

Penerapan Algoritma Decision Tree Untuk Klasifikasi Dan Prediksi Potabilitas Air Minum

Application Of A Decision Tree Algorithm For Classification And Prediction Of Drinking Water Potability

Bobby Khoir Nanda¹, Lili Tanti²

^{1,2} Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama
K.L. Yos Sudarso KM 6,5 No. 3A Tj.Mulia – Medan

Email : bobbykhoir818@gmail.com¹, Email : lilitanti82@gmail.com²

ABSTRAK

Ketersediaan air minum yang layak konsumsi (potabel) merupakan hal yang vital bagi kesehatan masyarakat. Namun, data menunjukkan bahwa masih banyaknya sumber air yang terkontaminasi, sehingga menimbulkan kebutuhan akan sistem deteksi yang akurat dan efisien. Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree dalam membangun model klasifikasi untuk memprediksi potabilitas air minum berdasarkan serangkaian parameter fisika dan kimia. Dataset yang digunakan bersumber dari Kaggle, terdiri dari 3.276 sampel data dengan 9 fitur atribut, yaitu: pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes, dan Turbidity, dengan variabel target Potability (0 = Tidak Layak, 1 = Layak). Metodologi penelitian meliputi Exploratory Data Analysis (EDA) untuk memahami karakteristik dari data, penanganan data hilang dengan teknik imputasi rata-rata berdasarkan kelompok potabilitas, standardisasi data menggunakan StandardScaler, dan pembagian dataset dengan rasio 80% untuk data latih dan 20% untuk data uji. Hasil evaluasi model menunjukkan bahwa algoritma Decision Tree mampu mengklasifikasikan potabilitas air dengan akurasi sebesar 74.7%. Analisis lebih lanjut menggunakan confusion matrix dan classification report mengungkapkan bahwa model lebih baik dalam memprediksi kelas "Tidak Layak" (Precision: 0.79, Recall: 0.81) dibandingkan kelas "Layak" (Precision: 0.66, Recall: 0.65), yang diidentifikasi sebagai dampak dari ketidakseimbangan jumlah data pada kedua kelas. Kesimpulan dari penelitian ini adalah algoritma Decision Tree dapat diimplementasikan sebagai alat yang efektif untuk memprediksi potabilitas air dengan kinerja yang baik dan model yang mudah diinterpretasikan, meskipun perbaikan lebih lanjut diperlukan untuk menangani data yang tidak seimbang.

Kata Kunci : **Klasifikasi, Decision Tree, Machine Learning, Potabilitas Air, Kualitas Air.**

ABSTRACT

The availability of potable drinking water is vital for public health. However, data indicates that many water sources remain contaminated, creating a need for accurate and efficient detection systems. This research aims to apply the Decision Tree algorithm to build a classification model for predicting drinking water potability based on a series of physical and chemical parameters. The dataset used, sourced from Kaggle, consists of 3,276 data samples with 9 feature attributes: pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes, and Turbidity, with the target variable Potability (0 = Not Potable, 1 = Potable). The research methodology includes Exploratory Data Analysis (EDA) to understand data characteristics, handling missing data using mean imputation based on potability groups, data standardization using StandardScaler, and splitting the dataset with an 80% ratio for training data and 20% for testing data. Model evaluation results show that the Decision Tree algorithm can classify water potability with an accuracy of 74.7%. Further analysis using a confusion matrix and classification report reveals that the model performs better at predicting the "Not Potable" class (Precision: 0.79, Recall: 0.81) compared to the "Potable" class (Precision: 0.66, Recall: 0.65), which is identified as an impact of data imbalance between the two classes. The conclusion of this study is that the Decision Tree algorithm can be implemented as an effective tool for predicting water potability with good performance and an easily interpretable model, although further improvements are needed to address the imbalanced data.

Kata Kunci : **Classification, Decision Tree, Machine Learning, Water Potability, Water Quality.**

Disubmit:

Info Artikel :
Direview:

Diterima :

Copyright © 2025 – senadimu. All rights reserved.

DOI: <https://www.doi.org/10.22303>

1. PENDAHULUAN

Air merupakan hasil dari alam yang menopang seluruh aspek kehidupan. Kualitas air minum yang aman dan layak konsumsi (potabel) secara langsung berdampak memengaruhi kesehatan masyarakat dan menjadi salah satu indikator inti dalam pembangunan suatu negara. Meskipun akses terhadap air minum layak di Indonesia menunjukkan peningkatan, kualitas air yang sampai ke masyarakat masih menjadi tantangan besar yang nyata. Data dari Badan Pusat Statistik[1] menunjukkan bahwa persentase rumah tangga dengan akses air minum layak telah mencapai 91,72%. Namun, indikator ini tidak sepenuhnya mencerminkan kualitas air yang dikonsumsi. Laporan dari [2] mengungkap fakta yang sangat mengkhawatirkan: sekitar 80% sumber air minum di Indonesia masih terkontaminasi bakteri *Escherichia coli* (*E. coli*), yang merupakan indikator utama pencemaran *fekal*(limbah / tinja).

Masalah ini semakin diperparah dengan menjamurnya tempat pengisian air minum isi ulang yang standar sanitasi dan pengawasannya seringkali tidak memenuhi standar. Sebuah studi oleh [3] menemukan bahwa 45,4% sampel dari air isi ulang terkontaminasi *E. coli*, menunjukkan angka yang lebih tinggi dibandingkan air dari Perusahaan Daerah Air Minum (PDAM) sebesar 33%. Kondisi ini dapat menciptakan urgensi untuk mengembangkan sebuah sistem pemantauan dan prediksi kualitas air yang tidak hanya sangat akurat tetapi juga efisien dari segi waktu dan biaya, mengingat melakukan pengujian laboratorium konvensional seringkali memakan waktu yang lama dan mahal.

Menurut [4], *Artificial Intelligence* atau kecerdasan buatan merupakan sebuah sistem komputer yang mampu melakukan berbagai tugas-tugas yang biasanya membutuhkan kemampuan kecerdasan manusia. Teknologi ini dapat membuat dan menimbang keputusan dengan cara menganalisis pola dan menggunakan data yang ada dan tersedia di dalam sistem yang sudah di kumpulkan.

Menurut [5], *Machine Learning* dimungkinkan untuk melakukan otomatisasi pada proses pemantauan, meningkatkan efisiensi dan mengurangi ketergantungan pada penilaian manusia. Teknologi ini mampu memahami pola tersembunyi dalam dan memberikan penilaian profesional berdasarkan aturan yang telah dipelajari, sehingga menghasilkan pengambilan keputusan yang lebih cepat dan tepat sesuai dengan yang di pelajari.

Di tengah permasalahan saat ini, kemajuan teknologi mengembangkan kecerdasan buatan (*Artificial Intelligence*) dan pembelajaran mesin (*Machine Learning*) menawarkan solusi yang sangat menjanjikan. Algoritma *machine learning* mampu menganalisis sekumpulan dataset besar yang berisi parameter kualitas air untuk menemukan sebuah pola kompleks dan membuat prediksi yang sangat akurat. Salah satu algoritma untuk melakukan klasifikasi yang populer dari segi efektivitasnya dan interpretasinya yang mudah adalah *Decision Tree* (Pohon Keputusan). Cara kerja *algoritma* ini dengan membangun model prediktif yang berbentuk struktur pohon, di mana setiap simpul internal mewakili "tes" pada sebuah atribut, setiap cabang mewakili hasil tes, dan setiap simpul daun mewakili label kelas. Keunggulan utamanya adalah kemampuannya untuk menghasilkan aturan-aturan secara logis dan mudah dipahami oleh manusia. Sejumlah penelitian sebelumnya telah menunjukkan potensi algoritma *Decision Tree* dan variasinya dalam klasifikasi kualitas air.

