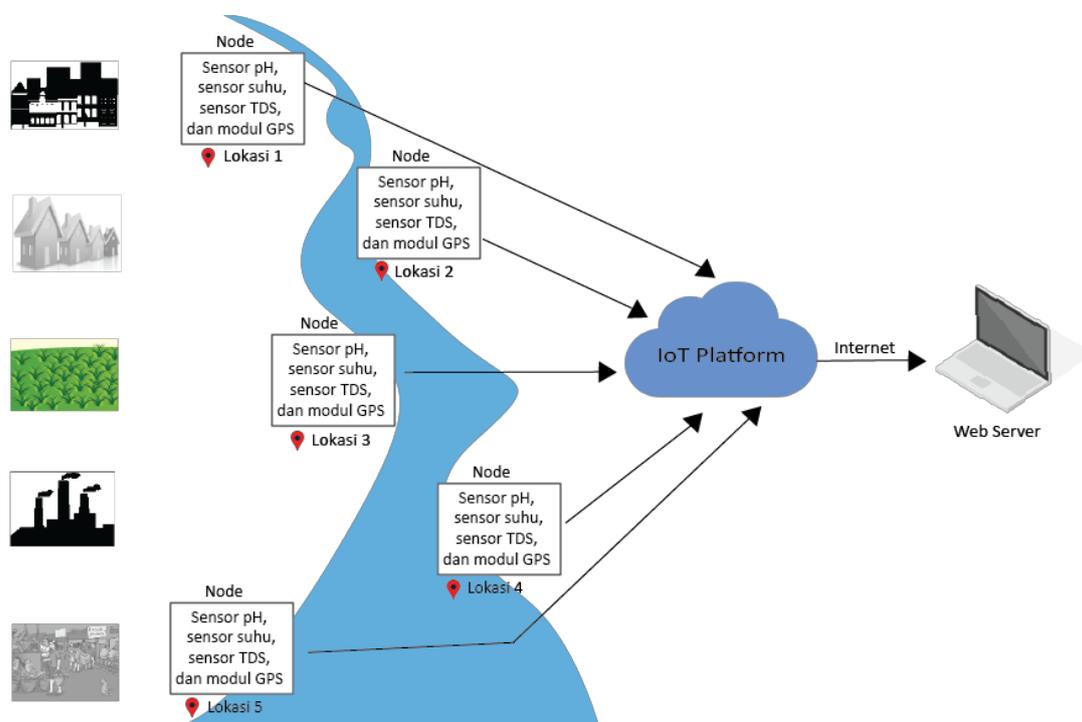
III. ANALISIS DAN PERANCANGAN SISTEM

A. Peninjauan Sistem

Dalam penelitian ini, keseluruhan arsitektur sistem didefinisikan untuk memenuhi tujuan berikut:

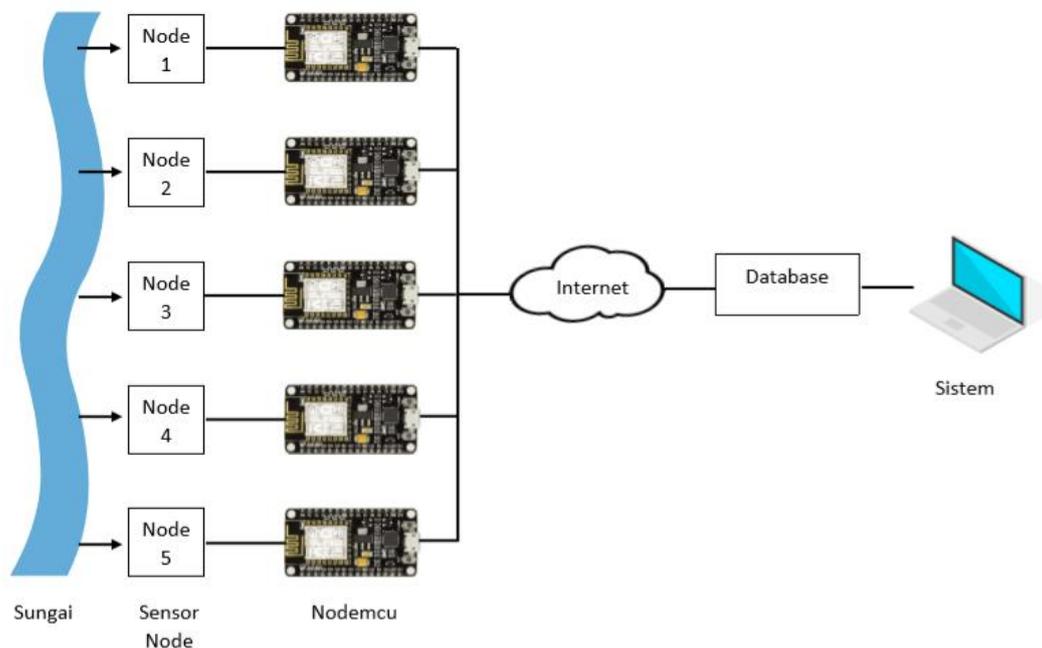
1. Membuat sistem deteksi lokasi pencemaran kualitas air sungai Citarum berbasis *Internet of Things* (IoT) menggunakan klasifikasi *Naive Bayes*.
2. Pengujian dilakukan untuk mendeteksi dan mengetahui lokasi tingkat pencemaran kualitas air sungai Citarum dengan indikator pH, suhu dan TDS penentu kualitas air.

