

Penerapan Algoritma Naïve Bayes Dalam Pengelompokkan Air Layak Minum

Dian Permata Sari¹, Wira Buana², Rozi Meri³, Rika Widya Perdana⁴

^{1,2}Program Studi Sistem Informasi, STMIK Jayanusa Padang, Sumatera Barat, Indonesia

³Program Studi Teknologi Informasi, Politeknik Negeri Padang, Sumatera Barat, Indonesia

⁴Program Studi Teknik Komputer, AMIK Kosgoro, Sumatera Barat, Indonesia

E-mail: ¹Langitbiru621@gmail.com, ² wira_buana59@yahoo.com,

³rozimeri88@gmail.com, ⁴rikaperdana9@gmail.com

Abstract

Water is a basic human need, without water humans cannot carry out activities. However, not all types of water can be consumed and are healthy for humans. So it is important to check water quality regularly. Because the influence of weather can make the water quality unfit consumption. To determine water quality, laboratory testing is required. This test requires complex analysis results so the time used in testing is relatively longer. In data mining there are classification techniques that can be used to group drinking water samples based on several variables that have been determined from water data. The aim of this research is to classify the types of drinking water that are suitable for consumption and the characteristics of healthy water. To process the data using the Naïve Bayes algorithm and produce an accuracy value of 75%.

Keywords: Drinking Water, Naïve Bayes, Data Mining

Abstrak

Air merupakan kebutuhan pokok manusia, tanpa air manusia tidak dapat beraktivitas. Namun tidak semua jenis air dapat dikonsumsi dan sehat untuk manusia. Sehingga penting dilakukan pemeriksaan kualitas air secara berkala. Karena pengaruh cuaca dapat membuat kualitas air menjadi tidak layak untuk dikonsumsi. Untuk mengetahui kualitas air diperlukan pengujian di laboratorium. Pengujian ini memerlukan hasil analisis yang kompleks sehingga waktu yang digunakan dalam pengujian relatif lebih lama. Dalam data mining terdapat teknik klasifikasi yang dapat digunakan untuk mengelompokkan sampel air minum berdasarkan beberapa variabel yang telah ditentukan dari data air. Tujuan dari penelitian ini adalah untuk mengelompokkan jenis air minum yang layak dikonsumsi serta ciri-ciri dari air yang sehat. Untuk mengolah data menggunakan algoritma naïve bayes dan menghasilkan nilai akurasi sebesar 75%.

Kunci: Air Minum, Naïve Bayes, Data Mining

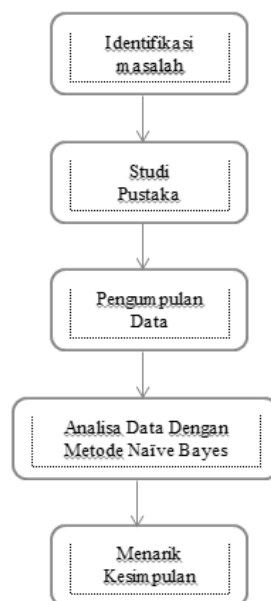
1. Pendahuluan

Air merupakan suatu elemen yang terdapat di bumi dan keberadaannya tidak dapat dipisahkan dari manusia dan makhluk hidup lainnya. Sehingga air sangat dibutuhkan untuk menjaga kelangsungan makhluk hidup. Kegunaan air dalam tubuh manusia untuk menambah cairan, menghilangkan haus ketika dahaga dan untuk menunjang kehidupan manusia air memiliki banyak manfaat seperti pembuatan bangunan, jalan, teknologi dan sebagainya. Dalam fungsi air untuk memenuhi cairan tubuh haruslah dapat dipastikan air yang dikonsumsi itu sehat dan bersih [1]. air bersih merupakan air yang digunakan untuk kebutuhan sehari-hari seperti memasak, mencuci, minum dan sebagainya [2].

