Pengembangan Sistem Klasifikasi Kualitas Air Minum Berbasis Web Menggunakan Algoritma *Decision Tree*

1st Syifa Melinda Naf'an
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
syifamelindanafan@student.telkomuniversity.ac.id

2th Meta Kallista
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
metakallista@telkomuniversity.ac.id

3th Ig.Prasetya Dwi Wibawa
Fakultas Teknik Elektro
Universitas Telkom
Bandung, Indonesia
prasdwibawa@telkomuniversity.ac.id

***Abstrak* — Kualitas air sangat penting untuk kehidupan manusia, namun tidak semua sumber air aman untuk dikonsumsi. Oleh karena itu, diperlukan identifikasi kelayakan air minum yang akurat dan cepat. Metode manual seperti STORET dan Indeks Pencemaran kurang efisien karena memakan waktu dan biaya yang tinggi. Oleh karena itu, digunakan teknologi *Machine Learning* dengan algoritma *Decision Tree* dan teknik SMOTE untuk menyeimbangkan data. Hasil penelitian menunjukkan bahwa model *Decision Tree* dengan *max_depth* = 4 menghasilkan performa yang paling optimal. Pada max_depth ini, model mencapai akurasi *training* sebesar 99.9% dan akurasi *testing* mencapai 100%. Waktu yang dibutuhkan untuk proses *training* adalah 0.03570 detik, sedangkan waktu *testing* adalah 0.00223 detik. Hasil evaluasi lainnya juga menunjukkan nilai AUC sebesar 1.00. Selain itu evaluasi juga dilakukan menggunakan *classification report* dan didapatkan hasil bahwa model memiliki presisi (*precision*) dan *recall* sebesar 1.00 untuk kelas "Air Layak Minum" dan "Air Tidak Layak Minum". Nilai f1-score juga sebesar 1.00 untuk kedua kelas, menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan data positif dan negatif.**

Kata kunci*—Decision Tree, Evaluasi. Kualitas Air, Teknologi Machine Learning.

I. PENDAHULUAN

Air yang bersih dan layak untuk dikonsumsi adalah sumber daya yang tak dapat diabaikan dalam kehidupan dan pembangunan. Kehadirannya dalam tubuh manusia memiliki peran penting, membentuk sekitar 50-70% dari total berat badan. Ketidakseimbangan air hanya sebesar 15% dari berat badan dapat berakibat fatal, bahkan berpotensi menyebabkan kematian[1-3]. Namun, tantangan dalam memenuhi kebutuhan akan air yang aman semakin meningkat. Penurunan kualitas air serta peningkatan pencemaran akibat aktivitas manusia dan perkembangan industri semakin menghambat akses terhadap air minum yang aman bagi manusia. Identifikasi kualitas air yang baik menjadi kunci untuk memastikan air yang aman dan layak dikonsumsi. Saat ini, metode tradisional seperti STORET dan WQI digunakan secara manual untuk menghitung kualitas air[4]. Namun, proses manual ini memakan waktu yang cukup lama, sekitar 10-30 hari tergantung pada jumlah parameter yang diukur. Selain itu, biaya yang diperlukan juga tinggi[4]. Oleh karena itu, pengembangan sistem otomatis yang efektif dan efisien diperlukan untuk mempercepat proses identifikasi kualitas air. Dalam hal ini, kemajuan teknologi dan penggunaan teknik *machine learning* memberikan peluang untuk mempercepat penentuan kualitas air melalui klasifikasi data berdasarkan parameter-parameter yang menggambarkan kualitas air.

Tujuan dari penelitian ini adalah mengembangkan sistem otomatis menggunakan algoritma *Decision Tree* untuk melakukan klasifikasi kualitas air secara cepat. Dengan menggunakan metode klasifikasi data dan teknik machine learning, sistem ini dapat memprediksi apakah air tersebut aman atau tidak untuk dikonsumsi berdasarkan parameter-parameter yang relevan. Keandalan dan kinerja model machine learning akan bergantung pada ketersediaan sampel air yang mencakup baik kualitas air yang layak minum maupun tidak, dalam jumlah yang seimbang. Masalah ketidakseimba ***kualitas air, teknologi machine learning,*** ngan data dalam klasifikasi kualitas air dapat diatasi dengan melakukan resampling, seperti menggunakan teknik SMOTE. Setelah model machine learning dibuat, sistem ini akan diimplementasikan ke dalam sebuah website yang memungkinkan pengguna untuk memasukkan parameter-parameter dan mendapatkan hasil kualitas air dengan cepat dan efisien. Diharapkan teknologi ini dapat mengurangi waktu dan biaya yang sebelumnya diperlukan dalam evaluasi kualitas air melalui tenaga ahli.

