



# Klasifikasi Kualitas Air Bersih Menggunakan Metode *Naïve baiyes*

Sutisna<sup>1\*</sup>, Mirsandi Nazar Yuniar<sup>2</sup>,

<sup>1,2,3</sup>Sistem Informasi, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika, Jakarta, Indonesia

Email Korespondensi : <sup>2</sup>[Mirsandiinaz09@gmail.com](mailto:Mirsandiinaz09@gmail.com),

**Abstrak**— Air merupakan kebutuhan penting bagi manusia, namun tidak semua air layak untuk dikonsumsi, sehingga penting untuk menentukan kualitas air yang digunakan untuk konsumsi. Kualitas air yang baik sangat penting, pencemaran air dapat menjadi resiko yang berbahaya. Uji laboratorium saat ini harus dilakukan untuk mengetahui kualitas air. Pengujian ini memerlukan analisis yang kompleks, sehingga penentuan kualitas air membutuhkan waktu yang lebih lama. Klasifikasi merupakan salah satu teknik data mining yang dapat digunakan untuk membangun model dari sampel data dalam kelompok yang sama. Oleh karena itu, peneliti melakukan penelitian untuk memprediksi kualitas air dengan melihat beberapa parameter yang terdiri dari data air. Tujuan dari penelitian ini untuk mengklasifikasi kualitas air bersih dan mengetahui ciri dari kualitas air yang baik. Metode penelitian ada 2 yaitu data sekunder diambil dari hasil wawancara dan observasi sedangkan primer diambil dari studi pustaka dan textbook. Hasil penelitian menggunakan RapidMiner menunjukkan akurasi sebesar 97,35%

**Kata Kunci** : Teluk naif, Air, Klasifikasi, Penambangan Data, RapidMiner

**Abstract**— Water is an essential need for humans, but not all water is suitable for consumption, so it is important to determine the quality of water used for consumption. Good water quality is very important, water pollution can be a dangerous risk. Laboratory tests must now be conducted to determine water quality. These tests require complex analysis, so determining water quality takes a longer time. Classification is one of the data mining techniques that can be used to build models from data samples in the same group. Therefore, researchers conducted a study to predict water quality by looking at several parameters that consist of water data. The purpose of this study is to classify the quality of clean water and find out the characteristics of good water quality. 2 research methods, namely secondary data taken from interviews and observations while primary data taken from literature studies and textbooks. The results of research using RapidMiner show an accuracy of 97.35%.

**Keywords**: Naïve Bayes, Water, Classification, Data Mining, RapidMiner

## I. PENDAHULUAN

Air sangat penting dalam menunjang keberlangsungan hidup manusia dan makhluk hidup lainnya yang ada di bumi. Air merupakan kebutuhan yang sangat penting bagi kehidupan manusia, namun tidak semua air aman untuk dikonsumsi, sehingga diperlukan adanya identifikasi kualitas air yang baik untuk di konsumsi[1]. Air sangat dibutuhkan oleh manusia dalam kebutuhan hidup sehari hari seperti memasak, mencuci, di konsumsi sebagai air minum dan lain lainnya. Dengan perannya yang sangat penting, air akan mempengaruhi dan dipengaruhi oleh kondisi/komponen lainnya.[2]

Seiring berkembangnya zaman, banyaknya pembangunan yang semakin pesat dapat menimbulkan kekhawatiran akan menurunnya kualitas air sehingga dapat menimbulkan dampak buruk bagi manusia, makhluk hidup lainnya dan akan menurunkan daya guna, hasil guna, produktivitas, daya dukung dan daya tampung dari sumber daya air yang pada akhirnya akan menurunkan sumber kekayaan alam.

Air merupakan salah satu media yang paling mudah dalam menularkan suatu penyakit dengan jarak jangkauan yang jauh. Kualitas air yang buruk diketahui menjadi salah satu faktor utama penyebab penyakit. Penurunannya kualitas air tidak pernah di permasalahan oleh masyarakat sehingga tidak tahu kadar kualitas air bagi kesehatan.[3] Beberapa faktor membuat menurunnya kualitas air disebabkan oleh limbah industri, limbah rumah tangga, pertambangan, dan sebagainya.