Air yang layak atau tidak dikonsumsi biasanya dipengaruhi oleh faktor zat yang terkandung didalam air dan harus memenuhi standar kalitas air bersih, untuk itu diperlukan pengawasan dan pengelolaan lingkungan terhadap kualitas air [3]. Terdapat banyak metode yang digunakan dalam pengelompokkan kualitas air bersih diantaranya adalah algoritma naive bayes. Pada penelitian sebelumnya oleh Muhammad Malik Mutoffar dkk menggunakan algoritma Random Forest dalam melakukan klasifikasi kualitas air sumur dengan tingkat akurasi 82% dari 83% data yang dapat diklasifikasikan. [4]. Penelitian lain yang dilakukan oleh Hakam Febtadianrano Putro dkk tentang alogritma naïve bayes untuk klasifikasi pelanggan, terdapat tiga variable yang digunakan yaitu jumlah pembelian, interval waktu dan lokasi dengan nilai akurasi mencapai 92%, nilai presisi mencapai 100% dan nilai recall mencapai 91%.[5]. Aldi Tangkelayuk dalam penelitiannya klasifikasi kualitas air menggunakan Metode KNN, Naïve Bayes Dan Decision Tree,hasil penelitian diperoleh dari membandingkan nilai akurasi dari ketiga algoritma. KNN memiliki tingkat akurasi tertinggi yaitu 86,88% kemudian Decision Tree sebesar 80,44% dan akurasi terendah sebesar 63,60% metode naïve bayes [6]. Pada penelitian Penerapan Algoritma Naïve Bayes untuk memprediksi tingkat kelulusan peserta Sertifikasi Microsoft Office Specialist (MOS) yang dilakukan oleh Mochamad Farid Rifai dkk menghasilkan nilai probabilitas sebesar 0.001042 dengan tingkat akurasi sebesar 99% [7]. Dari penjabaran di atas dapat disimpulkan bahwa algortima naïve bayes merupakan suatu metode yang memiliki akurasi cukup tinggi. Untuk itu penulis akan menerapkan naïve bayes dalam klasifikasi air layak minum.

2. Metodologi Penelitian

Dibawah ini adalah tahapan yang penulis gunakan dalam penelitian mulai dari identifikasi masalah sampai memperoleh hasil akhirnya.



Gambar 1. Kaidah Penelitian

Pada tahap identifikasi masalah menentukan varibel yang digunakan dalam penelitian yaitu TDS, Ph, Besi, Arsen, KadmiumTimbal dan Alumunium. Pada tahap studi pustaka mengumpulkan buku serta jurnal-jurnal referensi yang dapat dijadikan rujukan dalam penelitian ini. Kemudian mengolah data dengan metode naïve bayes dengan melakukan perhitungan secara manual menggunakan ms excel. Tahap terakhir dalam penelitian ini adalah menarik kesimpulan dari hasil penelitian yang telah dilakukan.

Naïve Bayes merupakan pengklasifikasian dengan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes [8][9]. Naïve Bayes adalah salah satu metode pengklasifikasian sederhana yang sering digunakan karena mudah diterapkan dan memiliki hasil akurasi yang cukup tinggi dibandingkan dengan metode klasifikasi yang lain [10][11]. Cara kerja metode Naive Bayes Classifier menggunakan perhitungan probabilitas [12], yaitu mencari nilai probabilitas paling tinggi untuk mengklasifikasikan sebuah data uji dengan kategori yang tepat. Teknik prediksi probabilitas yang sederhana didasarkan pada penerapan teorema bayes atau aturan bayes merupakan suatu teknik yang diimplementasikan pada algoritma naïve bayes. Rumus naïve bayes [13]:

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

Keterangan:

X adalah data sampel dengan kelas (label) yang tidak diketahui

H adalah hipotesa bahwa X adalah data dengan kelas (label).

P(H) adalah peluang dari hipotesa H.

P(X) adalah peluang data sampel yang akan diamati.

P(X|H) adalah peluang data sampel X, bila diasumsikan bahwa hipotesa benar (valid).

3. Hasil dan Pembahasan

Penelitian ini menerapkan algoritma naïve bayes dalam pengelompokkan air layak minum berdasarkan kualitas air yang dinilai dari beberapa variable yaitu TDS, PH, Besi, Arsen, Kadmium dan Alumunium. Terdapat enam varibael yang digunakan dalam menilai kualitas air layak minum. Adapun data yang digunakan dalam penelitian ini adalah data hasil dari pemeriksaan air minum yang diambil langsung dari rumah masyarakat kemudian dilakukan pemeriksaan di laboratorium.