Gambaran umum pada sistem deteksi lokasi pencemaran air sungai dapat dilihat pada Gambar 1.



Gambar 1. Gambaran Umum Sistem

Pada Gambar 1 merupakan gambaran umum sistem, bagaimana sistem dapat berjalan secara keseluruhan. Terdapat node yang diimplementasikan di sekitar aliran sungai untuk dapat mendeteksi lokasi dan tingkat pencemaran air sungai dengan sensor pH, suhu dan TDS, juga modul GPS yang berfungsi untuk mengetahui *latitude* dan *longitude* lokasi pengujian. Pada setiap node sensor terhubung dengan mikrokontroler untuk mengambil data, data disimpan pada *IoT platform* dengan koneksi internet atau *wifi*, data dikirim ke *database* untuk diklasifikasi dan mengetahui lokasi dan tingkat pencemaran kualitas air sungai, hasil data tersebut dapat ditampilkan melalui web server.



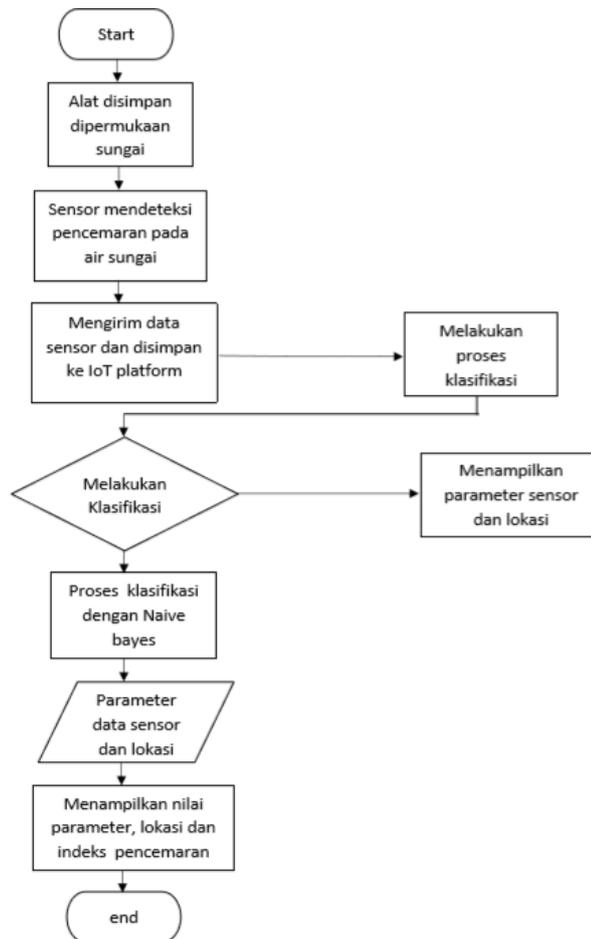
Gambar 2. Model Sistem

Pada Gambar 2 merupakan model sistem yang didalamnya terdapat bagian node sensor yang berfungsi sebagai pembaca nilai parameter kualitas air hingga menjadi data sensor dan satu modul GPS yang berfungsi untuk mengetahui *latitude* dan *longitude* pada lokasi. Bagian pengambilan data sensor oleh node kemudian dikirim melalui jaringan internet atau *wifi* dan disimpan ke *database*. Pada bagian sistem perangkat yang diproses lebih lanjut menjadi kategori tingkat pencemaran kualitas air sungai dengan klasifikasi dan lokasi untuk hasil akhir.

B. Diagram Alur Sistem

Terdapat diagram alur sistem pada sistem deteksi lokasi pencemaran air sungai Citarum ini untuk memperlihatkan urutan proses dalam suatu program agar berjalan dengan baik, diagram alur sistem dapat dilihat pada Gambar 3.

Pada Gambar 3 menjelaskan tentang alur proses pada sistem, terdapat proses untuk melakukan klasifikasi yang digunakan untuk mengetahui hasil klasifikasi dengan metode *Naive Bayes* dengan hasil kategori tingkat pencemaran kualitas air sungai Citarum. Sistem tersebut akan menampilkan hasil dengan nilai parameter yang digunakan yaitu pH, suhu dan TDS juga lokasi koordinat *longitude* dan *latitude*, untuk mengetahui lokasi yang tercemar dengan nilai parameter dan indeks kelas pencemaran air sungai.