II. KAJIAN TEORI

A. Kualitas Air

Kualitas air merupakan faktor penting yang memengaruhi keberlanjutan sumber daya air dan memiliki dampak langsung terhadap kesehatan manusia serta lingkungan. Ketika membahas kualitas air, tidak dapat dipisahkan dari pembahasan mengenai parameter kualitas air dan standar kualitas air.

1. Parameter Kualitas Air

Parameter kualitas air merujuk pada berbagai faktor yang digunakan untuk menggambarkan sifat-sifat air dan komponen yang ada di dalamnya. Penentuan kualitas air sangat penting dalam memastikan bahwa air yang digunakan aman dan sesuai untuk berbagai keperluan, termasuk minum, kegiatan pertanian, industri, dan ekosistem perairan. Adapun parameter-parameter air yang menjadi dataset untuk

mengevaluasi model *Decision Tree* meliputi *E.Colli*, *Coliform*, Arsen, Kromium, Kadmium, Nitrit, Nitrat, Sianida, Selenium, Alumunium, Besi, Kesadahan, Klorida, Mangan, pH, Seng, Sulfat, Tembaga, Amonia, *Chlor*, BOD5, COD, Bau, Rasa, Warna, TDS, Kekeruhan, Suhu dan Potabilitas.

2. Standar Kualitas Air

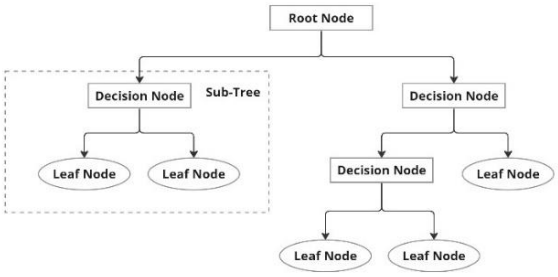
Kualitas air minum yang aman merupakan hal yang sangat penting untuk dipastikan. Untuk mencapai itu, perlu dilakukan identifikasi berdasarkan parameter air yang telah ditetapkan dalam Peraturan Menteri Kesehatan Republik Indonesia No. 492/MENKES/PER/IV/2010. Identifikasi ini melibatkan pengukuran kondisi fisik, karakteristik kimiawi, dan biologis air.

B. Algoritma Klasifikasi

Algoritma klasifikasi adalah suatu metode atau pendekatan yang digunakan untuk membedakan atau mengelompokkan data ke dalam kategori atau kelas yang berbeda berdasarkan atribut-atribut yang ada dalam dataset. Dalam konteks pengembangan sistem untuk mengevaluasi kualitas air, algoritma klasifikasi digunakan untuk memprediksi apakah air tersebut aman atau tidak berdasarkan nilai-nilai parameter yang diukur.

1. *Decision Tree*

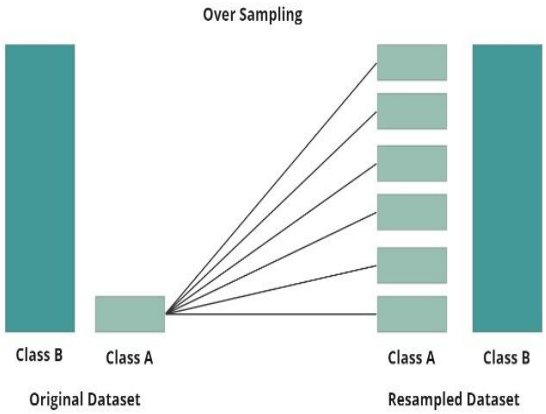
Algoritma *Decision Tree* banyak digunakan, terutama di bidang klasifikasi dan pemodelan regresi. Struktur *Decision Tree* meliputi simpul akar, simpul keputusan, dan simpul daun[5], seperti yang ditunjukkan pada Gambar 1. Simpul akar berfungsi sebagai titik awal algoritma, menandai dimulainya proses pengambilan keputusan. Simpul keputusan digunakan untuk membagi pohon keputusan menjadi beberapa cabang. Sementara itu, simpul daun adalah titik akhir di mana keputusan akhir dibuat[6]. Keuntungan dari pohon keputusan adalah kesederhanaan dan kemudahan dalam memahami model, karena disajikan dalam bentuk pohon dengan cabang-cabang, yang membuat interpretasi lebih mudah[7].