Tabel 1. Data Training

NO	Nama	TDS	Ph	Besi	Arsen	KadmiumTimbal	Alumunium	Keterangan
1	A	210	6,9	0,00	0,057	0,007	0,00	TIDAK MEMENUHI SYARAT
2	B	248	6,9	0,016	0,00	0,00	0,006	MEMENUHI SYARAT
3	C	146	6,6	0,003	0,00	0,00	0,009	MEMENUHI SYARAT
4	A	133	6,5	0,004	0,00	0,00	0,007	MEMENUHI SYARAT
5	B	110	6,9	0,006	0,00	0,00	0,004	MEMENUHI SYARAT
6	C	192	7,1	0,00	0,34	0,00	0,002	TIDAK MEMENUHI SYARAT
7	A	169	6,9	0,008	0,00	0,00	0,00	MEMENUHI SYARAT
8	B	159	6,9	0,00	0,00	0,00	0,005	MEMENUHI SYARAT
9	C	169	7,0	0,020	0,00	0,179	0,018	TIDAK MEMENUHI SYARAT
10	A	168	6,8	0,00	0,00	0,108	0,00	TIDAK MEMENUHI SYARAT
11	B	110	6,4	0,001	0,00	0,007	0,003	TIDAK MEMENUHI SYARAT
12	C	113	7,0	0,004	0,076	0,082	0,00	TIDAK MEMENUHI SYARAT
13	A	121	6,9	0,0	0,00	0,00	0,00	MEMENUHI SYARAT
14	B	158	6,7	0,00	0,096	0,004	0,004	TIDAK MEMENUHI SYARAT
15	C	168	6,3	0,0	0,01	0,058	0,007	TIDAK MEMENUHI SYARAT

Langkah perhitungan naïve bayes:

1. Menentukan nilai untuk setiap kelas

C1 (Kelas=" Memenuhi Syarat")= jumlah "Memenuhi syarat" pada kolom keterangan= 7

C2 (Kelas="Tidak Memenuhi Syarat")=jumlah "Tidak Memenuhi Syarat"pda kolom keterangan=8

Tabel 2. Nilai Probabilitas kelas

menghitung jumlah kelas			
label	jumlah data	jumlah seluruh data	hasil
memenuhi syarat	7	15	0.46667
tidak memenuhi syarat	8	15	0.5

2. Menentukan nilai pada setiap parameter

a. Nilai TDS

Jika kondisi baik, range nilai 0-250

Jika kondisi cukup, range nilai 250-300

Jika kondisi buruk, range nilai >300

Tabel 3. Probabilitas Nilai TDS

tds	memenuhi syarat	tidak memenuhi syarat	P(memenuhi syarat)	P(tidak memenuhi syarat)
baik	7	7	1	0.875
cukup	0	1	0	0.125

Pada Tabel 3 merupakan hasil perhitungan probabilitas untuk nilai TDS. Dapat dilihat bahwa TDS baik dengan kondisi memenuhi syarat memiliki nilai probabilitas sebesar 1 sedangkan TDS baik dengan kondisi tidak memenuhi syarat memiliki nilai 0,875. TDS cukup dengan kondisi memenuhi syarat nilai probabilitasnya 0 sedangkan TDS cukup dengan kondisi tidak memenuhi syarat memiliki dengan probabilitas sebesar 0,125.

b. Nilai PH

Jika kondisi netral, range nilai 6,5-8

Jika kondisi asam, range nilai <6

Jika kondisi basa, range nilai >8

Tabel 4. Probabilitas nilai PH

ph	memenuhi syarat	tidak memenuhi syarat	P(memenuhi syarat)	P(tidak memenuhi syarat)
netral	7	6	1	0.75
asam	0	2	0	0.25

Pada Tabel merupakan hasil perhitungan probabilitas untuk nilai PH. Dapat dilihat bahwa PH netral dengan kondisi memenuhi syarat memiliki nilai probabilitas sebesar 1 sedangkan PH netral dengan kondisi tidak memenuhi syarat memiliki nilai 0,75. TDS asam dengan kondisi memenuhi syarat nilai probabilitasnya 0 sedangkan PH asam dengan kondisi tidak memenuhi syarat memiliki dengan probabilitas sebesar 0,25.

c. Nilai Besi

Jika kondisi baik, range nilai 0,0-0,2

Jika kondisi buruk, range nilai >0,2

Tabel 5. Probabilitas Nilai Besi

besi	memenuhi syarat	tidak memenuhi syarat	P(memenuhi syarat)	P(tidak memenuhi syarat)
baik	7	8	1	1
buruk	0	0	0	0

Pada Tabel 5 merupakan hasil perhitungan probabilitas untuk nilai besi. Dapat dilihat bahwa Besi baik dengan kondisi memenuhi syarat memiliki nilai probabilitas sebesar 1 sedangkan Besi baik dengan kondisi tidak memenuhi syarat memiliki nilai 1. Besi Buruk dengan kondisi memenuhi syarat nilai probabilitasnya 0 sedangkan Besi buruk dengan kondisi tidak memenuhi syarat memiliki dengan probabilitas sebesar 0.

d. Nilai Arsen

Jika kondisi baik, range nilai 0,0-0,01

Jika kondisi buruk, range nilai >0,01

Tabel 6. Probabilitas Nilai Arsen

arsen	memenuhi syarat	tidak memenuhi syarat	P(memenuhi syarat)	P(tidak memenuhi syarat)
baik	7	4	1	0.5
buruk	0	4	0	0.5