Gambar 3. Diagram Alur Sistem

C. Parameter Kualitas Air

Kualitas air adalah kondisi kualitatif (kandungan nilai) air yang diukur atau diuji berdasarkan parameter tertentu dan metode tertentu (Pasal 1 keputusan Menteri Negara Lingkungan Hidup Nomor 115 tahun 2003). Kualitas air dapat dinyatakan dengan parameter kualitas air. Parameter ini meliputi parameter fisik, kimia, dan mikrobiologis [10].

Indikator yang sering kali digunakan dalam pemeriksaan pencemaran air adalah pH, oksigen terlarut (*Dissolved Oxygen*, DO), kebutuhan oksigen biokimia (*Biochemiycal Oxygen Demand*, BOD) serta kebutuhan oksigen kimiawi (*Chemical Oxygen Demand*, COD) [1]. Namun pada pengujian ini menggunakan beberapa parameter yang terdapat pada sensor air yaitu *potential of Hydrogen* (pH), suhu atau *Temperature* dan *Total Dissolved Solids* (TDS).

Pada Tabel I: merupakan nilai berdasarkan standar kualitas air yang dipakai dengan tingkat kondisi kualitas air. Pada *range* pH 6.5 – 8.5 menunjukkan kondisi alami sungai belum tercemar, *range* pH kurang dari 6 menunjukkan kondisi asam, dan *range* pH lebih dari 8 menunjukkan kondisi basa [1]. Pada *range* suhu 26°-32° menunjukkan kondisi netral, *range* suhu kurang dari 26° menunjukkan kondisi air dingin, dan *range* suhu lebih dari 32° menunjukkan kondisi air panas [21]. Pada *range* TDS 0 - 250

menunjukkan kondisi baik, *range* TDS 250 - 400 menunjukkan kondisi cukup dan *range* TDS lebih dari 400 menunjukkan kondisi buruk.

Tabel I. Nilai Standar Parameter Kualitas Air

Parameter Kualitas Air		
Nama Parameter	Nilai Range pH	Kondisi
pH	6.5 - 8.5 < 6 > 8	Kondisi Alami Kondisi Asam Kondisi Basa
Suhu	26 - 32 < 26 > 32	Netral Dingin Panas
TDS	0 - 250 250 - 400 > 400	Baik Cukup Buruk

D. Model Klasifikasi yang Digunakan

Naive Bayes Classifier merupakan salah satu metode di dalam data *mining* untuk mengklasifikasikan data. Cara kerja dari metode *Naive Bayes Classifier* menggunakan perhitungan probabilitas. Konsep dasar yang digunakan oleh *Naive Bayes* adalah teorema *Bayes*, yaitu teorema yang digunakan dalam statistika untuk menghitung suatu peluang, *Naive Bayes Classifier* menghitung peluang dari satu kelas dari masing-masing kelompok atribut yang ada dan menentukan kelas yang paling optimal. Proses pengelompokan atau klasifikasi dibagi menjadi dua fase yaitu *learning/training* dan *testing/classify*. Pada fase *learning*, sebagian data yang telah diketahui kelas, datanya diumpangkan untuk membentuk model perkiraan. Kemudian pada fase *testing*, model yang sudah terbentuk diuji dengan sebagian data.

Dasar dari *Naive Bayes* yang dipakai dalam pemrograman adalah rumus *Bayes* [15]:

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \quad (1)$$

Keterangan:

- X: Data dengan *class* yang belum diketahui
- H: Hipotesis data merupakan suatu *class* spesifik
- P(H|X): Probabilitas hipotesis H berdasarkan kondisi X (posteriori probabilitas)
- P(H): Probabilitas hipotesis H (Prior probabilitas)
- P(X|H): Probabilitas X berdasarkan kondisi pada hipotesis H
- P(X): Probabilitas X

Terdapat alur dari metode *Naive Bayes* digunakan untuk proses klasifikasi berikut:

1) Baca data *training*: Pada Tabel II. merupakan data latih atau data *training*, untuk menentukan data untuk dianalisis dengan metode naive bayes. Data *training* yang digunakan yaitu 1000 data, dengan masing-masing setiap kelas kondisi berjumlah 250 data. Kelas kondisi baik 250 data dari 1-250, kondisi tercemar ringan 250 data dari 251-500, kondisi tercemar sedang 250 data dari 501-750, dan kondisi tercemar berat 250 data dari 751-1000.

