

IMPLEMENTASI METODE *NAIVE BAYES* UNTUK KLASIFIKASI KUALITAS AIR BERSIH DI SUNGAI KAHAYAN

Lutfiah Nur Hasinah¹, Agatha Deolika², Danish Rio Wahyudi³, Nisa Afrilliska⁴

1) Teknik Lingkungan, Fakultas Teknik, Institut Teknologi dan Sains Nahdlatul Ulama Kalimantan, Indonesia

2) Teknik Komputer, Fakultas Teknik, Institut Teknologi dan Sains Nahdlatul Ulama Kalimantan, Indonesia

3)4) Teknik Lingkungan, Fakultas Teknik, Institut Teknologi dan Sains Nahdlatul Ulama Kalimantan, Indonesia

| Article Info | ABSTRACT |
|--|---|
| <p>Article history:</p> <p>Received: 24 Oktober 2025 Revised: 12 Desember 2025 Accepted: 23 Desember 2025</p> | <p>Abstrak</p> <p>Penelitian ini mengusulkan sebuah model klasifikasi untuk menentukan kualitas air bersih berdasarkan parameter pH, <i>Dissolved Oxygen</i> (DO), dan <i>Total Dissolved Solids</i> (TDS) menggunakan metode <i>Naive Bayes</i>. Penelitian dilakukan pada dataset air Sungai Kahayan, Kalimantan Tengah, yang penting sebagai sumber air baku bagi masyarakat sekitar. Dengan menganalisis data historis parameter kualitas air, model <i>Naive Bayes</i> diharapkan dapat mengklasifikasikan air ke dalam kategori 'bersih' atau 'tidak bersih' dengan akurasi tinggi. Hasil penelitian ini menghasilkan bahwa model <i>naive bayes</i> sudah mampu melakukan klasifikasi dengan optimal berdasarkan hasil evaluasi yang dilakukan menggunakan <i>Confusion Matrix</i>. Pengujian dilakukan dengan bermacam jumlah dataset (100, 250, dan 500), dan menghasilkan data yang tidak konsisten. Data paling optimal Adalah dataset 100 data dengan Akurasi 100%, Recall 100%, dan Presisi 100%. Walaupun hasilnya kurang konsiten tetapi model sudah mampu mengklasifikasi dengan baik dan benar. Penelitian ini diharapkan dapat memberikan kontribusi dalam pemantauan kualitas air secara efektif dan mendukung pengambilan keputusan untuk pengelolaan sumber daya air.</p> |

Kata Kunci: Kualitas air, *Naive Bayes*, Sungai Kahayan, klasifikasi, Sungai.

Abstract

This study proposes a classification model to determine clean air quality based on pH, Dissolved Oxygen (DO), and Total Dissolved Solids (TDS) parameters using the Naive Bayes method. The study was conducted on the Kahayan River water dataset, Central Kalimantan, which is important as a source of raw water for the surrounding community. By analyzing historical data of air quality parameters, the Naive Bayes model is expected to be able to classify air into the 'clean' or 'unclean' category with high accuracy. The results of this study indicate that the Naive Bayes model is able to perform optimal classification based on the evaluation results carried out using a Confusion Matrix. Testing was carried out with various numbers of datasets (100, 250, and 500), and produced inconsistent data. The most optimal data is the 100 data dataset with 100% Accuracy, 100% Recall, and 100% Precision. Although the results are less consistent, the model is able to classify well and correctly. This research is expected to contribute to effective air quality monitoring and support decision making for water resource management.

Keywords: Water quality, *Naive Bayes*, Kahayan River, classification, River.