Pada Tabel 6 merupakan hasil perhitungan probabilitas untuk nilai Arsen. Dapat dilihat bahwa Arsen baik dengan kondisi memenuhi syarat memiliki nilai probabilitas sebesar 1 sedangkan Arsen baik dengan kondisi tidak memenuhi syarat memiliki nilai probabilitas sebesar 0,5. Arsen buruk dengan kondisi memenuhi Syarat maka nilai probabilitasnya 0 sedangkan Arsen buruk dengan kondisi tidak memenuhi syarat maka nilai probabilitasnya adalah 0,5.

e. Nilai Kadmium Timbal

Jika kondisi baik, range nilai 0,0-0,003

Jika kondisi buruk, range nilai >0,003

Tabel 7. Probabilitas Nilai Kadmium Timbal

kadmium timbal	memenuhi syarat	tidak memenuhi syarat	P(memenuhi syarat)	P(tidak memenuhi syarat)
baik	7	1	1	0.125
buruk	0	7	0	0.875

Pada Tabel 7 merupakan hasil perhitungan probabilitas untuk nilai Kadmium timbal. Dapat dilihat bahwa kadmium timbal baik dengan kondisi memenuhi syarat memiliki nilai probabilitas sebesar 1 sedangkan kadmium timbal baik dengan kondisi tidak memenuhi syarat memiliki nilai probabilitas sebesar 0,125. Kadmium timbal buruk dengan kondisi memenuhi Syarat maka nilai probabilitasnya 0 sedangkan Kadmium timbal buruk dengan kondisi tidak memenuhi syarat maka nilai probabilitasnya adalah 0,875.

f. Nilai Alumunium

Jika kondisi baik, range nilai 0,0-0,2

Jika kondisi buruk, range nilai >0,2

Tabel 8. Probabilitas Nilai Alumunium

alumunium	memenuhi syarat	tidak memenuhi syarat	P(memenuhi syarat)	P(tidak memenuhi syarat)
baik	7	8	1	1
buruk	0	0	0	0

Pada Tabel 8 merupakan hasil perhitungan probabilitas untuk nilai Alumunium. Dapat dilihat bahwa Alumunium baik dengan kondisi memenuhi syarat memiliki nilai probabilitas sebesar 1 sedangkan Alumunium baik dengan kondisi tidak memenuhi syarat memiliki nilai probabilitas sebesar 1. Alumunium buruk dengan kondisi memenuhi Syarat maka nilai probabilitasnya 0 sedangkan Alumuniuml buruk dengan kondisi tidak memenuhi syarat maka nilai probabilitasnya adalah 0.

3. Pengukuran tingkat akurasi

Setelah menentukan nilai probabilitas pada tiap varibael maka tahap selanjutnya menentuka tingkat akurasi. Akurasi diperoleh dari jumlah data yang diprediksi benar dibagi dengan jumlah data yang diprediksi.

$$Akurasi = \frac{\sum i}{\sum n} * 100\%$$

(2)

Untuk data testing yang dilakukan untuk mencari akurasi digunakan sebanyak 5 data sampel dari 15 data training. Adapun prediksi nilai nya seperti pada tabel di bawah ini:

Tabel 9. Hasil Data Uji

data uji	
kelas sebenarnya	kelas prediksi
tidak memenuhi syarat	tidak memenuhi syarat
memenuhi syarat	tidak memenuhi syarat
tidak memenuhi syarat	tidak memenuhi syarat
tidak memenuhi syarat	memenuhi syarat
memenuhi syarat	memenuhi syarat

Dari hasil data uji maka dengan menggunakan rumus akurasi untuk mencari akurasi dari data uji.

Tabel 10. Hasil Perhitungan Akurasi, Presisi dan Recall

CONCLUSION MATRIX		
	PREDIKSI	
AKTUAL	memenuhi syarat	tidak memenuhi syarat
memenuhi syarat	1	1
tidak memenuhi syarat	1	2
akurasi	75%	
presisi	50%	
recall	50%	

Dari perhitungan berdasarkan tabel maka di peroleh nilai akurasi sebesar 75%, sedangkan nilai presisi sebesar 50%. Presisi menunjukkan seberapa baik model dapat mengklasifikasikan dengan benar untuk kelas yang di prediksi positif (memenuhi syarat). Sementara recall: seberapa baik model dapat mengklasifikasikan dengan benar untuk kelas positif (memenuhi syarat) yaitu sebesar 50%.