Tabel II. Data Training

No	pH	Suhu	TDS	Kondisi Kelas
1	7	31	95	Kondisi Baik
2	6	29	127	Kondisi Baik
...	Kondisi Baik
250	8	30	130	Kondisi Baik
251	8	27	259	Tercemar Ringan
252	7	31	264	Tercemar Ringan
...	Tercemar Ringan
500	6	28	384	Tercemar Ringan
501	28	33	300	Tercemar Sedang
502	8	31	783	Tercemar Sedang
...	Tercemar Sedang
750	5	30	363	Tercemar Sedang
751	4	30	949	Tercemar Berat
752	11	28	461	Tercemar Berat
...	Tercemar Berat
1000	5	18	496	Tercemar Berat

2) Menentukan *rules*: Pada klasifikasi *Naive Bayes* terdapat *rules* untuk menentukan nilai parameter yang masuk dalam kelas kategori kondisi air. Dengan melihat nilai standar parameter kualitas air pada Tabel I

Berikut merupakan penamaan label untuk nilai parameter kualitas air yang digunakan, terdapat tiga parameter kualitas air dan label setiap masing-masing nilai parameter yang digunakan:

- Nilai pH
 - a) 0 = Kondisi netral, *range* nilai= 6 - 8
 - b) 1 = Kondisi asam, *range* nilai = 0 - 5
 - c) 1 = Kondisi basa, *range* nilai = 9 - 14
- Nilai suhu
 - a) 0 = Kondisi baik, *range* nilai = 26° - 32°
 - b) 1 = Kondisi dingin, *range* nilai = < 26°
 - c) 1 = Kondisi panas, *range* nilai = > 32°
- Nilai TDS
 - a) 0 = Kondisi baik, *range* nilai = 0 - 250
 - b) 1 = Kondisi cukup, *range* nilai = 250 - 400
 - c) 2 = Kondisi buruk, *range* nilai = > 400

Berikut merupakan penamaan untuk label kelas kategori kondisi air, terdapat empat label dan kelas kategori kondisi air yang digunakan yaitu kondisi baik, tercemar ringan, tercemar sedang dan tercemar berat.

- 1) 0 = Kondisi baik
- 2) 1 = Tercemar ringan
- 3) 2 = Tercemar sedang
- 4) 3 = Tercemar berat

Penamaan label kelas dan label parameter untuk menentukan perhitungan *rules* pada *Naive Bayes*, *rules* digunakan untuk menganalisis data kualitas air yang ditentukan pada data *training*. Tabel III: merupakan hasil *rules* untuk perhitungan *Naive Bayes* yang telah ditentukan.

Tabel III. Rules Naive Bayes

Label pH	Label Suhu	Label TDS	Label Kelas
0	0	0	0
0	0	1	1
0	0	2	2
0	1	0	1
0	1	1	2
0	1	2	3
1	0	0	1
1	0	1	2
1	0	2	3
1	1	0	2
1	1	1	2
1	1	2	3

3) Kondisi dan probabilitas: Pada *Naive Bayes* terdapat nilai probabilitas setiap nilai parameter yang didapatkan dari data latih atau data *training* pada Tabell.

Adapun nilai probabilitas dengan kategori kondisi air dapat dilihat pada Tabel IV dan nilai probabilitas setiap parameter dapat dilihat pada Tabel V nilai pH, Tabel VI nilai suhu, dan Tabel VII nilai TDS berikut.

Tabel IV. Probabilitas Kategori Kelas Kondisi Air

Jumlah Kejadian "Kondisi Air"				Probabilitas			
Baik	Ringan	Sedang	Buruk	Baik	Ringan	Sedang	Buruk
250	250	250	250	0.25	0.25	0.25	0.25

Tabel IV dapat dilihat nilai probabilitas untuk masing-masing kelas kondisi air baik, ringan, sedang dan buruk menghasilkan nilai probabilitas yang sama yaitu 0.25. Berdasarkan 1000 data *training* untuk masing-masing jumlah kejadian kondisi air yaitu 250.

Tabel V. Probabilitas Nilai PH

Keadaan PH	Jumlah Kejadian "Kondisi Air"				Probabilitas			
	Baik	Ringan	Sedang	Buruk	Baik	Ringan	Sedang	Buruk
Netral	250	166	100	83	0.996	0.6667	0.2024	0.3333
Asam dan Basa	0	84	150	167	0.004	0.3333	0.7976	0.6667
Jumlah	250	250	250	250	1	1	1	1

Pada Tabel V: merupakan perhitungan probabilitas nilai pH. Dapat dilihat bahwa Keadaan pH netral dalam kondisi baik memiliki probabilitas 0.996, pH netral kondisi ringan memiliki probabilitas 0.6667, pH netral kondisi sedang memiliki probabilitas 0.2024, pH netral kondisi buruk 0.3333. Keadaan pH asam dan basa dalam kondisi baik memiliki probabilitas 0.004, pH asam dan basa kondisi ringan memiliki probabilitas 0.3333, pH asam dan basa kondisi sedang memiliki probabilitas 0.7976, pH asam dan basa kondisi buruk 0.6667.