GAMBAR 1
Arsitektur Decision Tree

C. Teknik *Oversampling* SMOTE

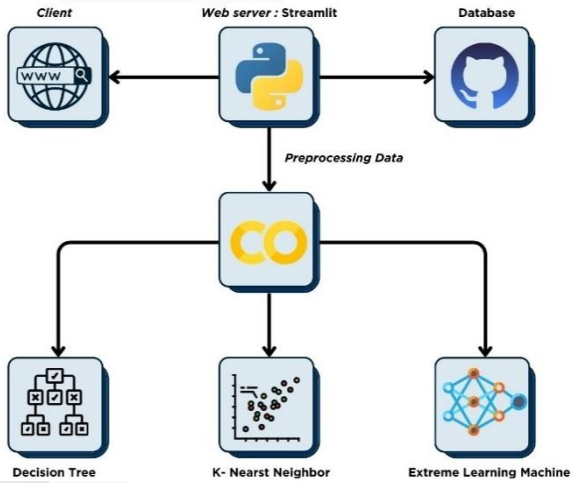
Synthetic Minority Over-sampling Technique (SMOTE) adalah teknik yang dapat digunakan untuk menghasilkan data tambahan pada kelas minoritas dalam kumpulan data perilaku. Konsep SMOTE melibatkan pembuatan salinan atau replika dari kelas minoritas agar mirip dengan kelas mayoritas. Replika ini disebut sebagai data sintetis[8]. SMOTE adalah metode *oversampling* yang awalnya dirancang untuk mengatasi masalah ketidakseimbangan dalam dataset[9], di mana jumlah sampel di kelas minoritas sangat kecil dibandingkan dengan kelas mayoritas seperti yang diilustrasikan pada Gambar 2.



GAMBAR 2
Ilustrasi Penerapan SMOTE

D. Aplikasi Berbasis Web

Aplikasi berbasis web merupakan jenis aplikasi yang dapat diakses dengan mudah melalui *browser*. Dalam hal penggunaannya, aplikasi ini tidak memerlukan banyak sumber daya baik dari segi perangkat lunak maupun perangkat keras. Pengguna hanya perlu memiliki koneksi internet dan *browser* untuk mengaksesnya, tanpa harus mengunduh aplikasi tambahan. Dalam proses pembuatannya, aplikasi web ini menggunakan *framework* Streamlit, sebuah *platform open source* yang berbasis Python dan memiliki fokus dalam pengembangan aplikasi web dibidang data sains dan *machine learning*.



GAMBAR 3
Arsitektur Aplikasi

Gambar 3 dalam Pengembangan Sistem Klasifikasi Kualitas Air Minum Berbasis Web Menggunakan Algoritma *Machine Learning* memberikan gambaran tentang arsitektur aplikasi yang akan dibuat. Website ini akan dirancang menggunakan *framework open source,* yaitu Streamlit, dengan Bahasa Python sebagai bahasa utamanya. Streamlit dapat dengan mudah berintegrasi dengan berbagai *library* yang tersedia dalam bahasa Python, membuatnya menjadi pilihan yang fleksibel untuk pengembangan aplikasi web. Untuk melakukan klasifikasi kualitas air minum, akan digunakan algoritma machine learning yaitu *Decision Tree*. Penggunaan algoritma-algoritma ini diharapkan dapat

memberikan hasil klasifikasi yang akurat dan beragam untuk mengidentifikasi kualitas air minum dengan lebih baik.