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



Corresponding Author:

E-mail : nurhasinahlutfiah@gmail.com

1. PENDAHULUAN

Air sangat dibutuhkan oleh manusia dalam kebutuhan hidup sehari-hari seperti memasak, mencuci, dikonsumsi sebagai air minum dan lain-lainnya. Dengan perannya yang sangat penting, air akan mempengaruhi dan dipengaruhi oleh kondisi/komponen lainnya.[1] Air bersih merupakan kebutuhan esensial[2] yang ketersediaannya terus terancam oleh aktivitas antropogenik dan perubahan lingkungan, sebagaimana yang juga terjadi pada Sungai Kahayan di Kalimantan Tengah yang vital sebagai sumber air baku namun rentan terhadap penurunan kualitas. Dibutuhkan pemantauan lingkungan sekitar sumber air diperlukan agar dapat menghasilkan kualitas air yang bersih sesuai standar kualitas air bersih dan layak dikonsumsi oleh manusia.[3] Mengingat pentingnya peran Sungai Kahayan bagi masyarakat, pemantauan dan klasifikasi kualitas air yang akurat adalah krusial untuk menjaga kesehatan publik dan kelestarian lingkungan. 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 klor, total klor.[4] Klasifikasi merupakan proses menemukan pola bertujuan untuk memperkirakan kelas dari objek yang belum diketahui.[5] Tingkat akurasi data setelah dilakukan evaluasi dengan menggunakan *Confusion Matrix* adalah sebesar 84,2411% yang masuk dalam kategori *Good Classification*. [6]

Proses pengklasifikasian berita dalam penelitian ini meliputi: pengumpulan data, text preprocessing, pembobotan kata, dan klasifikasi naïve bayes classifier. Nilai akurasi tertinggi yang diperoleh dalam penelitian ini sebesar 94% dengan pembagian data uji 10% dan data latih 90%.[7] Klasifikasi dengan naïve bayes melakukan perhitungan nilai probabilitas hasil dari penelitian ini memiliki akurasi sebesar 89.83%.[8] Penelitian ini berupaya menjawab permasalahan mengenai karakteristik kualitas air Sungai Kahayan berdasarkan parameter pH, *Dissolved Oxygen* (DO), dan *Total Dissolved Solids* (TDS), serta bagaimana membangun dan mengevaluasi model klasifikasi kualitas air (bersih atau tidak bersih) menggunakan metode *Naive Bayes* berdasarkan parameter-parameter tersebut, dengan tujuan akhir untuk mengidentifikasi dan menganalisis karakteristik kualitas air, mengembangkan model klasifikasi yang efektif, dan mengevaluasi kinerjanya dalam mendukung upaya pengelolaan dan pemantauan kualitas air Sungai Kahayan.

2. METODE PENELITIAN

2.1 Pengumpulan Data

Data yang akan digunakan dalam penelitian ini adalah data historis kualitas air Sungai Kahayan, yang meliputi parameter pH, DO, dan TDS. Data ini dapat diperoleh dari instansi terkait Balai Wilayah Sungai II Provinsi Kalimantan Tengah yang melakukan pemantauan rutin. Data akan dikumpulkan selama periode tertentu, misalnya 2-3 tahun terakhir, dengan frekuensi pengambilan sampel bulanan atau triwulanan. Setiap sampel data akan memiliki nilai pH, DO (dalam mg/L), dan TDS (dalam mg/L), serta label kelas kualitas air yang telah ditentukan berdasarkan standar baku mutu air bersih yang berlaku (PP No. 22 Tahun 2021).

2.2 Pra-pemrosesan Data

Data yang telah terkumpul akan melalui tahap pra-pemrosesan untuk memastikan kualitas dan konsistensi data. Tahap ini meliputi penanganan nilai yang hilang (missing values) melalui interpolasi atau penghapusan data jika jumlahnya sedikit. Deteksi dan penanganan *outlier* juga akan dilakukan untuk menghindari bias dalam model. Selanjutnya, data akan dinormalisasi atau distandarisasi jika diperlukan, meskipun *Naive Bayes* tidak terlalu sensitif terhadap skala fitur. Terakhir, data akan dibagi menjadi dua set: data pelatihan (*Training data*) dan data pengujian (*Testing data*), biasanya dengan rasio 80:20, untuk melatih dan mengevaluasi kinerja model secara independen.