Tabel VI. Probabilitas Nilai Suhu

Keadaan Suhu	Jumlah Kejadian "Kondisi Air"				Probabilitas			
	Baik	Ringan	Sedang	Buruk	Baik	Ringan	Sedang	Buruk
Baik	250	167	100	83	0.996	0.6667	0.3532	0.3333
Dingin dan Panas	0	83	150	167	0.004	0.3333	0.6468	0.6667
Jumlah	250	250	250	250	1	1	1	1

Pada Tabel VI: merupakan perhitungan probabilitas nilai suhu. Dapat dilihat bahwa Keadaan suhu baik dalam kondisi baik memiliki probabilitas 0.996, suhu baik kondisi ringan memiliki probabilitas 0.6667, suhu baik kondisi sedang memiliki probabilitas 0.3532, suhu baik kondisi buruk 0.3333.

Keadaan suhu dingin dan panas dalam kondisi baik memiliki probabilitas 0.004, suhu dingin dan panas kondisi ringan memiliki probabilitas 0.3333, suhu dingin dan panas kondisi sedang memiliki probabilitas 0.6468, suhu dingin dan panas kondisi buruk 0.6667.

Tabel VII. Probabilitas Nilai TDS

Keadaan TDS	Jumlah Kejadian "Kondisi Air"				Probabilitas			
	Baik	Ringan	Sedang	Buruk	Baik	Ringan	Sedang	Buruk
Baik	250	167	50	0	0.9921	0.664	0.2016	0.0039
Cukup	0	83	150	0	0.0039	0.332	0.5969	0.0039
Buruk	0	0	50	250	0.004	0.004	0.2015	0.9921
Jumlah	250	250	250	250	1	1	1	1

Pada tabel VII: merupakan probabilitas nilai TDS. Dapat dilihat bahwa Keadaan TDS baik dalam kondisi baik memiliki probabilitas 0.9921, TDS baik kondisi ringan memiliki probabilitas 0.664, TDS baik kondisi sedang memiliki probabilitas 0.2016, TDS baik kondisi buruk 0.0039. Keadaan TDS cukup kondisi baik memiliki probabilitas 0.0039, TDS cukup kondisi ringan memiliki probabilitas 0.332, TDS cukup kondisi sedang memiliki probabilitas 0.5969, TDS cukup kondisi buruk 0.0039. Keadaan TDS buruk kondisi baik memiliki probabilitas 0.004, TDS buruk kondisi ringan memiliki probabilitas 0.004, TDS buruk kondisi sedang memiliki probabilitas 0.2015, TDS buruk kondisi buruk 0.9921.

4) Melakukan pengujian *testing* atau probabilitas akhir yang menghasilkan solusi klasifikasi

Probabilitas akhir atau solusi akhir yang didapat digunakan untuk menentukan tingkat pencemaran kondisi air sungai yang termasuk dalam kelas kondisi baik, tercemar ringan, tercemar sedang dan tercemar buruk untuk data yang ingin diuji.

E. Pengukuran Tingkat Akurasi

Dalam melakukan klasifikasi diharapkan dapat melakukan klasifikasi dengan benar, sehingga dalam klasifikasi harus diukur tingkat akurasi. Akurasi didapatkan dari jumlah data yang diprediksi benar dibagi dengan jumlah data yang diprediksi [11]. Berikut merupakan formula untuk menghitung tingkat akurasi menggunakan persamaan (2).

$$Akurasi = \frac{\sum i}{\sum n} * 100\% \quad (2)$$

Keterangan:

- $\sum i$ = Jumlah data yang diprediksi benar
- $\sum n$ = Jumlah data yang diprediksi

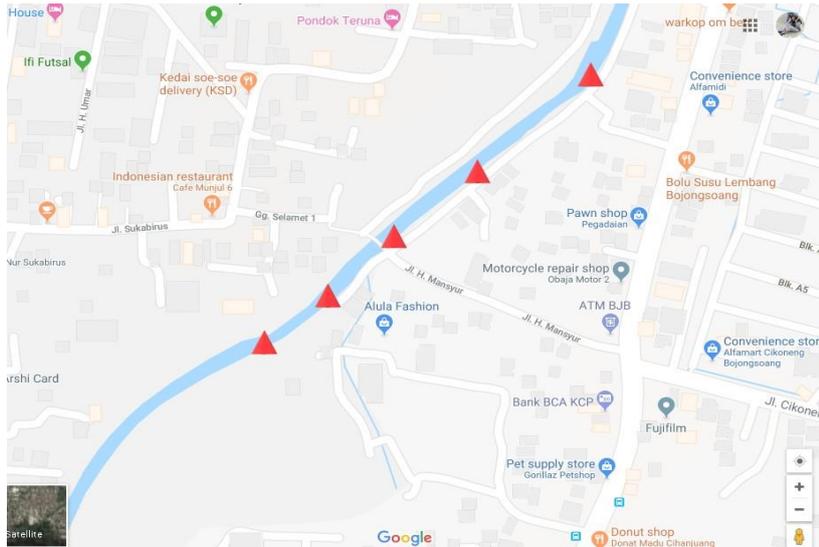
IV. HASIL DAN PEMBAHASAN

Pada pengujian sistem ini terdapat 5 node sensor yang tersebar di sepanjang aliran sungai Citarum di beberapa titik lokasi sesuai lokasi pengujian dan akan mengambil data dari sensor air. Titik-titik lokasi dapat dilihat pada Gambar 5: merupakan titik lokasi yang digunakan untuk pengujian daerah Citeureup, Dayeuhkolot, Bandung.