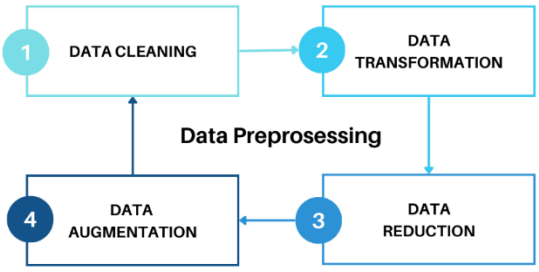
III. METODE

A. *Dataset*

Penelitian ini menggunakan *dataset* berisi parameter kualitas air yang diperoleh dari 24 tempat, termasuk PDAM dan perusahaan air minum. Data tersebut mencakup informasi tentang berbagai parameter kualitas air seperti *E.Colli*, *Coliform*, Arsen, Kromium, Kadmium, Nitrit, Nitrat, Sianida, Selenium, Alumunium, Besi, Kesadahan, Klorida, Mangan, pH, Seng, Sulfat, Tembaga, Amonia, *Chlor*, BOD5, COD, Bau, Rasa, Warna, TDS, Kekeruhan, Suhu, dan Potabilitas. Total terdapat 4.198 data mengenai potabilitas air dalam dataset ini, di mana hanya 961 data yang dikategorikan sebagai air tidak layak minum. Persentase air tidak layak minum hanya sebesar 23% dari total data potabilitas air. Dalam *dataset* ini, terdapat fitur khusus yang berfungsi sebagai variabel target, yaitu menandakan tingkat kelayakan pada air minum. Data tersebut diberi label potabilitas sebagai air layak minum (1) atau air tidak layak minum (0).

B. Preprosessing Data

Pada penelitian ini, dilakukan preprosessing data melalui beberapa tahapan penting seperti yang divisualisasikan pada Gambar 4. Tahap pertama adalah *data cleaning*, di mana langkah-langkah dilakukan untuk mengatasi *missing value* dengan mengisi nilai yang hilang menggunakan median. Selanjutnya, dilakukan *data transformation* dengan mengubah tipe data pada kolom yang relevan sesuai kebutuhan analisis. *Label encoding* juga diterapkan untuk mengubah kolom Rasa menjadi angka, dengan nilai "Berasa" direpresentasikan oleh angka 1 dan nilai "Tidak Berasa" direpresentasikan oleh angka 0. Hal serupa dilakukan untuk kolom Bau, dengan "Berbau" direpresentasikan oleh angka 0 dan "Tidak Berbau" direpresentasikan oleh angka 1.



GAMBAR 4
Tahapan Preprosessing Data

Tahap berikutnya adalah *data reduction,* di mana kolom-kolom yang memiliki korelasi rendah dengan kolom lainnya dihapus. Tujuannya adalah untuk memilih fitur yang memiliki kontribusi yang signifikan terhadap variabel target. Terakhir, dilakukan data augmentation menggunakan teknik *oversampling*, khususnya metode SMOTE. SMOTE (*Synthetic Minority Over-sampling Technique*) merupakan metode yang populer diterapkan dalam rangka menangani ketidakseimbangan kelas. Teknik ini menyintesis sampel baru dari kelas minoritas untuk menyeimbangkan data dalam kelas target dengan cara sampling ulang sampel kelas minoritas[15]. Setelah semua data berhasil diproses, lalu dilakukan pemisahan data menjadi data training dan data testing. Pembagian data training dan data testing dilakukan menggunakan pemrograman Python. Python menyediakan *library* yang dapat mengimplementasikan train atau test split yaitu *library* scikit-learn. Pembagian data yang akan digunakan adalah 80% untuk data *training* dan 20% untuk data *testing*.

C. Kriteria Evaluasi

Pada penelitian ini, dilakukan perbandingan antara algoritma *Decision Tree* dengan penerapan teknik penanganan ketidakseimbangan data SMOTE untuk meningkatkan akurasi. Evaluasi dilakukan dengan menggunakan berbagai metrik seperti akurasi, presisi, recall, dan F1-score, yang digunakan untuk menilai kinerja dan efektivitas masing-masing algoritma dalam mengklasifikasikan kualitas air minum. Selain itu, dilakukan visualisasi *confusion matrix* untuk memberikan gambaran yang lebih jelas tentang performa model, yang menampilkan empat output penting seperti yang terlihat pada Tabel 1. Metrik-metrik seperti sensitivitas, spesifisitas, akurasi, dan tingkat kesalahan dapat dihitung dari *confusion matrix* ini.