2.3 Penerapan Algoritma *Naive Bayes*

Algoritma *Naive Bayes* akan diimplementasikan menggunakan *tools* Python dengan *library scikit-learn*. Langkah-langkah penerapan meliputi:

1. Estimasi Probabilitas Priors

Menghitung probabilitas setiap kelas (bersih, tidak bersih) dari data pelatihan.

2. Estimasi Probabilitas Bersyarat

Menghitung probabilitas setiap fitur (pH, DO, TDS) untuk setiap kelas.

3. Klasifikasi

Untuk setiap sampel data pengujian, hitung probabilitas posterior untuk setiap kelas menggunakan Teorema Bayes, dan tetapkan sampel ke kelas dengan probabilitas posterior tertinggi.

2.4 Evaluasi

Kinerja model *Naive Bayes* akan dievaluasi menggunakan metrik standar klasifikasi, yaitu confusion matrik dengan parameter:

1. Akurasi: Rasio prediksi yang benar terhadap total prediksi.
2. Presisi: Proporsi kasus positif yang benar dari total kasus positif yang diprediksi.
3. *Recall* (Sensitivitas): Proporsi kasus positif yang benar dari total kasus positif yang sebenarnya.

2.5 Kualitas Air Bersih dan Parameter Indikator

Kualitas air bersih didefinisikan oleh serangkaian parameter fisik, kimia, dan biologis yang harus memenuhi standar tertentu agar aman untuk konsumsi dan penggunaan lainnya. pH, yang mengukur tingkat keasaman atau kebasaan air, penting karena nilai ekstrem dapat bersifat korosif atau beracun. *Dissolved Oxygen* (DO) adalah oksigen terlarut dalam air yang vital bagi kehidupan akuatik; rendahnya DO seringkali mengindikasikan pencemaran organik. *Total Dissolved Solids* (TDS) mencerminkan jumlah padatan terlarut anorganik dan organik yang terkandung dalam air, yang dapat memengaruhi rasa, bau, dan bahkan kesehatan jika konsentrasinya terlalu tinggi. Organisasi seperti *World Health Organization* (WHO) dan standar nasional (misalnya, Peraturan Pemerintah di Indonesia) menetapkan ambang batas untuk parameter-parameter ini guna menentukan kelayakan air untuk berbagai keperluan.

2.6 Sungai Kahayan sebagai Lokasi Penelitian

Sungai Kahayan memiliki peran strategis bagi masyarakat di Provinsi Kalimantan Tengah, terutama di Kota Palangka Raya, sebagai sumber air minum, transportasi, perikanan, dan aktivitas ekonomi lainnya. Namun, sungai ini juga menerima beban dari aktivitas perkotaan, pertanian, dan kadang-kadang pertambangan ilegal, yang dapat memengaruhi kualitas airnya. Data historis dari dinas lingkungan hidup setempat sering menunjukkan fluktuasi dalam parameter kualitas air, yang menggarisbawahi perlunya sistem pemantauan yang canggih dan metode klasifikasi yang akurat untuk menjaga integritas ekologis dan fungsional sungai. Studi sebelumnya di wilayah serupa telah menunjukkan bahwa variabilitas musiman dan aktivitas manusia secara signifikan memengaruhi parameter pH, DO, dan TDS.