Penelitian yang dilakukan [6] dengan judul “Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning” dengan *exploratory data analysis* dapat mempermudah menganalisis data. Hasil yang diperoleh dari serangkaian penelitian ini dapat disimpulkan bahwa algoritma *Random Forest Classifier* memiliki tingkatan akurasi yang paling baik dibandingkan beberapa *algoritma* lainnya yang diuji. Sistem identifikasi kualitas air minum memiliki akurasi sebesar 76.52% dengan menggunakan *algoritma Random Forest Classifier*.

Penelitian yang dilakukan [7] dengan judul “Membandingkan *Algoritma Decision Tree* Dengan Indeks Kualitas Air Kimiawi (CWQI) Untuk Mengklasifikasikan Kualitas Air Tanah Di Wilayah Rawan Kekeringan Di Odisha, India.” Menekukan bahwa fokus utama adalah pada kandungan *ion fluoride*. Hasilnya menunjukkan bahwa metode *Decision Tree* sangat efektif dalam mengidentifikasi zona kualitas air yang tinggi untuk manajemen pasokan air. Meskipun CWQI memberikan gambaran secara keseluruhan yang baik, algoritma *Decision Tree* memberikan hasil yang lebih akurat dan lebih mudah dimengerti dalam analisis kualitas air untuk wilayah yang rawan kekeringan.

Penelitian yang dilakukan [8] dengan judul “*Prediction of Cyanobacteria Using Decision Tree Algorithm and Sensor Monitoring Data*”. penelitian ini menunjukkan bahwa *Decision Tree* dapat diimplementasikan untuk memprediksi kejadian *cyanobacteria* di sungai, dengan menggunakan data dari sensor kualitas air. Model ini memberikan kemampuan detail prediksi yang cukup baik untuk mendeteksi potensi risiko kualitas air yang berpotensi berbahaya. *Algoritma Decision Tree* membantu dalam mendeteksi dini dan memberikan peringatan terhadap masalah kualitas air yang berpotensi membahayakan kesehatan, seperti munculnya *cyanobacteria*.

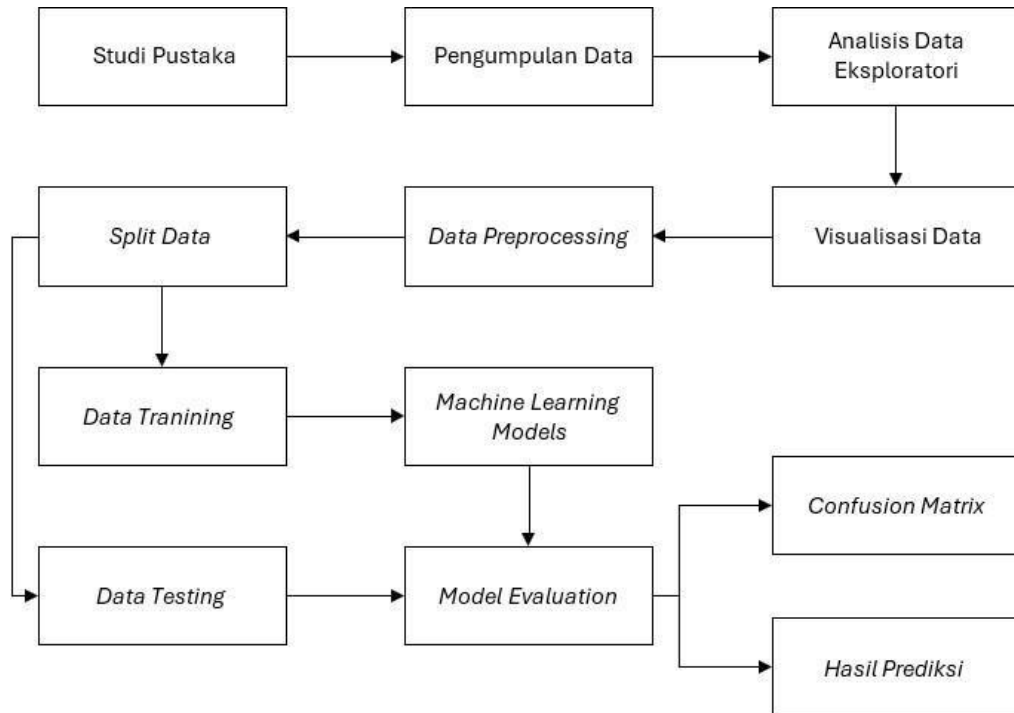
Penelitian yang dilakukan [9] dengan judul “*Machine Learning-Based Water Quality Classification Assessment*”. Penelitian ini menunjukkan model hibrida yang menggabungkan *algoritma Decision Tree* dengan algoritma lain menunjukkan akurasi sangat tinggi dalam mengklasifikasikan kualitas air. Model ini digunakan untuk mengklasifikasikan 15 macam parameter kualitas air, dan hasilnya menunjukkan hasil yang sangat baik. *Decision Tree*, bila dikombinasikan dengan teknik lain, dapat meningkatkan persentase akurasi dalam klasifikasi kualitas air dan dapat diandalkan untuk penentuan keputusan dalam pengelolaan air.

Penelitian yang dilakukan [10] dengan judul “*Machine Learning Approaches for Fish Pond Water Quality Classification: Random Forest, Gaussian Naive Bayes, and Decision Tree Comparison*”. Dalam penelitian ini, meskipun *Random Forest* dan *Gaussian Naive Bayes* menampilkan hasil yang lebih baik dari segi akurasi, *Decision Tree* tetap memberikan kinerja yang baik dan lebih mudah dipahami. *Algoritma Decision Tree* diidentifikasi sebagai alat yang bermanfaat untuk melakukan pemantauan kualitas air dengan multivariat, dengan akurasi yang dapat diterima. *Algoritma Decision Tree* efektif meskipun ada algoritma lain yang memberikan kinerja lebih baik, karena interpretabilitasnya yang menjanjikan dalam mengaplikasinya pada klasifikasi kualitas air.

Meskipun demikian, masih terdapat celah penelitian dalam menerapkan *algoritma Decision Tree* murni pada dataset potabilitas air yang umum digunakan, dengan mekalikan analisis yang mendalam terhadap fitur-fitur yang paling berpengaruh dan evaluasi kinerja yang transparan, terutama pada data yang tidak seimbang. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma *Decision Tree* untuk mengklasifikasikan dan memprediksi potabilitas air minum. Fokus penelitian adalah membangun model yang tidak hanya akurat tetapi juga dapat diimplementasikan, serta mengevaluasi kinerjanya secara komprehensif menggunakan metrik standar. Diharapkan, bahwa hasil penelitian ini dapat memberikan kontribusi pada bidang ilmiah dan menjadi dasar bagi pengembangan sistem pendukung keputusan yang praktis untuk pemantauan kualitas air yang lebih baik.

1. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif berbasis data dengan metodologi yang terstruktur untuk memastikan validitas dan reproduibilitas hasil. Tahapan penelitian dirancang secara sistematis, mulai dari pengumpulan data hingga evaluasi model, seperti yang digambarkan dalam diagram alur berikut.