Permasalahan utama yang terjadi pada sumber daya air seperti banjir, kekeringan, erosi dan sedimentasi, intrusi air laut, dan pencemaran air. Standar air bersih sebagai acuan kualitas air yang aman dan dapat di konsumsi oleh masyarakat. Standar ini dibuat untuk memastikan bahwa air yang di konsumsi masyarakat memenuhi persyaratan kualitas yang ditetapkan oleh pemerintah atau badan pengatur lainnya

Saat ini untuk mengetahui kualitas air diperlukannya uji laboratorium. Uji laboratorium ini memiliki analisis yang cukup rumit yang mengakibatkan diperlukannya waktu yang cukup lama untuk menentukan kualitas dan kelayakan air.[4] Standar air bersih bertujuan untuk memastikan bahwa air aman untuk di konsumsi dan tidak membahayakan kesehatan, untuk melindungi lingkungan dan ekosistem dari pencemaran air. Dibutuhkan pemantauan lingkungan sekitar sumber air diperlukan agar dapat menghasilkan kualitas air yang bersih sesuai standar kualitas air bersih dan layak dikonsumsi oleh manusia.[5]

Untuk klasifikasi kualitas air bersih menggunakan teknik data mining. Data mining merupakan proses teknik statistika, matematika, kecerdasan buatan, dan machine learning untuk mengidentifikasi informasi yang bermanfaat dan pengetahuan terkait big data[6]. Data mining di bagi menjadi beberapa kelompok yaitu estimasi, prediksi, klasifikasi, clustering dan asosiasi. Data mining merupakan data yang disajikan dengan cara yang paling mudah dipahami sehingga dapat menjadi acuan untuk pengambilan keputusan[7]. Teknik yang digunakan



dalam penulisan ini menggunakan teknik klasifikasi. Klasifikasi merupakan proses menemukan pola bertujuan untuk memperkirakan kelas dari objek yang belum diketahui[8]. Klasifikasi dapat dideskripsikan sebagai metode untuk membuktikan sebuah objek data sebagai salah satu jenis yang telah dideskripsikan sebelumnya[9]. Cara pengelompokan klasifikasi berdasarkan ciri-ciri yang dimiliki suatu objek dan dapat dilakukan oleh manusia ataupun dengan bantuan teknologi[10]. Teknik klasifikasi ini menggunakan algoritma naive bayes.

Algoritma naïve bayes merupakan pengklasifikasian statistik yang digunakan untuk memprediksi probabilitas keanggotaan suatu class[11]. Algoritma teorema Bayes mengasumsikan semua atribut independen atau tidak saling ketergantungan yang diberikan oleh nilai pada variabel kelas[12]. Naive Bayes sering bekerja jauh lebih baik dalam kebanyakan situasi dunia nyata yang kompleks dari pada yang diharapkan.[13]. Pada metode ini, semua atribut akan memberikan kontribusinya dalam pengambilan keputusan, dengan bobot atribut yang sama penting dan setiap atribut saling bebas satu sama lain[14]. Tools yang digunakan dalam penelitian ini menggunakan tools Rapidminer. RapidMiner adalah sebuah lingkungan machine learning data mining, text mining dan predictive analytics[15].

## II. METODE PENELITIAN

Metode penelitian juga dilakukan untuk menemukan jawaban dari permasalahan yang terjadi dalam sistem, penyebab permasalahan tersebut dan cara mengatasi permasalahan yang ada. Dalam penyusunan laporan, menggunakan beberapa metode untuk memperoleh informasi dalam mengklasifikasi kualitas air bersih menggunakan metode naïve bayes

### a. Data Penelitian

#### 1) Data primer

Data primer adalah data yang diperoleh atau dikumpulkan langsung di lapangan oleh peneliti dari orang yang bersangkutan. Data primer terdiri dari :

- a) Wawancara adalah percakapan yang terjadi antara narasumber dengan pewawancara
- b) Observasi adalah suatu aktivitas pengamatan yang dilakukan secara langsung

#### 2) Data Sekunder

Data skunder adalah berbagai informasi yang dikumpulkan oleh peneliti untuk kepentingan penelitian. Dara skunder terdiri dari :

- a) Studi Pustaka adalah pengumpulan data pustaka, membaca dan mencatat serta mengolah bahan penelitian.
- b) Textbook adalah pengumpulan informasi yang dilakukan oleh peneliti dengan membaca buku-buku yang berkaitan dengan penelitian.