Adapun titik *latitude* dan *longitude* dari tiap lokasi dapat dilihat pada Tabel VIII: yang tersebar di sepanjang aliran sungai Citarum daerah Citeureup, Dayeuhkolot, Bandung.

Tabel VIII. Titik Latitude dan Longitude lokasi node sensor

No	Latitude	Longitude
1	-6.979747	107.632318
2	-6.979191	107.632931
3	-6.978432	107.633829
4	-6.980187	107.631896
5	-6.980449	107.631611



Gambar 5. Peta Lokasi node sensor

A. Skenario Hasil Pengujian Alat

Pada pengujian ini dilakukan dengan menguji fungsionalitas pada pengujian alat sensor yang dipakai. Pengujian alat sensor dipakai untuk mengukur pencemaran kualitas air dengan mengukur parameter sensor pH, suhu dan TDS, hasilnya dibandingkan dengan hasil pengukuran dari alat pengukur baku *hydrometer*, pH, suhu dan TDS.

Tabel IX merupakan hasil pengujian sensor pH dan alat ukur pH

Tabel IX. hasil pengujian sensor pH

No	Lokasi	Sensor pH	pH hydrometer	Error
1	Node 1	8.4	9.1	7.69%
2	Node 2	8.19	8.1	1.2%
3	Node 3	7.88	8	1.5%
4	Node 4	6.98	7.27	3.98%
5	Node 5	7.48	7.58	1.31%
Error Rata-rata				1.56%

Tabel X merupakan hasil pengujian sensor suhu dan alat ukur suhu

Tabel X. hasil pengujian sensor Suhu

No	Lokasi	Sensor Suhu	Suhu hydrometer	Error
1	Node 1	27	27.7	2.52%
2	Node 2	26.5	27.6	3.98%
3	Node 3	27.5	28	1.78%
4	Node 4	26.6	27	1.78%
5	Node 5	26.38	27	1.48%
Error Rata-rata				1.21%

Tabel XI merupakan hasil pengujian sensor TDS dan alat ukur TDS

Tabel XI. hasil pengujian sensor TDS

No	Lokasi	Sensor TDS	TDS hydrometer	Error
1	Node 1	980	893	9.74%
2	Node 2	425.76	536	3.04%
3	Node 3	354.25	383	7.50%
4	Node 4	357.64	395	9.4%
5	Node 5	351.12	381	7.84%
Error Rata-rata				3.76%

Hasil dari ke 5 node atau 5 lokasi pengujian, pengujian sensor pH dibandingkan dengan pH meter memiliki rata-rata *error* 1.56%, pengujian sensor suhu dengan suhu meter memiliki rata-rata *error* 1.21% dan pengujian sensor TDS dan TDS meter memiliki rata-rata *error* sebesar 3.76%. Sehingga dapat dikatakan sensor pH, suhu dan TDS dapat bekerja dengan baik yang memiliki tingkat kesalahan yang kecil.

B. Skenario Hasil Pengujian Proses Sistem

Pada pengujian ini dilakukan proses pembacaan data yang dikirim dari perangkat ke server untuk mengetahui banyaknya data dan waktu yang didapatkan pada pengujian yang kemudian dilakukan untuk klasifikasi lanjut. Pengujian proses sistem ini dilakukan dengan 5 node perangkat dan server sistem dari hasil pengujian. Hasil pengujian proses sistem dapat dilihat pada Tabel XII dan Tabel XIII.

Tabel XII. Rata-rata Pembacaan 5 node

Jumlah Node	Jumlah Semua Data	Rata-rata Jumlah Data	Waktu
5	506	109	800 detik = 13 menit 20 detik

Pada Tabel XII dapat dilihat hasil pengujian dalam waktu 800 detik atau 13 menit 20 detik menghasilkan data dari semua 5 node yang berjumlah 506 data. Dapat dikatakan bahwa setiap node memiliki jumlah data dengan rata-rata 109 data yang didapatkan.