Gambar 1. Diagram Alur Metodologi Penelitian.

2.1. Sumber Data dan Deskripsi Dataset

Penelitian ini menggunakan dataset sekunder berjudul "Water Potability" yang diperoleh melalui *platform online repositori* data publik, Kaggle merupakan *platform* berbasis komunitas yang sangat terkenal di kalangan *data scientist* dan pengembang *machine learning*, yang menyediakan ruang untuk berkolaborasi, berbagi, dan bertanding dalam tantangan analisis data. Kaggle menawarkan berbagai *dataset* yang dapat digunakan untuk eksperimen [11]. *Dataset* ini total berisi 3.276 sampel air, di mana setiap sampel dianalisis berdasarkan 9 parameter fisika-kimia sebagai fitur (variabel independen) dan 1 variabel target (variabel dependen).

	A	B	C	D	E	F	G	H	I	J
	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
1		2.048 904.554 713 360	20.791 318 980 747.000	7.300 211 873 184 750	36.851 644 134 980 300	5.643 086 541 722 430	103.797 830 780 847	869 909 704 615 088	29.631 353 806 316 400	0
2	371 608 007 538 699	12.942 292 051 494 400	18.630 057 857 970 300	6.835 245 883 862	5.928 855 591 348 520	15.180 015 118 357 200	5.632 907 628 451 760	4.500 656 274 942 400	0	0
3	8.099 124 189 298 390	22.423 625 939 355 700	19.909 541 732 292 300	9.275 883 602 694 080	4.186 062 130 644 810	16.868 636 929 550 900	6.642 009 251 176 360	30.559 337 496 641 600	0	0
4	8.316 765 884 214 670	21.437 339 408 562 200	22.018 417 440 775 200	805 933 237 743 854	35.688 613 564 305 600	3.632 665 161 642 430	18.436 524 495 493 300	10.034 167 436 508 000	4.628 770 536 877 080	0
5	9.092 223 456 290 960	18.110 150 923 612 500	1.797 898 633 892 620	6.546 599 974 207 940	31.013 573 752 420 400	39.841 081 338 184 400	11.558 279 443 446 300	31.997 992 727 424 700	4.075 075 425 430 030	0
6	5.584 086 638 456 080	2.883 133 237 695 160	2.874 868 773 904 610	754 486 878 877 965	3.266 783 629 116 730	2.804 679 159 334 870	8.399 734 640 152 750	54.917 861 841 994 400	25.597 082 275 565 200	0
7	10.223 862 164 528 700	24.807 173 527 013 900	28.749 716 543 528 200	75.134 084 658 313 000	39.366 339 551 509 600	2.836 516 335 078 440	13.789 695 317 519 800	8.460 355 617 402 350	2.672 988 736 934 770	0
8	8.635 848 718 500 730	20.336 152 258 657 000	13.672 091 763 901 600	4.563 008 605 599 700	3.033 097 711 591 810	47.460 764 484 244 800	1.236 381 669 870 520	62.798 308 962 925 100	4.401 434 715 445 480	0
9	11.898 857 909 025 100	11.898 857 909 025 100	14.285 583 854 224 500	7.804 173 553 073 090	288.646 940 746 221	3.893 755 658 712 610	1.270 604 896 865 790	53.928 845 767 512 200	35.900 171 809 576 100	0
10	11.180 284 470 721 500	22.723 146 923 797 400	2.548 450 849 098 780	9.077 200 016 914 390	40.404 163 468 408 900	5.638 854 814 810 940	1.792 780 641 128 500	7.197 660 103 221 910	4.370 561 936 655 490	0
11	7.360 640 105 838 250	16.552 079 725 952 800	32.452 614 409 143 800	7.550 700 906 704 110	32.662 435 345 560 100	42.538 341 949 538 700	15.586 810 438 033 100	7.874 001 566 430 470	36.622 917 838 524 500	0
12	7.974 521 648 923 860	21.869 330 048 866 600	1.876 765 668 181 340	8.110 384 501 123 870	36.409 823 046 204 800	14.525 745 697 593 200	7.648 591 117 965 150	4.011 718 108 339 780	0	0
13	7.119 824 384 264 550	15.670 499 334 039 200	18.730 813 653 342 700	36.060 360 905 057 200	2.823 440 504 739 600	34.771 502 726 194 300	15.929 535 908 825 600	795 007 783 369 744	3.445 756 223 321 890	0
14	7.49 623 220 797 336	1.501 749 233 951 360	27.331 361 961 927 700	6.838 223 470 687 500	29.941 578 134 685 800	37.976 183 482 577 200	19.370 807 181 232 100	7.65 099 955 279 583	4.413 974 182 974 900	0
15	6.347 271 760 539 310	18.674 288 066 057 600	4.106 523 476 453 930	9.629 596 276 480 580	3.644 876 872 467 600	516 743 281 893 657	11.539 781 191 539 400	7.507 161 728 663 770	4.376 348 290 691 890	0
16	7.051 785 800 016 840	21.104 940 606 054 500	30.980 600 786 788 800	10.094 796 011 661 400	3.151 412 672 443 020	2.039 702 184 072 240	5.665 160 378 979 330	4.288 428 857 506 180	0	0
17	9.181 560 007 151 530	27.381 380 665 980 000	2.404 132 628 006 120	6.904 989 726 470 090	3.983 505 168 222 770	4.779 746 418 621 770	13.387 340 780 225 500	714 573 622 129 516	4.503 660 796 179 120	0
18	8.975 464 347 533 960	27.935 716 677 009 200	19.460 398 131 232 100	6.204 320 858 892 470	43.144 398 999 034 800	1.288 875 905 430 390	6.382 123 709 666 390	24.360 855 903 052 700	0	0
19	7.371 050 302 429 530	21.449 661 045 715 600	25.630 320 036 999 700	4.432 669 290 372 120	33.575 443 859 606 500	46.991 455 147 923 500	12.509 163 940 498 600	6.279 727 715 266 120	25.602 991 476 149 100	0
20	22.743 504 835 115 500	2.230 556 741 374 140	10.333 917 888 218 600	31.093 085 831 787 800	5.548 200 864 605 430	1.633 169 312 269 440	45.382 815 177 870 900	413 342 264 357 917	0	0
21	6.660 212 026 118 100	16.828 374 685 651 800	30.944 363 591 242 600	5.858 769 130 547 580	5.236 712 975 009 440	1.788 423 519 296 480	770 423 180 517 003	37.487 012 410 996 100	0	0
22										

Gambar 2. Dataset Water Potability.

```
df = pd.read_csv('/content/water_potability.csv')
df.head()
```

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability
0	NaN	204.890455	20791.318981	7.300212	368.516441	564.308654	10.379783	86.990970	2.963135	0
1	3.716080	129.422921	18630.057858	6.635246	NaN	592.885359	15.180013	56.329076	4.500656	0
2	8.099124	224.236259	19909.541732	9.275884	NaN	418.606213	16.868637	66.420093	3.055934	0
3	8.316766	214.373394	22018.417441	8.059332	356.886136	363.266516	18.436524	100.341674	4.628771	0
4	9.092223	181.101509	17978.986339	6.546600	310.135738	398.410813	11.558279	31.997993	4.075075	0

Gambar 3. lima baris pertama dari *Dataset Water Potability*.