### b. Atribut

Data penelitian ini merupakan data private dimana data tersebut merupakan data perusahaan bukan data public yang dapat digunakan secara bebas. Data mentah yang diperoleh berjumlah 226 record dengan 8 atribut

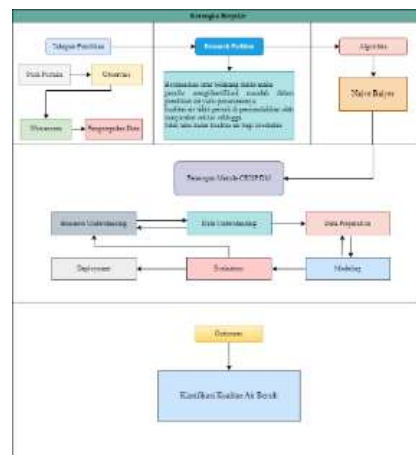
Tabel 1 Deskripsi Nama Atribut Pada Data Awal

Atribut	Keterangan
Tanggal	Tanggal Pengecekan Air
Rasa	Tidak ada rasa dan berasa
Bau	Tidak ada bau dan bau
Kekeruhan	Keruh air
Ph	Ph kadar air
Suhu	Suhu air
Sisa Khlor	Sisa Khlor Air
Analyzer	Analyzer Air

Setelah data terkumpul diperlukan penyiapan data dari data mentah menjadi data siap olah untuk proses selanjutnya yang akan dilakukan pendeskripsian data, evaluasi pemilihan data dan pemilihan atribut.

### c. Penerapan Metodologi

Pada penelitian ini metode yang digunakan adalah algoritma naïve bayes. Data yang digunakan pada penelitian ini data kualitas air. Pada tahapan penelitian ini akan dilakukan pengumpulan data awal maka proses selanjutnya akan dilakukan Pendeskripsian Data, kemudian setelah itu dilakukan Evaluasi Pemilihan Data, lalu pada tahap selanjutnya dilakukan Pemilihan Atribut, tahap berikutnya melakukan penerapan metode CRISP-DM yang pada proses penelitian ini mengacu pada enam tahap CRISP-DM yaitu pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi dan penyebaran. Berikut tahapan metodologi penelitian yang dapat dilihat pada Gambar



Gambar 1 Tahapan Penerapan Metodologi

### d. Rancangan Pengujian



Gambar 2. CRISP DM



Metode yang digunakan pada tahapan pendekatan ini menggunakan metode Cross Industry Standard for Data Mining (CRISP-DM). CRISP-DM merupakan metode yang menggunakan model proses pengembangan data yang banyak digunakan para ahli untuk memecahkan masalah. Proses penelitian ini mengacu pada enam tahap terdapat dalam CRISP-DM ini yakni dijelaskan sebagai berikut :

a. Bussines Understanding (Pemahaman Bisnis)

Dalam tahap ini tujuan penelitian adalah untuk menentukan strategi dalam meningkatkan kualitas air jernih. Untuk dilakukan pengelompokan data yang di dapatkan hasil observasi supaya dapat mengetahui kualitas air jernih dan tidak jernih.

b. Data Understanding (Pemahaman Data)

Berdasarkan data yang telah didapat dari transaksi penjualan, tahap selanjutnya dalam metode CRISP-DM yaitu pemahaman terhadap kebutuhan data terkait dengan pencapaian tujuan untuk menentukan strategi dalam meningkatkan kualitas air jernih. Data kualitas air tersebut memiliki jumlah 226 dengan 8 atribut atribut yang akan dilakukan dengan cara pemilihan atribut menjadi 5 atribut yaitu kekeruhan, ph, suhu, sisa klor, analyzer.