TABEL 1
Confusion Matrix untuk Evaluasi Model

TP (*True Positif Rate*)	FP (*False Positif Rate*)
FN (*False Negative Rate*)	TN (*True Negative Rate*)

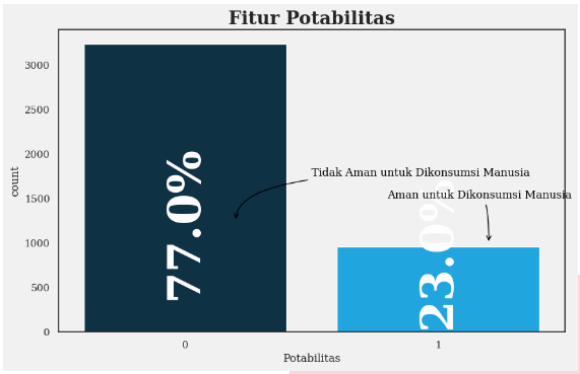
Confusion matrix digunakan untuk mengevaluasi kinerja sistem dalam klasifikasi kelayakan air minum. Dalam konteks ini, *true positive* menunjukkan jumlah air minum yang benar-benar "layak" dan berhasil diidentifikasi dengan tepat oleh sistem sebagai "layak". *True negative* menandakan jumlah air minum yang sebenarnya "tidak layak" dan berhasil diidentifikasi dengan benar oleh sistem sebagai "tidak layak". Di sisi lain, *false positive* menggambarkan air minum yang sebenarnya "tidak layak", tetapi disalahklasifikasikan oleh sistem sebagai "layak". Sedangkan, *false negative* mencerminkan air minum yang sebenarnya "layak", namun disalahklasifikasikan oleh sistem sebagai "tidak layak". Dengan *confusion matrix* ini, akan dievaluasi sejauh mana sistem dapat mengklasifikasikan kelayakan air minum secara akurat dan mengidentifikasi potensi kesalahan yang mungkin terjadi.

Selain itu, kinerja model dievaluasi menggunakan kurva ROC, yang menggambarkan hubungan antara tingkat positif benar dan tingkat positif salah pada berbagai ambang klasifikasi[10]. Nilai AUC, yang mewakili luas area di bawah kurva ROC, menjadi ukuran kapasitas model dalam membedakan antara kelas positif dan negatif. Nilai AUC yang lebih tinggi menandakan peningkatan kinerja klasifikasi, menunjukkan kemampuan model dalam mengidentifikasi dengan akurat antara kelas-kelas yang diprediksi[11].

IV. HASIL DAN PEMBAHASAN

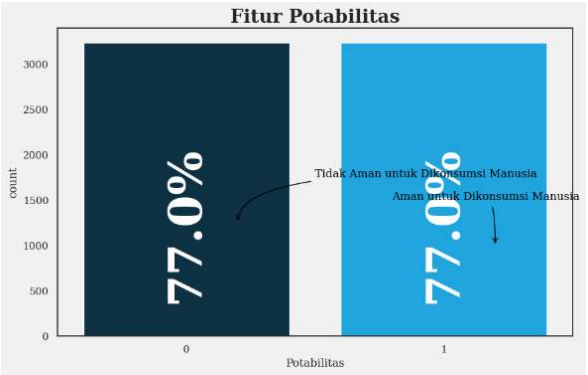
Setelah melewati tahap pra-pemrosesan data, diketahui bahwa data tidak mengandung nilai Null, NA, NaN, atau *missing value*, sehingga dianggap bersih. Seperti yang disebutkan sebelumnya, *dataset* kelayakan air minum memiliki ketidakseimbangan pada kelas target. Untuk

mengatasi hal ini, teknik SMOTE dipilih sebagai solusi. Sebelum *oversampling*, terdapat sekitar 961 data air tidak layak minum yang hanya menyumbang sekitar 23% dari total transaksi, sementara sisanya 3.237 data (77%) merupakan data dengan kategori air layak minum. Hal ini dapat dilihat pada Gambar 5 di bawah ini.



GAMBAR 5
Distribusi Data Sebelum SMOTE

Namun, setelah dilakukan *oversampling*, jumlah data pada kolom target menjadi seimbang dengan data air minum layak 3.237 dan data air minum tidak layak 3.237 seperti yang terlihat pada Gambar 6.