2.1.1. Variabel Independen (Fitur):

Variabel independen adalah variabel bebas yang memengaruhi variabel dependen (dalam hal ini status potabilitas air: layak/tidak layak) parameter kualitas air yang digunakan sebagai input model prediksi [12].

1. *Ph* : Merupakan mengukur tingkat keasaman atau kebasaan air. Skala pH berkisar dari 0 hingga 14.
2. *Hardness* (Kekerasan) : Diukur dalam mg/L, merepresentasikan konsentrasi *ion kalsium* dan *magnesium*.
3. *Solids* (Total Padatan Terlarut) : Diukur dalam ppm (*parts per million*), menunjukkan jumlah total zat padat yang larut dalam air.
4. *Chloramines* : Konsentrasi *Kloramin* (mg/L), senyawa desinfektan yang digunakan dalam pengolahan air.
5. *Sulfate* : Konsentrasi *sulfat* (mg/L), senyawa yang secara alami ada di air.
6. *Conductivity* (Konduktivitas) : Diukur dalam $\mu S/cm$, menunjukkan kemampuan air untuk menghantarkan listrik, yang berkaitan dengan kandungan *ion*.
7. *Organic Carbon* (Karbon Organik) : Konsentrasi total karbon organik (ppm), indikator bahan organik.
8. *Trihalomethanes* : Konsentrasi *trihalometana* ($\mu g/L$), produk sampingan dari proses klorinasi.
9. *Turbidity* (Kekeruhan) : Diukur dalam NTU (*Nephelometric Turbidity Units*), menunjukkan tingkat kejernihan air.

2.1.2. Variabel Dependen (Target):

Variabel dependen merupakan status potabilitas air (apakah air tersebut layak konsumsi atau tidak yang di pengaruhi oleh beberapa parameter [13]. 1. *Potability*: Variabel biner yang menunjukkan kelayakan air untuk dikonsumsi.

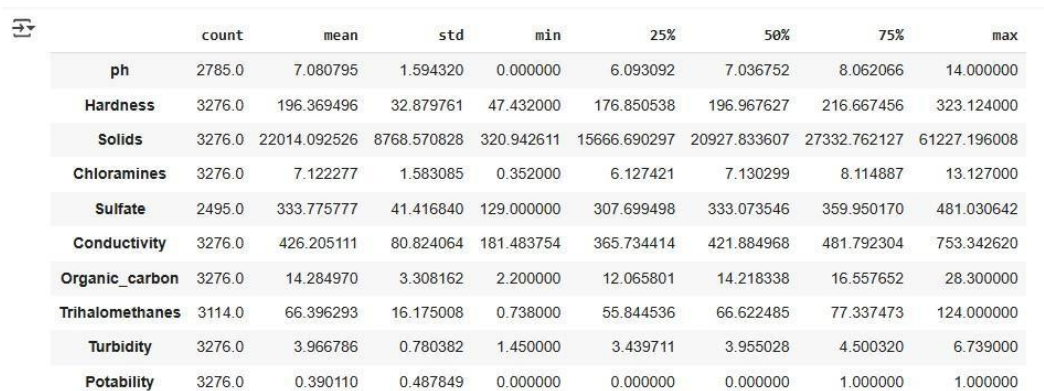
- 1) 1: *Potable* (Layak Minum)
- 2) 0: *Not Potable* (Tidak Layak Minum)

2.2. Pra-pemrosesan Data (*Data Preprocessing*)

Kualitas data mentah seringkali tidak langsung ideal untuk langsung melakukan pemodelan. Oleh sebab itu, serangkaian langkah pra-pemrosesan data dijalankan untuk membersihkan dan menyiapkan dataset sesuai dengan yang di harapkan.

1. *Analisis Data Eksploratori (Exploratory Data Analysis - EDA)*: Pada tahap awal ini bertujuan untuk mengeksplorasi data melalui visualisasi, ringkasan statistik, dan mengidentifikasi pola untuk memahami karakteristik dari dataset [14]. Ini mencakup pemeriksaan dari dimensi data (3.276 baris, 10 kolom), tipe data yang ada setiap kolom (9

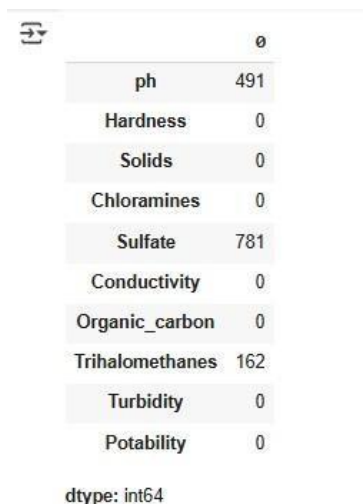
float64 dan 1 *int64*), dan ringkasan statistik (rata-rata, standar *deviasi*, *min*, *maks*, *kuartil*) untuk setiap fitur. Visualisasi seperti *histogram*, *boxplot*, dan *heatmap* korelasi juga dibuat untuk mengidentifikasi distribusi data, outlier, dan hubungan antar variabel. Ditemukan bahwa distribusi kelas target tidak seimbang, dengan 61% data termasuk dalam kelas "Tidak Layak Minum" dan 39% dalam kelas "Layak Minum".



	count	mean	std	min	25%	50%	75%	max
ph	2785.0	7.080795	1.594320	0.000000	6.093092	7.036752	8.062066	14.000000
Hardness	3276.0	196.369496	32.879761	47.432000	176.850538	196.967627	216.667456	323.124000
Solids	3276.0	22014.092526	8768.570828	320.942611	15666.690297	20927.833607	27332.762127	61227.196008
Chloramines	3276.0	7.122277	1.583085	0.352000	6.127421	7.130299	8.114887	13.127000
Sulfate	2495.0	333.775777	41.416840	129.000000	307.699498	333.073546	359.950170	481.030642
Conductivity	3276.0	426.205111	80.824064	181.483754	365.734414	421.884968	481.792304	753.342620
Organic_carbon	3276.0	14.284970	3.308162	2.200000	12.065801	14.218338	16.557652	28.300000
Trihalomethanes	3114.0	66.396293	16.175008	0.738000	55.844536	66.622485	77.337473	124.000000
Turbidity	3276.0	3.966786	0.780382	1.450000	3.439711	3.955028	4.500320	6.739000
Potability	3276.0	0.390110	0.487849	0.000000	0.000000	0.000000	1.000000	1.000000

Gambar 4. Ringkasan Statistik Dataset.

2. Penanganan Nilai yang Hilang (*Handling Missing Values*): Merupakan proses mengatasi data yang hilang agar data hasil analisis dan pemodelan tidak bias [15]. Hasil *EDA* menunjukkan adanya sejumlah nilai yang hilang (*missing values*) pada tiga kolom: *ph* (491 nilai hilang), *Sulfate* (781 nilai hilang), dan *Trihalomethanes* (162 nilai hilang). Untuk menangani masalah ini, digunakan strategi imputasi rata-rata berdasarkan kelompok. Nilai yang hilang pada setiap kolom akan diisi dengan nilai rata-rata dari kolom tersebut, yang dihitung secara terpisah untuk setiap kelas target (*Potability* 0 dan 1). Pendekatan ini dinilai lebih baik daripada imputasi rata-rata global karena mampu mempertahankan struktur data asli dengan lebih baik.



	0
ph	491
Hardness	0
Solids	0
Chloramines	0
Sulfate	781
Conductivity	0
Organic_carbon	0
Trihalomethanes	162
Turbidity	0
Potability	0

dtype: int64

Gambar 5. Nilai hilang Pada Dataset.