c. Data Preparation (Persiapan Data)

Semua data yang mencakup semua kegiatan pada persiapan data yang diteliti untuk membangun dataset kualitas air yang diterapkan ke dalam alat pemodelan, dari data mentah berupa dataset kualitas air dan selanjutnya akan dilakukan proses data mining. Tahap selanjutnya dilakukan tahap pemodelan klasifikasi pada data mining. Pada proses ini akan dilakukan persiapan data kualitas air yang telah dilakukan pada tahap evaluasi dan dilakukan perbaikan data berdasarkan hasil evaluasi tersebut kemudian dilakukan penetapan tujuan untuk meningkatkan strategi kualitas air setiap tahunnya berdasarkan atribut pada tahap persiapan data merupakan implementasi awal dari penerapan tujuan tersebut.



Gambar 3 Pemodelan menggunakan Rapidminer

d. Modelling

Pada tahap ini akan digunakan pemodelan menggunakan metode klasifikasi dengan algoritma naïve bayes. Dalam penerapan metode klasifikasi akan dibagi menjadi 2 kategori yaitu kualitas air jernih dan tidak jernih. Rapidminer adalah alat pemodelan yang digunakan pada penelitian ini.

e. Evaluation

Fase evaluasi merupakan tahap untuk mengetahui apakah model yang di rancang telah sesuai atau belum dengan tujuan pada fase awal. Tujuan awal di

dirancangnya model ini yaitu agar menghasilkan nilai akurasi yang tinggi, sehingga dapat membuktikan bahwa penelitian yang dilakukan telah berhasil.

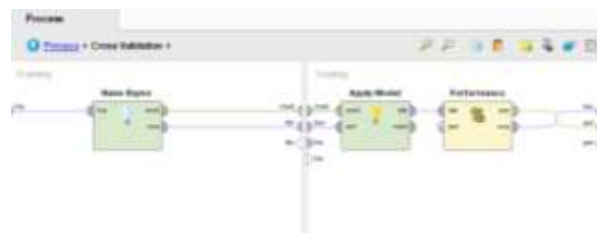
f. Deployment

Tahap selanjutnya dari penelitian yang sudah dilakukan yaitu tahap penyebaran hasil dan dijadikan sebagai presentasi atau laporan dari pengetahuan yang sudah didapat berdasarkan evaluasi dan pemodelan pada proses data mining

### III. HASIL DAN PEMBAHASAN

#### a. Implementasi Pengujian Menggunakan Rapidminer

Berikut ini merupakan proses pengolahan dataset menggunakan RapidMiner dengan metode Naive Bayes.



Gambar 4 Split Validation

Tujuan utama dilakukannya penelitian ini adalah untuk mengetahui nilai akurasi dari pemodelan menggunakan algoritma Naive Bayes dengan tools yang digunakan yaitu RapidMiner. Pola data baru yang dihasilkan digunakan untuk proses klasifikasi kualitas air bersih. Dalam kolom training terdapat algoritma yang digunakan yaitu Naive Bayes. Sedangkan dalam kolom testing, terdapat operator Apply Model yang berfungsi untuk menjalankan model Naive Bayes serta Performance yang berfungsi untuk mengukur performa dari model Naive Bayes yang digunakan tersebut.

#### b. Hasil Pengujian

Pada percobaan pengolahan dataset dengan algoritma Naive Bayes dibantu tools RapidMiner, diperoleh waktu eksekusi proses yaitu 0 detik. Hal ini membuktikan bahwa perhitungan menggunakan Naive Bayes berjalan sangat cepat.



Gambar 5. Simple Distribution

Model distribusi untuk label atribut Class adalah sebagai berikut :

- a) Class “Jernih” : 5 distributions
- b) Class “Tidak Jernih” : 5 distributions

Percobaan pada penelitian ini menggunakan Rapidminer 9.10 algoritma yang digunakan adalah Naive Bayes. Validasinya menggunakan cross validation dan



untuk testing menggunakan Apply Model untuk menjalankan algoritma atau model Naive Bayes serta Performance untuk mengukur performa dari model Naive Bayes tersebut.