2.7 Metode Klasifikasi *Naive Bayes*

Naive Bayes adalah algoritma klasifikasi probabilistik yang didasarkan pada Teorema Bayes dengan asumsi "naive" bahwa setiap fitur independen satu sama lain. Meskipun asumsi independensi ini jarang terpenuhi sepenuhnya dalam praktiknya, *Naive Bayes* seringkali menunjukkan kinerja yang sangat baik dalam berbagai aplikasi klasifikasi, termasuk dalam bidang lingkungan. Keunggulannya meliputi efisiensi komputasi, kemampuannya menangani data dengan dimensi tinggi, dan kemampuannya untuk bekerja dengan dataset yang relatif kecil. Dalam konteks klasifikasi kualitas air, *Naive Bayes* dapat memprediksi probabilitas suatu sampel air termasuk dalam kategori 'bersih' atau 'tidak bersih' berdasarkan nilai-nilai pH, DO, dan TDS yang diamati. Penerapan *Naive Bayes* telah banyak diteliti untuk masalah klasifikasi lingkungan dan diagnostik. *Naive bayes classifier* (NBC) merupakan model yang menggunakan metode statistik dan probabilitas dalam proses klasifikasi yang dikenalkan oleh Thomas Bayes (ilmuwan Inggris), yaitu tentang prediksi probabilitas yang terjadi dimasa depan dengan mempertimbangkan pengalaman dari masa yang telah terjadi.[9] Rumus Teorema Bayes memiliki persamaan sebagai berikut: [10]

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

Keterangan :

X = Data dengan class yang belum diketahui

H = Hipotesis data X merupakan suatu class spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi x (posteriori prob.)

$P(H)$ = Probabilitas hipotesis H (prior prob.)

$P(X|H)$ = Probabilitas X berdasarkan kondisi tersebut

$P(X)$ = Probabilitas dari X

3. HASIL DAN PEMBAHASAN

Dalam penelitian ini dilaksanakan pengujian model *navie bayes* dengan 3 variable input yang menghasilkan klasifikasi data air Sungai air bersih dan kotor. Pengujian akan dilakukan dengan pengujian beberapa data yaitu 100, 250, dan 500 data. Berikut adalah dataset yang digunakan dapat dilihat pada Tabel 1:

Tabel 1. Dataset

| No | pH | DO | TDS | Hasil |
|-----|------------|------------|------------|--------|
| 1 | 8.98834457 | 7.32380979 | 11.4869846 | Bersih |
| 2 | 5.57834305 | 6.30897269 | 6.28790366 | Bersih |
| 3 | 9.34962426 | 7.44670043 | 7.88519636 | Kotor |
| 4 | 6.71253002 | 6.53850775 | 11.7820594 | Bersih |
| 5 | 4.68769326 | 7.5299105 | 13.359933 | Kotor |
| 6 | 5.32264194 | 6.28193548 | 7.41652641 | Bersih |
| 7 | 6.13487504 | 7.33912256 | 11.2226649 | Bersih |
| 8 | 4.18829804 | 7.18313851 | 15.6310398 | Kotor |
| 9 | 9.76222616 | 7.38844901 | 12.0825798 | Kotor |
| 10 | 5.63273803 | 6.16096002 | 13.9057699 | Bersih |
| ... | | | | |
| 500 | 7.47724353 | 7.79333707 | 10.584223 | Bersih |

Table 2. Komposisi Data

| No | Jenis Data | Jumlah |
|-------------------|----------------------|--------|
| Pungujain Pertama | | |
| 1 | Data <i>Training</i> | 80 |
| 2 | Data <i>Testing</i> | 20 |
| Pengujuan Kedua | | |
| 1 | Data <i>Training</i> | 200 |
| 2 | Data <i>Testing</i> | 50 |
| Pengujuan Ketiga | | |
| 1 | Data <i>Training</i> | 400 |
| 2 | Data <i>Testing</i> | 100 |

Pada Tabel 2 dapat dilihat komposisi data yang dibagi untuk penelitian ini Adalah 80:20. Dengan data 80% atau 400 data, model dilatih dengan 3 variabel input dan menghasilkan output kualitas air.