4. Kesimpulan

Dari hasil perhitungan terhadap kasus klasifikasi air layak minum menggunakan algoritma naïve bayes maka diperoleh tingkat akurasi sebesar 75% dengan perhitungan manual menggunakan ms excel.

Daftar Pustaka

- [1] N. R. Saragi, A. Sembiring, And Nurhayati, "Sistem Pakar Mendiagnosa Kelayakan Air Minum Untuk Dikonsumsi Menggunakan Metode Certainty Factor Pada Pdam Tirta Sari Kota Binjai," *J. Sains Teknol.*, Vol. 2, No. 1, Pp. 23–26, 2022.
- [2] Y. S. Sari, "Penerapan Metode Naïve Bayes Untuk Mengetahui Kualitas Air Di Jakarta," *J. Ilm. Fiso*, Vol. 13, No. 2, P. 222, 2021, Doi: 10.22441/Fifo.2021.V13i2.010.
- [3] M. Djana, "Analisis Kualitas Air Dalam Pemenuhan Kebutuhan Air Bersih Di Kecamatan Natar Hajimena Lampung Selatan," *J. Redoks*, Vol. 8, No. 1, Pp. 81–87, 2023, Doi: 10.31851/Redoks.V8i1.11853.
- [4] M. M. Mutoffar And A. Fadillah, "Klasifikasi Kualitas Air Sumur Menggunakan Algoritma Random Forest," *Naratif J. Nas. Riset, Apl. Dan Tek. Inform.*, Vol. 4, No. 2, Pp. 138–146, 2022, Doi: 10.53580/Naratif.V4i2.160.
- [5] H. F. Putro, R. T. Vlandari, And W. L. Y. Saptomo, "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan," *J. Teknol. Inf. Dan Komun.*, Vol. 8, No. 2, 2020, Doi: 10.30646/Tikomsin.V8i2.500.
- [6] A. Tangkelayuk, "The Klasifikasi Kualitas Air Menggunakan Metode Knn, Naïve Bayes, Dan Decision Tree," *Jatisi (Jurnal Tek. Inform. Dan Sist. Informasi)*, Vol.

- 9, No. 2, Pp. 1109–1119, 2022, Doi: 10.35957/Jatisi.V9i2.2048.
- [7] M. F. Rifai, H. Jatnika, And B. Valentino, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (Mos),” *Petir*, Vol. 12, No. 2, Pp. 131–144, 2019, Doi: 10.33322/Petir.V12i2.471.
 - [8] Syarif And A. A. Muin, “Metode Naive Bayes Untuk Prediksi Kelulusan,” *J. Ilm. Ilmu Komput.*, Vol. 2, No. 1, Pp. 22–26, 2016, [Online]. Available: <https://media.neliti.com/media/publications/283828-metode-naive-bayes-untuk-prediksi-kelulu-139fcfea.pdf>
 - [9] R. Rachman And R. N. Handayani, “Klasifikasi Algoritma Naive Bayes Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras Umkm,” *J. Inform.*, Vol. 8, No. 2, Pp. 111–122, 2021, Doi: 10.31294/Ji.V8i2.10494.
 - [10] N. Chatrina Siregar, R. Ruli, A. Siregar, ; M. Yoga, And D. Sudirman, “Implementasi Metode Naive Bayes Classifier (Nbc) Pada Komentar Warga Sekolah Mengenai Pelaksanaan Pembelajaran Jarak Jauh (Pjj),” *J. Teknol. Aliansi Perguru. Tinggi Bumn*, Vol. 3, No. 1, Pp. 102–110, 2020.
 - [11] A. Gusderia, M. Ramadhan, And M. I. Perangin-Angin, “Data Mining Untuk Klasifikasi Data Penjualan Alat Teknik Menggunakan Metode Naive Bayesian Clacifier,” *J. Saintikom (Jurnal Sains Manaj. Inform. Dan Komputer)*, Vol. 21, No. 2, P. 73, 2022, Doi: 10.53513/Jis.V21i2.5973.
 - [12] M. Muqorobin, K. Kusri, And E. T. Luthfi, “Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah,” *J. Ilm. Sinus*, Vol. 17, No. 1, P. 1, 2019, Doi: 10.30646/Sinus.V17i1.378.
 - [13] A. B. P. Et Al. Swi Indriyaningtyasa1, “Implementasi Data Mining Menggunakan Algoritma Naïve Bayes Classifier Untuk Mendeteksi Minat Beli Pelanggan Terhadap Produk Paket Internet,” *J. Ilmu Komput.*, Vol. 15, No. 2, Pp. 1–7, 2021.