Tabel XIII. Hasil Pengujian Pengiriman Data ke Server

Node	Jumlah Data	waktu
1	113	166 detik = 2 menit 46 detik
2	105	193 detik = 3 menit 13 detik
3	106	114 detik = 1 menit 54 detik
4	108	167 detik = 2 menit 47 detik
5	74	160 detik = 2 menit 40 detik

Pada Tabel XIII hasil pengujian pengiriman data dari setiap node perangkat ke server sistem. Dapat dilihat jumlah data yang didapatkan dari node pertama dalam waktu 166 detik atau 2 menit 46 detik berjumlah 113 data. Node kedua dalam waktu 193 detik atau 3 menit 13 detik data yang didapatkan berjumlah 105 data. Node ketiga dalam waktu 114 detik atau 1 menit 54 detik data yang didapatkan berjumlah 106 data. Node keempat dalam waktu 167 detik atau 2 menit 47 detik data yang didapatkan berjumlah 108 data. Dan node kelima dalam waktu 160 detik atau 2 menit 40 detik data yang didapatkan berjumlah 74 data. Hal tersebut dapat dikatakan bahwa node kelima memiliki jumlah data yang sedikit dibandingkan dengan node yang lainnya.

C. Hasil Pengujian Klasifikasi *Naive Bayes*

Pada pengujian ini bertujuan untuk mengetahui hasil prediksi tingkat pencemaran air sungai. Prediksi dilakukan dengan metode klasifikasi *Naive Bayes* dengan data *testing* atau data uji yang didapatkan oleh perangkat dari 5 lokasi pengujian yang dikirim ke server. Hasil pengujian data *testing* dapat dilihat pada Tabel XIV.

Tabel XIV. Pengujian Data *Testing*

Lokasi	Waktu	Nilai pH	Nilai suhu	Nilai TDS	Latitude	Longitude	Tingkat Pencemaran
1	12:25:05	8.4	27	1027.1	-6.979747	107.632318	Tercemar Berat
2	12:26:00	8.19	26.5	525.75	-6.979191	107.632931	Tercemar Berat
3	12:26:05	7.88	27.5	334.25	-6.978432	107.633829	Tercemar Ringan
4	12:27:00	6.98	26.5	357.64	-6.980187	107.631896	Tercemar Ringan
5	12:27:05	7.48	26.38	351.12	-6.980449	107.631611	Tercemar Ringan

Tabel XIV hasil akhir menampilkan informasi kondisi tingkat pencemaran air sungai Citarum dari proses pengujian dan klasifikasi akhir, data informasi yang ditampilkan yaitu nilai pH, nilai suhu, nilai TDS, *latitude* dan *longitude* lokasi pengujian, dan tingkat pencemaran air sungai citarum dengan kondisi lokasi 1 dan 2 tercemar berat, lokasi 3, 4, dan 5 dengan kondisi tingkat pencemaran air sungai yang tercemar ringan.

D. Hasil Uji Akurasi *Naive Bayes*

Dalam melakukan prediksi klasifikasi sistem dapat melakukan klasifikasi pada semua objek dengan benar. Untuk menghitung tingkat akurasi (ketepatan) penelitian ini menggunakan Persamaan (2). Percobaan dilakukan dengan menggunakan 506 data *testing* atau uji dan 1000 data *training* yang memberikan hasil akurasi sebesar 96%. Untuk hasil akurasi yang baik dalam klasifikasi sistem ini dengan melakukan perhitungan rules *Naive Bayes* yang tepat dan jumlah data *training* yang lebih banyak dibandingkan dengan jumlah data *testing* pengujian.

V. KESIMPULAN

Berdasarkan hasil pengujian yang diimplementasikan pada skenario yang dibuat, sistem menghasilkan 506 data dengan waktu 800 detik atau 13 menit 20 detik dari lokasi pengujian. Data diambil dengan menggunakan sensor node yang dikirim melalui jaringan internet ke server yang sudah tersedia dan diproses untuk klasifikasi tingkat pencemaran air sungai Citarum dengan hasil akurasi sebesar 96%. Sistem mampu memberikan informasi kondisi tingkat pencemaran air sungai Citarum dengan akurat. Hal ini menunjukkan bahwa sistem prediksi kondisi pencemaran air sungai Citarum berbasis IoT dapat dilakukan dengan klasifikasi *Naive Bayes*.

PUSTAKA

- [1] Riyadh Arridha. Design and implementation of iot-big data analytic for smart environment monitoring system. 2018
- [2] Saeful Bachrein. Pengembangan daerah aliran sungai (das) cikapundung: Diagnostik wilayah. *Jurnal Bina Praja: Journal of Home Affairs Governance*, 4(4):227–236, 2015
- [3] Pablo Bermejo, José A Gámez, and José M Puerta. Speeding up in cremental wrapper feature subset selection with naïve bayes classifier. *Knowledge-Based Systems*, 55:140–147, 2014
- [4] Nana Terangna Bukit and Iskandar A Yusuf. Beban pencemaran limbah industri dan status kualitas air sungai citarum. *J. Teknologi Lingkungan*, 3(2):98–106, 2002
- [5] Hefni Effendi. *Telaah kualitas air, bagi pengelolaan sumber daya dan lingkungan perairan*. Kanisius, 2003.
- [6] Dewan Md Farid and Mohammad Zahidur Rahman. Anomaly network intrusion detection based on improved self adaptive bayesian algorithm. *JCP*, 5(1):23–31, 2010.
- [7] S Geetha and S Gouthami. Internet of things enabled real time water quality monitoring system. *Smart Water*, 2(1):1, 2016.