GAMBAR 6
Distribusi Data Setelah SMOTE

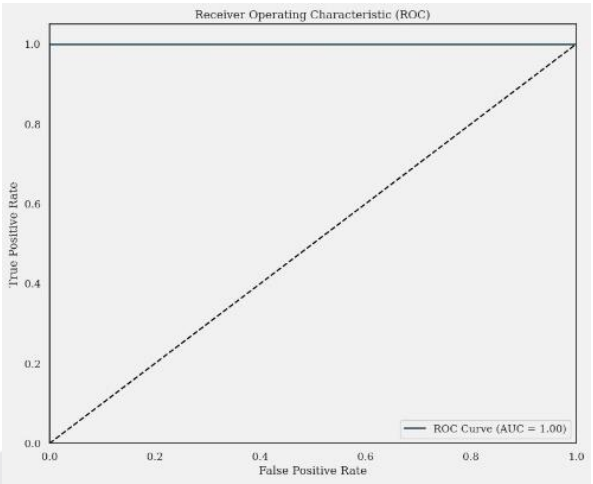
Setelah melakukan pemodelan menggunakan algoritma Decision Tree dan menerapkan teknik *oversampling* SMOTE untuk menangani ketidakseimbangan data, kinerja dari setiap eksperimen dievaluasi dengan menggunakan berbagai metrik, termasuk akurasi *training,* akurasi *testing*, waktu *training,* waktu *testing*, dan ROC *Curve* (AUC *Score*). Hasil evaluasi ini disajikan dalam Tabel 2 berikut ini.

TABEL 2
Evaluasi Kinerja Model Klasifikasi dengan Penggunaan SMOTE dan Algoritma Decision Tree

max_depth	Akurasi Training	Akurasi Testing	Waktu Training	Waktu Testing	Nilai AUC
1	0.990	0.993	0.02026s	0.00226s	0.99
2	0.996	0.997	0.03325s	0.00247s	1.00
3	0.998	0.999	0.03456s	0.00272s	1.00
4	**0.999**	**1.0**	**0.03570s**	**0.00223s**	**1.00**
5	0.999	0.999	0.05908s	0.00223s	1.00

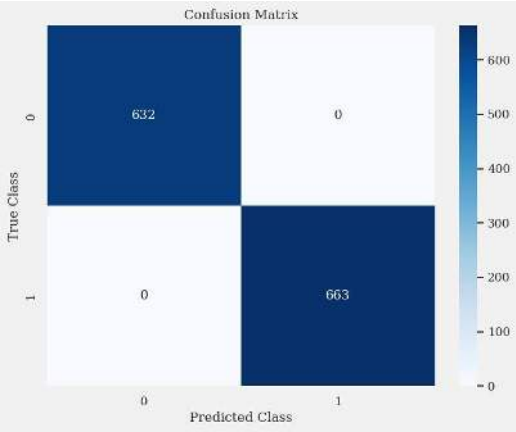
Dari hasil evaluasi di atas, dapat dilihat bahwa peningkatan kedalaman maksimum (*max_depth*) dari Decision Tree memberikan pengaruh positif terhadap akurasi training dan akurasi testing. Semakin tinggi nilai *max_depth*, semakin tinggi juga akurasi yang dihasilkan baik pada data *training* maupun data *testing*. Selain itu, waktu yang dibutuhkan untuk proses *training* dan *testing* juga meningkat seiring dengan peningkatan *max_depth.* Namun, perlu diingat bahwa peningkatan kedalaman maksimum hingga nilai 5 telah memberikan akurasi yang tinggi dan stabil (hampir 100%) tanpa mengorbankan waktu yang signifikan. Selain itu, nilai AUC (*Area Under the ROC Curve*) juga menunjukkan bahwa model Decision Tree yang dibangun memiliki kemampuan yang sangat baik dalam membedakan antara kelas positif dan negatif. Semua nilai AUC mendekati 1, yang menunjukkan bahwa model memiliki tingkat keberhasilan yang tinggi dalam mengklasifikasikan kualitas air minum.

Hasil evaluasi menunjukkan bahwa model Decision Tree dengan *max_depth* sebesar 4 menghasilkan performa yang paling optimal. Pada *max_depth* ini, model mencapai akurasi training sebesar 99.9% dan akurasi *testing* mencapai 100%. Waktu yang dibutuhkan untuk proses training adalah 0.03570 detik, sedangkan waktu *testing* adalah 0.00223 detik. Hasil evaluasi lainnya juga menunjukkan nilai AUC sebesar 1.00 yang divisualisasikan pada Gambar 7, menunjukkan kemampuan model yang sangat baik dalam membedakan antara kelas positif dan negatif.