3. Pemeriksaan Duplikat: Pemeriksaan ini dilakukan untuk memastikan bahwa tidak ada data duplikat yang terdapat di dalam dataset. Hasilnya menunjukkan tidak ada baris yang identik, sehingga tidak ada data yang perlu dihapus.

```
[15] df.duplicated().sum()
```

```
np.int64(0)
```

Gambar 6. Hasil Pemeriksaan Duplikat.

4. Standardisasi Fitur (*Feature Standardization*): Fitur-fitur dalam dataset memiliki skala dan rentang nilai yang berbeda (misalnya, *Solids* memiliki nilai puluhan ribu, sementara *pH* berkisar antara 0-14). Untuk memastikan setiap fitur memberikan kontribusi yang setara selama pelatihan model, dilakukan standardisasi menggunakan *StandardScaler* dari pustaka *scikitlearn*. Proses ini mengubah setiap fitur sehingga memiliki rata-rata 0 dan standar deviasi 1.

	ph	Hardness	Solids	Chloramines	Sulfate	\
0	7.085378	204.890455	20791.318981	7.300212	368.516441	
1	3.716080	129.422921	18630.057858	6.635246	334.564290	
2	8.099124	224.236259	19909.541732	9.275884	334.564290	
3	8.316766	214.373394	22018.417441	8.059332	356.886136	
4	9.092223	181.101509	17978.986339	6.546600	310.135738	
...	...	...	...	...	...	
3271	4.668102	193.681735	47580.991603	7.166639	359.948574	
3272	7.808856	193.553212	17329.802160	8.061362	332.566990	
3273	9.419510	175.762646	33155.578218	7.350233	332.566990	
3274	5.126763	230.603758	11983.869376	6.303357	332.566990	
3275	7.874671	195.102299	17404.177061	7.509306	332.566990	
...						
	Conductivity	Organic_carbon	Trihalomethanes	Turbidity	Potability	
0	564.308654	10.379783	86.990970	2.963135	0	
1	592.885359	15.180013	56.329076	4.500656	0	
2	418.606213	16.868637	66.420093	3.055934	0	
3	363.266516	18.436524	100.341674	4.628771	0	
4	398.410813	11.558279	31.997993	4.075075	0	
...	...	...	...	...	...	
3271	526.424171	13.894419	66.687695	4.435821	1	
3272	392.449580	19.903225	66.539684	2.798243	1	
3273	432.044783	11.039070	69.845400	3.298875	1	
3274	402.883113	11.168946	77.488213	4.708658	1	
3275	327.459760	16.140368	78.698446	2.309149	1	

[3276 rows x 10 columns]

Gambar 7. Data Standardization.

2.3. Pemodelan dengan Algoritma *Decision Tree*

Algoritma yang digunakan dalam penelitian ini adalah *Decision Tree Classifier* keunggulan *algoritma* ini adalah kemampuannya dalam menghasilkan model yang mudah diinterpretasi dan di visualkan ini sangat efektif dalam konteks yang menggunakan pemahaman yang jelas bagaimana keputusan dibuat [11]. *Algoritma* ini dipilih karena model yang dihasilkannya mudah diinterpretasikan dan efisien secara komputasi.

1. Pembagian Data (*Train-Test Split*): Dataset yang telah diproses dibagi menjadi dua set: lalu data latih (*training set*) dan data uji (*testing set*) dengan rasio 80:20. Ini menghasilkan 2.620 sampel untuk dilakukan pelatihan dan 656 sampel untuk pengujian. Parameter *random_state=42* digunakan untuk memastikan hasil dari pembagian data konsisten dan dapat direproduksi.

```
[ ] X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Menampilkan jumlah data untuk masing-masing set
print(f"Jumlah data untuk training set: {X_train.shape[0]} sampel")
print(f"Jumlah data untuk testing set: {X_test.shape[0]} sampel")

Jumlah data untuk training set: 2620 sampel
Jumlah data untuk testing set: 656 sampel
```

Gambar 8. Hasil Split Data.

2. Pelatihan Model: *Model Decision Tree Classifier* dari *scikit-learn* diinisialisasi (dengan *random_state=42* untuk konsistensi) dan dilatih menggunakan data latih (*X_train* dan *y_train*). Selama proses pelatihan, *algoritma* secara *rekursif* akan membagi data berdasarkan fitur yang memberikan pemisahan terbaik, diukur menggunakan kriteria *Gini Impurity*. Tujuannya adalah dengan menciptakan node-node yang dihasilkan se-"murni" mungkin (berisi mayoritas satu kelas). Proses ini berlangsung hingga kriteria penghentian telah terpenuhi, seperti kedalaman pohon maksimum atau jumlah sampel minimum yang dimiliki per daun.

2.4. Evaluasi Model

Kinerja model yang telah dilatih dievaluasi menggunakan data uji, yang belum pernah dilihat oleh model sebelumnya. Metrik evaluasi yang digunakan adalah:

1. Akurasi (*Accuracy*): Persentase total prediksi yang benar.

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Total Data}}$$

2. Presisi (*Precision*): Dari semua yang diprediksi sebagai kelas positif, berapa persen yang benar-benar positif. Metrik ini penting untuk meminimalkan *False Positives*.

$$\text{Presisi} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}}$$

3. *Recall (Sensitivitas)*: Dari semua yang sebenarnya kelas positif, berapa persen yang berhasil diprediksi sebagai positif. Metrik ini penting untuk meminimalkan *False Negatives*.

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Negative (FN)}}$$

4. *F1-Score*: Rata-rata harmonik dari Presisi dan *Recall*, memberikan skor tunggal yang menyeimbangkan kedua metrik tersebut.

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5. *Confusion Matrix*: Sebuah tabel yang merangkum hasil prediksi, menunjukkan jumlah *True Positives (TP)*, *True Negatives (TN)*, *False Positives (FP)*, dan *False Negatives (FN)*.

	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	TP	FN
<i>Actual Negative</i>	FP	TN

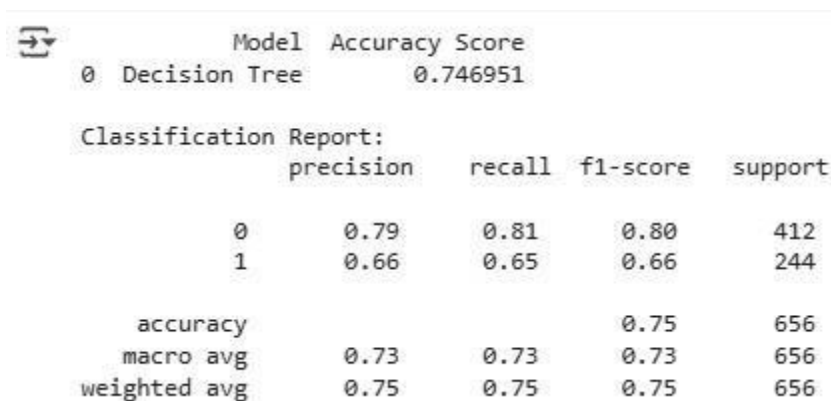
Tabel 1. Confusion Matrix.

2. HASIL DAN PEMBAHASAN

Bagian ini akan menyajikan hasil dari penerapan model *Decision Tree* pada dataset potabilitas air dan analisis mendalam terhadap hasil tersebut.