Tabel 3. Data *Testing* Pengujuan Ketiga

| No | pH | DO | TDS |
|----|------------|------------|------------|
| 1 | 4.14082127 | 6.39222969 | 15.9734494 |
| 2 | 4.78931646 | 6.33957572 | 19.2161356 |
| 3 | 8.87107748 | 7.13794002 | 7.45173157 |
| 4 | 4.72271341 | 7.06194672 | 15.6213142 |
| 5 | 7.76106952 | 7.25084874 | 6.9340319 |
| 6 | 4.17837996 | 7.87092145 | 6.77541689 |
| 7 | 8.0872126 | 7.13467391 | 17.4973554 |
| 8 | 9.22377363 | 6.29559728 | 15.0227439 |
| 9 | 6.38562839 | 6.13846197 | 9.57438501 |

| | | | |
|-----|------------|------------|------------|
| 10 | 6.40721928 | 7.08512564 | 7.05164603 |
| ... | | | |
| 100 | 8.03495442 | 7.17797276 | 17.9326246 |

Setelah proses Training data menggunakan metode *Naive Bayes* selesai, selanjutnya pengujian data uji yang berjumlah 25, 50, dan 100 data pelanggan dapat dilihat pada Tabel 4, Tabel 5, dan Tabel 6.

Tabel 4. Data Klasifikasi Pertama (100)

| No | Hasil Sebenarnya | Hasil Klasifikasi |
|-----|------------------|-------------------|
| 1 | Bersih | Bersih |
| 2 | Bersih | Bersih |
| 3 | Kotor | Kotor |
| 4 | Bersih | Bersih |
| 5 | Bersih | Bersih |
| 6 | Bersih | Bersih |
| 7 | Kotor | Kotor |
| 8 | Bersih | Bersih |
| 9 | Bersih | Bersih |
| 10 | Bersih | Bersih |
| ... | | |
| 20 | Bersih | Bersih |

Tabel 5. Data Klasifikasi Kedua (250)

| No | Hasil Sebenarnya | Hasil Klasifikasi |
|-----|------------------|-------------------|
| 1 | Bersih | Bersih |
| 2 | Kotor | Kotor |
| 3 | Bersih | Bersih |
| 4 | Bersih | Kotor |
| 5 | Bersih | Bersih |
| 6 | Bersih | Bersih |
| 7 | Kotor | Kotor |
| 8 | Bersih | Bersih |
| 9 | Bersih | Bersih |
| 10 | Bersih | Bersih |
| ... | | |
| 50 | Bersih | Bersih |

Berdasarkan hasil (20 data) pada table 4, hasil klasifikasi menunjukan hasil 20 data benar di klasifikasikan dan 0 data yang salah. Berdasarkan hasil (50 data) pada table 5, hasil klasifikasi menunjukan hasil 46 data benar di klasifikasikan dan 4 data yang salah.

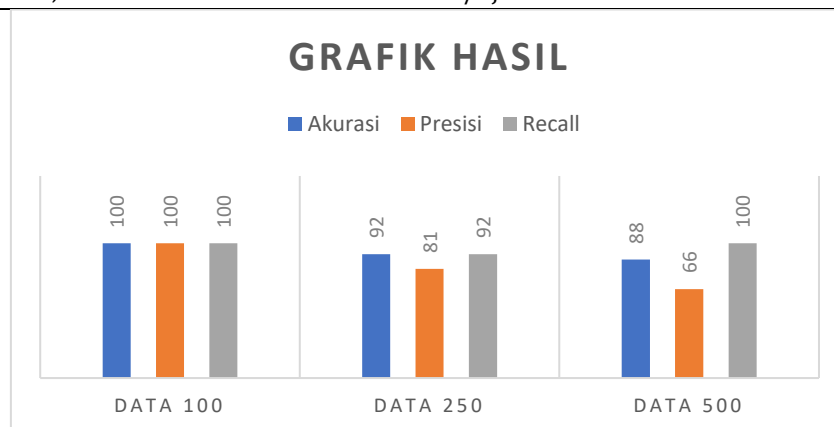
Tabel 6. Data Klasifikasi Ketiga (500)

| No | Hasil Sebenarnya | Hasil Klasifikasi |
|-----|------------------|-------------------|
| 1 | Kotor | Kotor |
| 2 | Kotor | Kotor |
| 3 | Bersih | Bersih |
| 4 | Kotor | Bersih |
| 5 | Bersih | Bersih |
| 6 | Kotor | Bersih |
| 7 | Bersih | Kotor |
| 8 | Kotor | Bersih |
| 9 | Bersih | Kotor |
| 10 | Bersih | Bersih |
| ... | | |
| 100 | Bersih | Kotor |