GAMBAR 7
ROC Curve (AUC Score) max_depth = 4

Selain itu, hasil evaluasi juga menunjukkan bahwa model memiliki nilai *True Positive* (TP) sebesar 632 dan *True Negative* (TN) sebesar 663. Model tidak menghasilkan *False Positive* (FP) atau *False Negative* (FN), yang berarti tidak ada data yang salah diklasifikasikan sebagai positif atau negatif terlihat pada Gambar 8.



GAMBAR 8
Confusion Matrix max_depth = 4

Hasil evaluasi model juga dievaluasi menggunakan *classification report*, yang memberikan informasi rinci tentang berbagai metrik evaluasi, termasuk presisi, recall, dan f1-score, seperti yang ditampilkan dalam Gambar 9 di bawah ini.

	precision	recall	f1-score	support
0	1.00	1.00	1.00	632
1	1.00	1.00	1.00	663
accuracy			1.00	1295
macro avg	1.00	1.00	1.00	1295
weighted avg	1.00	1.00	1.00	1295

GAMBAR 9
Classification Report max_depth = 4

Berdasarkan *classification report* tersebut, dapat dilihat bahwa model memiliki presisi (precision) dan recall sebesar 1.00 untuk kelas "Air Layak Minum" dan "Air Tidak Layak Minum". Nilai f1-score juga sebesar 1.00 untuk kedua kelas, menunjukkan bahwa model memiliki performa yang sangat baik dalam mengklasifikasikan data positif dan negatif.

Setelah model berhasil dibuat dan diperoleh hasil yang optimal, kemudian model tersebut diimplementasikan ke dalam *website*. Pada *website* tersebut, terdapat tiga fitur utama, yaitu Beranda, Informasi, dan Credits, yang divisualisasikan pada Gambar 10 di bawah ini.

GAMBAR 10
Tampilan Fitur Beranda

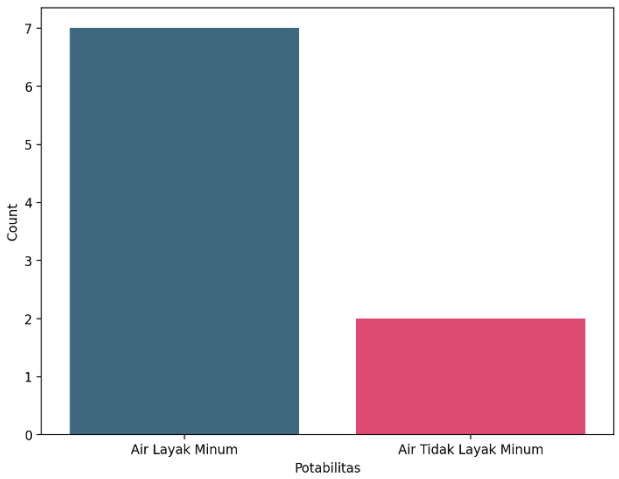
Fitur Beranda menjadi fitur pokok yang memungkinkan pengguna untuk melakukan klasifikasi kualitas air. Pengguna dapat memilih untuk melakukan klasifikasi secara manual dengan memasukkan data sendiri, atau menggunakan fitur unggah *file* untuk memproses data dari *file* yang diunggah. Hasil klasifikasi manual akan ditampilkan dalam bentuk visualisasi seperti yang terlihat pada Gambar 8. Sementara hasil klasifikasi menggunakan unggah *file* akan ditampilkan seperti pada Gambar 11.

GAMBAR 11
Tampilan Hasil Klasifikasi Metode Manual

Sementara hasil klasifikasi menggunakan unggah file akan ditampilkan seperti pada Gambar 12 di bawah ini. Selain itu hasil proporsi antara air minum layak dan air minum tidak layak pun akan ditampilkan seperti yang terlihat pada Gambar 13.

	E.Colli	Coliform	Arsen	Kromium	Kadmium	Nitrit	Nitrat	Sianida	Selenium	Alumunium
0	0	0	0.0082	0.0338	0.0027	1.341	26.1831	0.0291	0.0044	0.118
1	7.3	7.2673	0.9112	0.8501	0.7751	2.889	21.779	0.7373	0.0148	0.1719
2	0	0	0.0072	0.0238	0.0025	1.351	25.1831	0.0293	0.0054	0.128
3	8.0	7.3673	0.9212	0.8401	0.7651	2.789	21.679	0.7373	0.0248	0.1719
4	0	0	0.0072	0.0138	0.0024	1.351	25.1731	0.0293	0.0044	0.128
5	0	0	0.0072	0.0128	0.0024	1.341	25.1731	0.0273	0.0044	0.128
6	0	0	0.0074	0.0148	0.0024	1.341	24.1731	0.0293	0.0044	0.128
7	0	0	0.0072	0.0348	0.0026	1.341	26.1831	0.0291	0.0044	0.118
8	0	0	0.0082	0.0338	0.0027	1.331	26.1831	0.0281	0.0044	0.128