3.1. Hasil Pelatihan dan Evaluasi Model

Setelah model *Decision Tree* dilatih menggunakan 2.620 sampel data latih, kinerjanya diuji pada 656 sampel data uji. Hasil evaluasi kinerja model dirangkum dalam Laporan Klasifikasi (*Classification Report*) pada Gambar 8.



Model	Accuracy Score
0 Decision Tree	0.746951

Classification Report:				
	precision	recall	f1-score	support
0	0.79	0.81	0.80	412
1	0.66	0.65	0.66	244
accuracy			0.75	656
macro avg	0.73	0.73	0.73	656
weighted avg	0.75	0.75	0.75	656

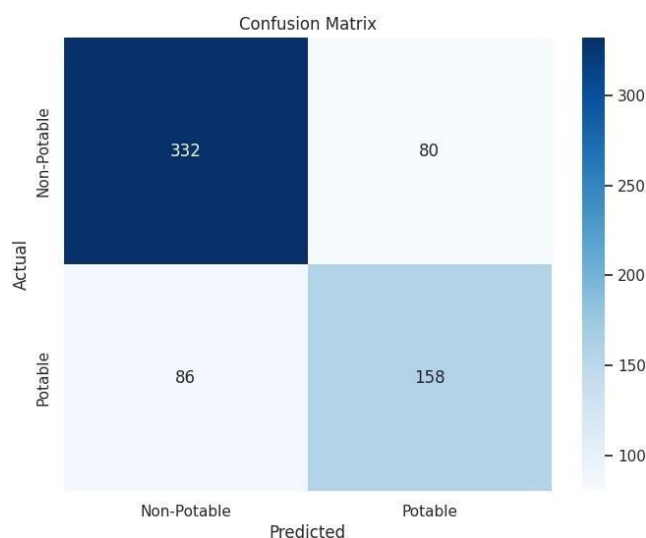
Gambar 8. Laporan Klasifikasi Kinerja Model *Decision Tree*.

Secara keseluruhan, model mencapai akurasi sebesar 74.7% (dibulatkan menjadi 75% dalam laporan), yang berarti model mampu memprediksi dengan benar sekitar tiga dari empat sampel air dalam data uji. Namun, akurasi saja tidak cukup untuk memberikan gambaran lengkap, terutama pada dataset yang tidak seimbang. Analisis lebih detail pada laporan klasifikasi menunjukkan adanya perbedaan kinerja yang signifikan antara kedua kelas:

1. Untuk Kelas 0 (Tidak Layak Minum): Model menunjukkan kinerja yang kuat. Dengan *Precision* 0.79, artinya ketika model memprediksi air tidak layak minum, 79% di antaranya benar. *Recall* sebesar 0.81 menunjukkan bahwa model berhasil mengidentifikasi 81% dari semua sampel air yang sebenarnya tidak layak minum. *F1-Score* sebesar 0.80 mengonfirmasi kinerja yang baik untuk kelas ini.
2. Untuk Kelas 1 (Layak Minum): Kinerja model untuk kelas ini lebih rendah secara signifikan. *Precision* 0.66, *Recall* 0.65, dan *F1-Score* 0.66 mengindikasikan bahwa model mengalami lebih banyak kesulitan dalam mengidentifikasi air yang layak minum.

3.2. Analisis *Confusion Matrix*

Untuk memahami pola kesalahan model lebih dalam, *confusion matrix* disajikan pada Gambar 9 dibawah ini.



Gambar 9. Confusion Matrix Hasil Prediksi Model.

Dari *confusion matrix* tersebut, dapat dijabarkan:

1. True Negatives (TN) = 332: Model dengan benar memprediksi 332 sampel sebagai "Tidak Layak Minum".
2. True Positives (TP) = 158: Model dengan benar memprediksi 158 sampel sebagai "Layak Minum".
3. False Positives (FP) = 80: Model salah memprediksi 80 sampel sebagai "Layak Minum", padahal sebenarnya "Tidak Layak Minum" (Kesalahan Tipe I).
4. False Negatives (FN) = 86: Model salah memprediksi 86 sampel sebagai "Tidak Layak Minum", padahal sebenarnya "Layak Minum" (Kesalahan Tipe II).

3.3. Pembahasan

Hasil evaluasi menunjukkan bahwa model *Decision Tree* memiliki kemampuan prediktif yang cukup baik, namun kinerjanya bias terhadap kelas mayoritas (kelas 0, "Tidak Layak Minum"). Fenomena ini umum terjadi pada dataset yang tidak seimbang. Karena terdapat lebih banyak contoh air tidak layak minum (61% dari data), model menjadi lebih "terlatih" untuk mengenali pola kelas tersebut, yang mengakibatkan nilai *Recall* dan *Precision* yang lebih tinggi untuk kelas 0.

Sebaliknya, jumlah sampel yang lebih sedikit untuk kelas 1 (39% dari data) membuat model kurang mampu menggeneralisasi pola untuk air yang layak minum. Hal ini tercermin dari nilai *Recall* yang lebih rendah (0.65), yang berarti model gagal mengidentifikasi sekitar 35% dari semua air yang sebenarnya aman untuk dikonsumsi (*False Negatives*). Dari perspektif kesehatan publik, kesalahan *False Positives* (menganggap air tercemar sebagai air aman) mungkin lebih berbahaya, namun *False Negatives* yang tinggi juga menjadi masalah karena dapat menyebabkan sumber air bersih yang potensial tidak dimanfaatkan.

Meskipun terdapat bias, keunggulan utama dari model *Decision Tree* adalah interpretasinya. Aturan-aturan keputusan yang dihasilkan oleh pohon (misalnya, if Sulfate $\leq$ X and pH $>$ Y, then...) dapat diekstrak dan dianalisis untuk memahami fitur mana yang paling berpengaruh pada setiap titik keputusan. Berdasarkan struktur pohon yang dihasilkan, fitur seperti *Sulfate*, *pH*, dan *Solids* sering muncul di level atas pohon, menunjukkan bahwa ketiganya adalah pembeda yang relatif penting dalam *dataset* ini, meskipun korelasi liniernya rendah.

Secara keseluruhan, penelitian ini berhasil menunjukkan bahwa *Decision Tree* adalah algoritma yang layak untuk tugas ini. Akurasi 74.7% merupakan titik awal yang solid. Keterbatasan utama terletak pada sifat *dataset* yang tidak seimbang, yang secara langsung memengaruhi kemampuan model untuk memprediksi kelas minoritas.

Hasil Prediksi:

	ph	Hardness	Solids	Chloramines	Sulfate	Conductivity	\
0	-0.001626	-0.396889	-0.181736	0.141929	-0.015839	-0.862984	
1	-0.304021	-0.231339	1.223451	-0.199252	0.004311	-1.106997	
2	0.518536	0.847637	0.135567	-0.746663	-0.919693	-0.473669	
3	0.049721	-0.410145	-1.734696	-2.081479	-1.549677	0.019900	
4	-0.323037	-0.528300	0.489763	1.386439	-0.031102	0.861638	
..	...	...	...	...	...	...	
651	2.009315	0.854007	-0.804269	0.201826	0.090619	1.122180	
652	-0.151935	1.074396	0.058776	-0.901588	-0.031102	0.864593	
653	0.253960	0.652421	-0.297745	-1.554892	0.517491	0.833683	
654	-0.837748	-0.349484	2.490852	-1.716377	2.182425	-1.168487	
655	-0.001626	0.739991	0.686993	0.548097	-0.719405	-0.061063	

	Organic_carbon	Trihalomethanes	Turbidity	Actual	Predicted
0	1.804635	0.041876	1.189622	0	0
1	0.138518	0.093942	0.773151	1	1
2	-0.257313	-1.469386	-1.889486	0	1
3	-1.355564	1.494488	-0.096291	0	0
4	-0.446127	0.751439	0.633546	1	1
..	...	...	...	...	...
651	1.402581	-0.062671	0.285457	0	0
652	0.579435	1.584510	1.227598	1	1
653	0.063462	-0.578144	-0.152566	0	1
654	0.320976	-0.426062	1.986511	1	0
655	0.998789	1.293324	-0.501061	0	1

[656 rows x 11 columns]

Gambar 10. Hasil Prediksi Model.

4. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *Decision Tree* untuk membangun model klasifikasi dan prediksi potabilitas air minum. Berdasarkan serangkaian analisis pada dataset publik, model yang dikembangkan mampu mencapai tingkat akurasi keseluruhan sebesar 74.7%. Hasil ini menunjukkan bahwa *machine learning*, khususnya *Decision Tree*, memiliki potensi signifikan sebagai alat bantu untuk penilaian kualitas air secara otomatis dan efisien.

Analisis kinerja yang lebih mendalam mengungkap bahwa model menunjukkan performa yang lebih unggul dalam mengidentifikasi air yang tidak layak minum (*Non-Potable*) dibandingkan dengan air yang layak minum (*Potable*). Fenomena ini dapat diatribusikan pada ketidakseimbangan distribusi kelas dalam dataset yang digunakan, di mana jumlah sampel untuk kelas "Tidak Layak Minum" lebih dominan. Meskipun demikian, model ini memberikan dasar yang kuat dan, yang terpenting, aturan keputusan yang dapat diinterpretasikan, yang merupakan keunggulan utama dari algoritma *Decision Tree*. Fitur-fitur seperti *Sulfate*, *pH*, dan *Solids* teridentifikasi sebagai beberapa parameter kunci dalam proses klasifikasi.

5. SARAN

Berdasarkan temuan dan keterbatasan dalam penelitian ini, beberapa saran untuk penelitian di masa depan dapat diajukan guna meningkatkan kinerja dan robustisitas model:

1. Penanganan Data Tidak Seimbang: Kinerja model, terutama untuk kelas minoritas ("Layak Minum"), dapat ditingkatkan secara signifikan dengan menerapkan teknik penanganan data tidak seimbang. Metode seperti *oversampling* (misalnya, *SMOTE - Synthetic Minority Over-sampling Technique*) atau *undersampling* pada data latih perlu dieksplorasi.

2. Eksplorasi *Algoritma Lain*: Meskipun *Decision Tree* menawarkan interpretasi yang baik, *algoritma ensemble* yang dibangun di atasnya, seperti *Random Forest* atau *gradient boosting (XGBoost, LightGBM)*, seringkali memberikan akurasi yang lebih tinggi karena kemampuannya mengurangi *overfitting* dan menangkap pola yang lebih kompleks. Studi komparatif dengan *algoritma-algoritma* ini akan sangat bermanfaat.
3. Rekayasa dan Seleksi Fitur (*Feature Engineering & Selection*): Melakukan analisis yang lebih mendalam untuk memilih fitur yang paling relevan atau menciptakan fitur baru dari kombinasi fitur yang ada dapat membantu meningkatkan daya prediksi model.
4. Validasi dengan Dataset yang Lebih Besar dan Bervariasi: Menguji model pada dataset yang lebih besar atau dataset dari lokasi geografis yang berbeda akan membantu memvalidasi kemampuan generalisasi model dan memastikan model tidak hanya berkinerja baik pada data spesifik yang digunakan dalam penelitian ini.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Dr. Lili Tanti, M. Kom, selaku dosen pembimbing, atas arahan, bimbingan, dan masukan yang konstruktif selama proses penyusunan dan penyelesaian penelitian ini. Ucapan terima kasih juga ditujukan kepada seluruh jajaran dosen di Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Potensi Utama, atas ilmu dan dukungan akademis yang telah diberikan.

REFERENSI

- [1] BPS (Badan Pusat Statistik), "Akses Terhadap Air Minum Layak Memang Menunjukkan Tren Peningkatan.," Indonesia, 2023. Accessed: Oct. 17, 2025. [Online]. Available: <https://www.bps.go.id/id/statistics-table/2/ODU0IzI=/persentase-rumah-tangga-yangmemiliki-akses-terhadap-sumber-air-minum-layak-menurut-provinsi-dan-klasifikasidesa--persen-.html>
- [2] Kementrian Kesehatan, "Kemenkes Luncurkan Hasil Surveilans Kualitas Air Minum Rumah Tangga di Indonesia Tahun 2023," <https://kemkes.go.id/id/kemenkes-luncurkanhasil-surveilans-kualitas-air-minum-rumah-tangga-di-indonesia-tahun-2023>, Jakarta, Dec. 20, 2024.
- [3] Kemenkes RI., "Kemenkes: Lebih dari 40 Persen Air Minum Isi Ulang di RI Positif Mengandung E Coli Baca artikel detikHealth, 'Kemenkes: Lebih dari 40 Persen Air Minum Isi Ulang di RI Positif Mengandung E Coli' selengkapnya <https://health.detik.com/berita-detikhealth/d-7695484/kemenkes-lebih-dari-40-persenair-minum-isi-ulang-di-ri-positif-mengandung-e-coli>. Download Apps Detikcom Sekarang <https://apps.detik.com/detik/>," *Detik Health*, Dec. 20, 2024.
- [4] M. Sobron *et al.*, *IMPLEMENTASI ARTIFICIAL INTELLIGENCE PADA SYSTEM MANUFAKTUR TERPADU*. 2021.
- [5] S. Prasetyo and T. Dewayanto, "PENERAPAN MACHINE LEARNING, DEEP LEARNING, DAN DATA MINING DALAM DETEKSI KECURANGAN LAPORAN KEUANGAN-A SYSTEMATIC LITERATURE REVIEW," *DIPONEGORO JOURNAL OF ACCOUNTING*, vol. 13, no. 3, pp. 1–12, 2024, [Online]. Available: <http://ejournals1.undip.ac.id/index.php/accounting>
- [6] L. Savitri and R. Nursalim, "Klasifikasi Kualitas Air Minum menggunakan Penerapan Algoritma Machine Learning dengan Pendekatan Supervised Learning," *Diophantine Journal of Mathematics and Its Applications*, vol. 2, no. 01, pp. 30–36, Jun. 2023, doi: 10.33369/diophantine.v2i01.28260.