Berdasarkan hasil (100 data) pada table 6, hasil klasifikasi menunjukan hasil 88 data benar di klasifikasikan dan 12 data salah.

Tabel 7. Data Evaluasi (*Confusion Matrix*)

| No | Parameter | Hasil |
|-------------------------|---------------|-------|
| Pengujian Pertama (100) | | |
| 1 | Akurasi | 100% |
| 2 | Presisi | 100% |
| 3 | Recall | 100% |
| Pengujian Kedua (250) | | |
| 1 | Akurasi | 92% |
| 2 | Presisi | 81% |
| 3 | Recall | 92% |
| Pengujian Ketiga (500) | | |
| 1 | Akurasi | 88% |
| 2 | Presisi | 66% |
| 3 | <i>Recall</i> | 100% |

Gambar 1. Grafik *Confusion Matrix*

Hasil Evaluasi menggunakan *Confusion Matrix* pada tabel 7 dan Gambar 1 menunjukkan hasil klasifikasi dataset 100 data mendapatkan rata-rata 100%, sedangkan dataset 250 menunjukkan hasil akurasi 92%, recall 92%, dan presisi 81%. Dataset 500 data menghasilkan akurasi 88%, recall 66%, presisi 100%.

4. SIMPULAN

Hasil penelitian ini menghasilkan bahwa model naïve bayes sudah mampu melakukan klasifikasi dengan optimal berdasarkan hasil evaluasi yang dilakukan menggunakan *Confusion Matrix*. Pengujian dilakukan dengan bermacam jumlah dataset (100, 250, dan 500) menghasilkan data yang tidak konsisten, terjadi penurunan akurasi, recall, dan presisi ketika dataset di tingkatkan. Data paling optimal Adalah dataset 100 data dengan Akurasi 100%, Recall 100%, dan Presisi 100%. Walaupun hasilnya kurang konsisten tetapi model sudah mampu mengklasifikasi dengan baik dan benar. Penelitian ini diharapkan dapat memberikan kontribusi dalam pemantauan kualitas air secara efektif dan mendukung pengambilan keputusan untuk pengelolaan sumber daya air.

UCAPAN TERIMA KASIH

Kami Ucapkan terima kasih kepada Kementerian Pendidikan Tinggi, Sains, dan Teknologi telah memberikan hibah penelitian kepada saya dan juga saya berterima kasih kepada Institut Teknologi dan Sains Nahdlatul Ulama Kalimantan, tanpa mereka peneliti ini tidak bisa terlaksana dengan baik. Semoga penelitian ini dapat bermanfaat bagi masyarakat untuk kemajuan pengetahuan Indonesia