GAMBAR 12
Hasil Klasifikasi Metode Unggah File



GAMBAR 13
Proporsi Kelas Target

V. KESIMPULAN

Pemodelan menggunakan algoritma *Decision Tree* dengan penanganan ketidakseimbangan data menggunakan teknik SMOTE telah menghasilkan model yang cukup efektif dalam mendeteksi transaksi penipuan dengan akurasi yang tinggi dan waktu eksekusi yang wajar. Peningkatan kedalaman maksimum hingga nilai 5 telah memberikan keseimbangan yang baik antara kinerja model dan waktu eksekusi. Selain itu, berdasarkan hasil Tabel 2 di atas model *Decision Tree* dengan *max_depth* 4 adalah model yang paling optimal dan dapat diandalkan dalam mengklasifikasikan kualitas air minum dengan tingkat akurasi yang sangat tinggi dan tanpa kesalahan dalam klasifikasi.

REFERENSI

[1] R. Shyamala, M. Shanthi, and P. Lalitha, “Physicochemical Analysis of Borewell Water Samples of Telungupalayam Area in Coimbatore District, Tamilnadu, India,” 2008.

[2] M. N. B. Momba, V. K. Malakate, and J. Theron, “Abundance of pathogenic Escherichia coli, Salmonella typhimurium and vibrio cholerae in Nkonkobe drinking water sources,” J Water Health, vol. 4, no. 3, pp. 289–296, Sep. 2006, doi: 10.2166/wh.2006.011.

[3] J. Eshcol, P. Mahapatra, and S. Keshapagu, “Is fecal contamination of drinking water after collection associated with household water handling and hygiene practices? A study of urban slum households in Hyderabad, India,” J Water Health, vol. 7, no. 1, pp. 145–154, 2009, doi: 10.2166/wh.2009.094.

[4] 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

[5] J. Liang, Z. Qin, S. Xiao, L. Ou, and X. Lin, “Efficient and Secure Decision Tree Classification for Cloud-Assisted Online Diagnosis Services,” IEEE Trans Dependable Secure Comput, vol. 18, no. 4, pp. 1632–1644, Jul. 2021, doi: 10.1109/TDSC.2019.2922958.

[6] A. Prayoga Permana, K. Ainiyah, and K. Fahmi Hayati Holle, “Analisis Perbandingan Algoritma Decision Tree, kNN, dan Naive Bayes untuk Prediksi Kesuksesan Start-up,” 2021. [Online]. Available: https://www.kaggle.com/manishkc06/startup-success-prediction.

[7] S. Bagga, A. Goyal, N. Gupta, and A. Goyal, “Credit Card Fraud Detection using Pipeling and Ensemble Learning,” in Procedia Computer Science, Elsevier B.V., 2020, pp. 104–112. doi: 10.1016/j.procs.2020.06.014.

[8] Hairani, N. A. Setiawan, & T. B. Adji. 2013. Metode Klasifikasi Data Mining dan Teknik Sampling Smote.Seminar Nasional Sains dan Teknologi, 168–172.

[9] J. Peng, R. Gao, L. Nguyen, Y. Liang, S. Thng and Z. Lin, "Classification of Non-Tumorous Facial Pigmentation Disorders Using Improved Smote and Transfer Learning," 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 2019, pp. 220-224, doi: 10.1109/ICIP.2019.8802993.

[10] I.H. Witten, E. Frank, and M.A. Hall, “Data Mining Practical Machine Learning Tools and Techniques”, Third Edition, Elsevier Publisher, USA, 2011.

[11] T. Y. Hadiwandra, “SATIN-Sains dan Teknologi Informasi Perbandingan Kinerja Model Klasifikasi Decission Tree , Bayesian Classifier, Instance Base, Linear Function Base, Rule Base pada 4 Dataset Berbeda,” vol. 5, no. 1, 2019, [Online]. Available: http://jurnal.stmik-amik-riau.ac.id.