- [7] S. Barad *et al.*, “Comparative approach of decision tree and CWQI analysis for classification of groundwater with a special reference to fluoride ion in drought-prone Boudh district of Odisha, India,” *Sustain Water Resour Manag*, vol. 7, no. 6, p. 94, Dec. 2021, doi: 10.1007/s40899-021-00582-0.
- [8] B.-G. Jo, W.-S. Jung, S.-H. Nam, and Y.-D. Kim, “Prediction of Cyanobacteria Using Decision Tree Algorithm and Sensor Monitoring Data,” *Applied Sciences*, vol. 13, no. 22, p. 12266, Nov. 2023, doi: 10.3390/app132212266.
- [9] W. Chen, D. Xu, B. Pan, Y. Zhao, and Y. Song, “Machine Learning-Based Water Quality Classification Assessment,” *Water (Basel)*, vol. 16, no. 20, p. 2951, Oct. 2024, doi: 10.3390/w16202951.
- [10] D. Danuri and M. Mohd Pozi, “Machine Learning Approaches for Fish Pond Water Quality Classification: Random Forest, Gaussian Naive Bayes, and Decision Tree Comparison,” in *Proceedings of the 11th International Applied Business and Engineering Conference, ABEC 2023, September 21st, 2023, Bengkalis, Riau, Indonesia*, EAI, 2024. doi: 10.4108/eai.21-9-2023.2342964.
- [11] developerghost, “Water Potability,” <https://www.kaggle.com/datasets/developerghost/water-potability>.
- [12] K. Abdi, A. Warjaya, I. Muthmainnah, and P. H. Pahutar, “Penerapan Algoritma Random Forest dalam Prediksi Kelayakan Air Minum,” *Jurnal Ilmu Komputer dan Informatika*, vol. 3, no. 2, pp. 81–88, Jan. 2024, doi: 10.54082/jiki.81.
- [13] I. Ivanov and B. Toleva, “Predicting the Water Potability Index Using Machine Learning,” *Environment and Ecology Research*, vol. 11, no. 4, pp. 537–542, Aug. 2023, doi: 10.13189/eer.2023.110402.
- [14] S. Dhummad, “The Imperative of Exploratory Data Analysis in Machine Learning,” *Scholars Journal of Engineering and Technology*, vol. 13, no. 01, pp. 30–44, Jan. 2025, doi: 10.36347/sjet.2025.v13i01.005.
- [15] J. Peng, J. Hahn, and K.-W. Huang, “Handling Missing Values in Information Systems Research: A Review of Methods and Assumptions,” *Information Systems Research*, vol. 34, no. 1, pp. 5–26, Mar. 2023, doi: 10.1287/isre.2022.1104.
- [16] Dewi, R., & Rizki, B. (2025). Pengaruh penggunaan fitur SpayLater dan flash sale Shopee terhadap perilaku konsumtif mahasiswa/i Muslim (Studi kasus pada mahasiswa/i FEB Universitas Potensi Utama). *Jurnal Media Informatika*, 6(3), 1634–1641. <https://doi.org/10.55338/jumin.v6i3.5174>
- [17] Dermawan, D. A., & Pohan, Y. A. (2024). Pengaruh disiplin, pelatihan kerja dan pengembangan karir terhadap kinerja karyawan pada PT. Bank Sumut Cabang Imam Bonjol. *Jurnal Dunia Pendidikan*, 5(2), 450–465. <https://doi.org/10.55081/jurdip.v5i2.3125>
- [18] Pohan, Y. A. (2024). Pengaruh kompensasi, lingkungan kerja non fisik dan stress kerja terhadap turnover intention karyawan pada PT. Mitra Tsunami Marine. *Jurnal Widya*, 5(2).
- [19] Gianing, E., & Pohan, Y. A. (2024). Pengaruh kepemimpinan transformasional, lingkungan kerja, pengembangan karir terhadap kinerja pegawai di Rumah Sakit Sinar Husni Medan. *Jurnal Widya*, 5(2), 1972–1978. <https://doi.org/10.54593/awl.v5i2.407>
- [20] Ramadhan, F., Wahyuni, E. S., & Prabowo, A. (2024). Perencanaan pensiun pada pekerja manufaktur di Kota Batam: Mengidentifikasi pengaruh sosial dan perilaku

- menabung. *Jurnal Menara Ekonomi: Penelitian dan Kajian Ilmiah Bidang Ekonomi*, 10(2), 65–78. <https://doi.org/10.31869/me.v10i2.5935>
- [21] Pane, S. H., & Prabowo, A. (2024). The influence of service quality and physical evidence on student satisfaction Potensi Utama University. *ePaper Bisnis: International Journal of Entrepreneurship and Management*, 1(4), 01–16. <https://doi.org/10.61132/epaperbisnis.v1i4.129>
- [22] Prabowo, A., Tanjung, Y., Wahyuni, E. S., & Aspan, H. (2025). Transformasi destinasi pemasaran: Model integratif pengembangan pariwisata berkelanjutan di kawasan objek wisata Sidebu-Debu, Kabupaten Karo Sumatera Utara. *Economic and Education Journal (Ecoducation)*, 7(2), 455–476. <https://doi.org/10.33503/ecoducation.v7i2.1687>
- [23] Lubis, K. S., Adam, A. A., Yusriadi, Y., Tanjung, A. P., & Azzahra, N. S. (2025). Analisis SWOT sebagai strategi penguatan daya saing UMKM depot air isi ulang pasca pendampingan manajerial dan digitalisasi. *Jurnal Bisnis Kompetitif*, 4(2), 154–159. <https://doi.org/10.35446/bisniskompetitif.v4i2.2368>
- [24] Gulo, M. K., & Tanjung, A. P. (2025). Analisis brand trust, sales promotion, shopping emotion, terhadap keputusan pembelian konsumen brand fashion UNIQLO. *Jurnal Dunia Pendidikan*, 5(4), 1289–1298. <https://doi.org/10.55081/jurdip.v5i4.3552>
- [25] Kadariah, S., Syahrizal, R., & Sari, S. D. N. (2024). Eksistensi dan transformasi budaya lokal dalam strategi ketahanan ekonomi di Kota Padang Sumatera Barat. *HUMAN FALAH: Jurnal Ekonomi dan Bisnis Islam*, 11(2), 22–35. <https://doi.org/10.30829/hf.v11i2.21763>
- [26] Nazli, P., & Kadariah, S. (2024). Implementasi akad al-qardh pada perbankan syariah (Studi kasus PT. Bank Sumut Cabang Syariah Medan). *Jurnal Widya*, 5(2), 2266–2276. <https://doi.org/10.54593/awl.v5i2.436>
- [27] Aulanda, N. F., Hutagalung, M. A. K., Ridwan, M., & Aini, R. (2025). Effect of servant leadership on employee performance with work motivation as an intervening variable. *Digital Business: Tren Bisnis Masa Depan*, 16(1), 48–53. <https://doi.org/10.59651/digital.v16i1.264>
- [28] Gemilang, F., & Hutagalung, M. A. K. (2025). Analysis of Murabahah financing risk management strategies at PT. BPRS Al-Washliyah Medan. *BISEI: Jurnal Bisnis dan Ekonomi Islam*, 10(2), 139–147. <https://doi.org/10.33752/bisei.v10i2.10756>
- [29] Salsabilla, S., & Hutagalung, M. A. K. (2025). Analysis of the implementation of Islamic business ethics at Manna Supermarket Yogyakarta Campus. *BISEI: Jurnal Bisnis dan Ekonomi Islam*, 10(2), 108–116. <https://doi.org/10.33752/bisei.v10i2.10757>
- [30] Eka Putra, S., Nurmala Sari, T., Simatupang, J., Murtani, A., & Tanjung, Y. (2025). Analysis of the benefits of the 7P marketing mix strategy on the brand image of Butong Tea at PT Perkebunan Nusantara IV Bah Butong. *Jurnal Ilmiah Sosio-Ekonomika Bisnis*, 28(2), 30–37. <https://doi.org/10.22437/jiseb.v28i02.41506>