Hasil akurasi perhitungan menggunakan algoritma Naive Bayes menunjukkan persentase nilai sebesar 97,35%. Artinya nilai akurasi menggunakan algoritma ini sudah baik berdasarkan tingkat akurasinya yang mencapai 90%

#### IV. KESIMPULAN

Berdasarkan hasil penelitian dapat disimpulkan bahwa Metode bayes dapat menentukan kualitas air jernih dan tidak jernih. Data yang digunakan untuk penelitian ini sebanyak 226 record dengan atribut turbidity,ph,temperature,sisa khlor,total khlor yang akan menjadi database pada Ms.Excel,data tersebut kemudian dikoneksikan pada tools RapidMiner,dan diolah dalam bentuk Naive Bayes. Analisa data menggunakan algoritma Naive Bayes dalam menyusun strategi kualitas agar tetap stabil. Klasifikasi yang digunakan terdapat dua kelas, yaitu air jernih dan tidak jernih. Hasil akhir pada pengujian menggunakan RapidMiner menghasilkan akurasi sebesar 97,35%

#### V. REFERENSI

[1] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN , Naive Bayes Dan Decision Tree," vol. 9, no. 2, pp. 1109–1119, 2022.

[2] Y. S. Sari, "Penerapan Metode Naive Bayes Untuk Mengetahui Kualitas Air Di Jakarta," vol. XIII, no. 2, pp. 222–228, 2021.

[3] U. Optimasi and S. Pemasaran, "Jurnal Teknologi Informasi dan Komunikasi STMIK Subng, April 2018 ISSN: 2252-4517," no. April, pp. 84–95, 2018.

[4] R. Forest, "url : <http://studentjournal.umpo.ac.id/index.php/komputer>," 2022.

[5] M. A. Rahman, N. Hidayat, and A. A. Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naive Bayes Untuk Klasifikasi Kualitas Air Bersih ( Studi Kasus PDAM Tirta Kencana Kabupaten Jombang )," vol. 2, no. 12, 2018.

[6] I. Ramadhan, "Data Mining untuk Klasifikasi Penderita Kanker Payudara Berdasarkan Data dari University Medical Center Menggunakan Algoritma Naive Bayes  $P(H|X) = \frac{P(X|H) \chi P(H)}{P(H)}$ ," vol. 7, no. 1, pp. 21–27, 2020, doi: 10.30865/jurikom.v7i1.1755.

[7] E. T. Lestari and J. Adhiva, "Implementation Naive Bayes Classifier Algorithm and K-Nearest Neighbor For Obesity Nutritional Status of Children with Disabilities Implementasi Algoritma Naive Bayes Classifier dan K-Nearest Neighbor Untuk Klasifikasi Status Gizi Obesitas Anak Disabilitas," pp. 1–11, 2022.

[8] M. Yoshe and W. Hadikurniawati, "Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita," vol. 2019, 2021.

[9] F. Alghifari and D. Juardi, "Penerapan Data Mining

Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naive Bayes," 2021.

[10] M. M. Mutoffar, M. Naseer, and A. Fadillah, "KLASIFIKASI KUALITAS AIR SUMUR MENGGUNAKAN," vol. 04, no. 02, pp. 138–146, 2022.

[11] H. Annur, "KLASIFIKASI MASYARAKAT MISKIN MENGGUNAKAN METODE," vol. 10, pp. 160–165, 2018.

[12] G. P. Kawani, "Implementasi Naive Bayes," J. Informatics, Inf. Syst. Softw. Eng. Appl., vol. 1, no. 2, pp. 73–81, 2019, doi: 10.20895/inista.v1i2.73.

[13] "1jurnal," vol. 10, pp. 191–197, 2019.

[14] S. Kusumadewi, "KLASIFIKASI STATUS GIZI MENGGUNAKAN," vol. 3, no. 1, pp. 6–11, 2009.

[15] D. Nuswantoro, "DATA MINING MENGGUNAKAN ALGORITMA NAIVE BAYES UNTUK KLASIFIKASI KELULUSAN MAHASISWA UNIVERSITAS," pp. 1–11, 2009.