- [8] Hartatik Hartatik, Andri Syafrianto, and Wiwi Widayani. Perbandingan klasifikasi pencemaran air sungai dengan metode backpropagation dan naïve bayes. *Data Manajemen dan Teknologi Informasi (DASI)*, 18(4):67–71, 2018.
- [10] Donald L Johnson, Stanley H Ambrose, Thomas J Bassett, Merle L Bowen, Donald E Crummey, Jeffry S Isaacson, Daniel N Johnson, Peter Lamb, Mahir Saul, and Alex E Winter-Nelson. Meanings of environmental terms. *Journal of environmental quality*, 26(3):581–589, 1997.
- [11] Ali Masduqi, Eddy S Soedjono, Noor Endah, and Wahyono Hadi. Prediction of rural water supply systems sustainability using a mathematical model. *Jurnal Purifikasi*, 10(2):155–164, 2009.
- [12] Burhan Alfironi Mukhtar, Noor Akhmad Setiawan, and Teguh Bharata Adji. Analisis perbandingan tingkat akurasi algoritma naïve bayes classifier dengan correlated-naïve bayes classifier. *SEMNASTEKNOMEDIA ONLINE*, 3(1):2–1, 2015.
- [13] Alieja Muhammad Putrada, Maman Abdurrohman, and Aji Gautama Putrada. Increasing smoke classifier accuracy using naïve bayes method on internet of things. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(1):19–26, 2018.
- [14] Utomo Sarjono Putro. Drama theory sebagai model dari dinamika konflik dalam permasalahan das citarum. *Journal of Technology Management*, 4(2), 2005.
- [15] Mohamad Sakizadeh. Assessment the performance of classification methods in water quality studies, a case study in karaj river. *Environmental monitoring and assessment*, 187(9):573, 2015.
- [16] Alfa Saleh. Implementasi metode klasifikasi naïve bayes dalam memprediksi besarnya penggunaan listrik rumah tangga. *Creative Information Technology Journal*, 2(3):207–217, 2015.
- [17] Hilmi Salim. Beban pencemaran limbah domestik dan pertanian di das citarum hulu. *Jurnal Teknologi Lingkungan*, 3(2):107–111, 2002.
- [18] Budi Santosa. Data mining teknik pemanfaatan data untuk keperluan bisnis. *Yogyakarta: Graha Ilmu*, 978(979):756, 2007.
- [19] Ade Sudradjat. Peran industri dan produk tekstil pada kelestarian sumberdaya lingkungan perairan das citarum. *Jurnal Teknologi Lingkungan*, 3(2):92–97, 2002.
- [20] Sritrusta Sukaridhoto, Dadet Pramadihanto, Muhammad Alif, Andrie Yuwono, Nobuo Funabiki, et al. A design of radio- controlled submarine modification for river water quality monitoring. In *2015 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pages 75–80. IEEE, 2015.
- [21] Sritrusta Sukaridhoto, Rahardhita Widyatra Sudiby, Widi Sarinastiti, Rizky Dharmawan, Atit Sasono, Ahmad Andika Saputra, and Shiori Sasaki. Design and development of a portable low-cost cots-based water quality monitoring system. In *2016 International Seminar on Intelligent Technology and Its Applications (ISITIA)*, pages 635–640. IEEE, 2016.
- [22] Pilipus A Urbasa, Suzanne L Undap, and Robert J Rompas. Dampak kualitas air pada budi daya ikan dengan jaring tancap di desa toulimembet danau tondano. *e-Journal BUDIDAYA PERAIRAN*, 3(1), 2015.
- [23] Savitri Wanabuliandari and Sekar Dwi Ardianti. Pengaruh modul e-jas edutainment terhadap karakter peduli lingkungan dan tanggung jawab. *Scholaria: Jurnal Pendidikan Dan Kebudayaan*, 8(1):70–79, 2018.
- [24] Shumei Wang, Zhaoji Zhang, Zhilong Ye, Xiaojun Wang, Xiangyu Lin, and Shaohua Chen. Application of environmental internet of things on water quality management of urban scenic river. *International Journal of Sustainable Development & World Ecology*, 20(3):216–222, 2013.
- [25] Xindong Wu and Vipin Kumar. The top ten algorithms in data mining. CRC press, 2009.
- [26] Daniela Xhemali, Chris J Hinde, and Roger G Stone. Naïve bayes vs. decision trees vs. neural networks in the classification of training web pages. 2009.
- [27] Chongsheng Zhang, Changchang Liu, Xiangliang Zhang, and George Almpandis. An up-to-date comparison of state-of-the-art classification algorithms. *Expert Systems with Applications*, 82:128–150, 2017.