REFERENCES

- [1] Y. S. Sari, "Penerapan Metode Naïve Bayes Untuk Mengetahui Kualitas Air Di Jakarta," *Jurnal Ilmiah FIFO*, vol. 13, no. 2, p. 222, Nov. 2021, doi: 10.22441/fifo.2021.v13i2.010.
- [2] A. Tangkelayuk and E. Mailoa, "Klasifikasi Kualitas Air Menggunakan Metode KNN, Naïve Bayes Dan Decision Tree," vol. 9, no. 2, pp. 1109–1119, 2022, [Online]. Available: <http://jurnal.mdp.ac.id>
- [3] M. A. Rahman, N. Hidayat, and A. A. Supianto, "Komparasi Metode Data Mining K-Nearest Neighbor Dengan Naïve Bayes Untuk Klasifikasi Kualitas Air Bersih (Studi Kasus PDAM Tirta Kencana Kabupaten Jombang)," 2018. [Online]. Available: <http://j-ptiik.ub.ac.id>
- [4] M. Nazar Yuniar, "Klasifikasi Kualitas Air Bersih Menggunakan Metode Naïve bayies," *Jurnal Sains dan Teknologi*, vol. 5, no. 1, pp. 243–246, 2023, doi: 10.55338/saintek.v5i1.1383.
- [5] M. Y. Titimeidara and W. Hadikurniawati, "Monica Yoshe Titimeidara Implementasi Metode Naive Bayes Implementasi Metode Naive Bayes Classifier Untuk Klasifikasi Status Gizi Stunting Pada Balita."
- [6] N. Alfiah, "Klasifikasi Penerima Bantuan Sosial Program Keluarga Harapan Menggunakan Metode Naive Bayes".
- [7] S. Muhammad Habib, E. Haerani, S. Kurnia Gusti, S. Ramadhani, and T. H. Informatika UIN Sultan Syarif Kasim Riau Jl Soebrantas, "Klasifikasi Berita Menggunakan Metode Naïve Bayes Classifier," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 2, 2022.
- [8] H. Susana and N. Suarna, "PENERAPAN MODEL KLASIFIKASI METODE NAIVE BAYES TERHADAP PENGGUNAAN AKSES INTERNET Program Studi Teknik Informatika STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon 3) Program Studi Rekayasa Perangkat Lunak STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon 4) Program Studi Komputerisasi Akuntansi STMIK IKMI Cirebon Jl Perjuangan No 10B Kesambi Kota Cirebon," *Jurnal Sistem Informasi dan Teknologi Informasi*, vol. 4, no. 1, pp. 1–8, 2022.
- [9] I. Rasila, U. Ristian, J. Rekayasa Sistem Komputer, and F. H. MIPA Universitas Tanjungpura Jl Hadari Nawawi, "IMPLEMENTASI METODE NAIVE BAYES CLASSIFIER PADA SISTEM PENGKLASIFIKASI BERITA OTOMATIS BERBASIS WEBSITE (STUDI KASUS: BERITA LOKAL DARI MEDIAMASSA ONLINE KALIMANTAN BARAT)," 2019.
- [10] H. Al Rasyid Harpizon *et al.*, "Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naïve Bayes," *Jurnal Nasional Komputasi dan Teknologi Informasi*, vol. 5, no. 1, 2022.
- [11] R. Watrionthos dkk., "Analisis Algoritma Naive Bayes Untuk Klasifikasi Kualitas Air Sungai," *Jurnal Informatika dan Teknologi Komputer*, 2020.
- [12] A. P. Windarto, "Penerapan Data Mining Pada Data Ekspor Menggunakan K-Means Clustering," *Jurnal Teknologi Informasi dan Ilmu Komputer*, 2018.
- [13] F. S. Jumeilah, "Penerapan Algoritma Naive Bayes Untuk Klasifikasi Kualitas Air," *Jurnal Teknologi dan Sistem Informasi*, 2017.
- [14] M. W. Kasrani, "Water Quality Classification Using Machine Learning Algorithms," *Int. Journal of Adv. Science and Tech*, 2020.
- [15] S. N. Rizki, "Analisis Perbandingan Algoritma C4.5 dan Naive Bayes Dalam Klasifikasi Kualitas Air," *Jurnal Sistem Informasi*, 2021.
- [16] I. Gunawan dkk., "Implementation of Naive Bayes Algorithm for Water Quality Prediction," *Journal of Physics*, 2020.
- [17] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*, 2nd ed. Wiley, 2014.
- [18] T. M. Mitchell, *Machine Learning*, McGraw-Hill Education, 1997.
- [19] B. Santosa, *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*, Graha Ilmu, 2007.
- [20] A. Saleh, "Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Penggunaan Listrik," *Jurnal INTI*, 2018.