

PENERAPAN ALGORITMA DECISION TREE (CART) UNTUK KLASIFIKASI KONSUMSI DAN KEPUASAN PELANGGAN PDAM TIRTA PAKUAN KOTA BOGOR

Andika Rizky Harahap¹⁾, Gina Purnama Insany²⁾

^{1,2} Program Studi Teknik Informatika, Fakultas Teknik Komputer dan Desain, Universitas Nusa Putra.

Corresponding Author: ¹ andika.rizky_ti22@nusaputra.ac.id

Article Info	ABSTRACT
Article history: Received: Jul 26, 2026 Revised: Jul 29, 2026 Accepted: Oct 11, 2026 Published: Oct 15, 2026	Artikel ini memaparkan penerapan algoritma Decision Tree tipe Classification and Regression Trees (CART) untuk mengelompokkan pelanggan Perumda Tirta Pakuan Kota Bogor berdasarkan volume pemakaian air, sebagai bagian dari kegiatan magang yang berjalan Agustus–Desember 2025. Persoalan yang melatarbelakangi adalah rekapitulasi dan penelaahan data pelanggan yang masih dikerjakan manual sehingga rawan keliru dan menyita waktu. Data historis 474 pelanggan dihimpun, diproses awal, lalu dipilah ke dalam tiga kelompok konsumsi (rendah, sedang, tinggi) memakai algoritma CART yang dijalankan dengan Python. Kelompok sedang mendominasi dengan porsi 53,4%, disusul kelompok rendah 25,5% dan tinggi 21,1%. Saat diuji pada 95 data, model mencapai akurasi 100%, menandakan volume pemakaian dan riwayat tagihan mampu memisahkan kategori pelanggan secara efektif, meskipun ukuran data yang masih terbatas mengharuskan pengujian lanjutan. Hasil ini menunjukkan model dapat membantu manajemen PDAM merencanakan distribusi air, menyusun strategi tarif, dan mendeteksi lebih awal anomali pemakaian.
Keywords: Analisis Data, Decision Tree/CART, Konsumsi Air, Segmentasi Pelanggan, PDAM.	



This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY SA 4.0)

1. PENDAHULUAN

Perkembangan teknologi informasi yang cepat mendorong instansi pelayanan publik, termasuk Perusahaan Daerah Air Minum (PDAM), mempercepat digitalisasi demi efisiensi dan keterbukaan layanan. Salah satu bentuk transformasi itu adalah sistem pemantauan terpadu berbasis Command Center yang memantau kondisi jaringan pipa, tekanan air, dan gangguan operasional secara real-time. Sistem tersebut membantu menjaga mutu air bersih, mempercepat penanganan keluhan, serta menekan kehilangan air non-pendapatan (non-revenue water).

Perumda Tirta Pakuan Kota Bogor, yang telah melayani sejak masa kolonial dan berstatus resmi PDAM sejak 1977, kerap menjadi subjek penelitian mengenai mutu layanan dan kepuasan pelanggannya [1]. Riset-riset sebelumnya umumnya mengandalkan survei dan analisis regresi untuk menilai kepuasan terhadap kualitas produk maupun distribusi [2], [5], [9], sementara faktor kompetensi SDM dan tata kelola turut disorot pengaruhnya terhadap kinerja layanan [4]. Secara umum, temuan tersebut sepakat bahwa kualitas, kuantitas, dan kontinuitas air menjadi penentu utama kepuasan pelanggan PDAM Tirta Pakuan.

Namun demikian, pendekatan yang dipakai kebanyakan penelitian tersebut masih berupa statistik deskriptif dan survei; penggunaan pembelajaran mesin untuk mengelompokkan pola konsumsi pelanggan secara otomatis belum banyak diujicoba. Padahal proses rekapitulasi dan penelaahan data pelanggan yang masih manual berisiko menyita waktu dan rawan salah, terlebih pada perusahaan berpelanggan banyak dengan ragam kategori (domestik, komersial, industri).

Menyikapi hal tersebut, kegiatan magang di Divisi Command Center dan Ruang Rapat PDAM Tirta Pakuan sepanjang Agustus–Desember 2025 difokuskan pada penerapan algoritma Decision Tree jenis Classification and Regression Trees (CART) untuk mengelompokkan pelanggan menurut tingkat konsumsi air. Algoritma ini dipilih sebab dapat melakukan klasifikasi kategori sekaligus memprediksi nilai numerik dalam satu proses, serta menghasilkan aturan keputusan yang relatif mudah dipahami pihak manajemen non-teknis.

Artikel ini disusun untuk mendokumentasikan penerapan model CART dalam mensegmentasi pelanggan berdasar volume konsumsi, menilai performanya lewat confusion matrix, serta mengulas dampaknya terhadap keputusan operasional PDAM seperti penentuan tarif, deteksi anomali, dan perencanaan distribusi. Pembahasan disusun

mengikuti kerangka IMRaD, mulai dari landasan teori, metode, hasil dan pembahasan, hingga simpulan.

Decision tree adalah salah satu metode klasifikasi dalam pembelajaran mesin terbimbing (supervised learning) yang menyatakan aturan keputusan dalam bentuk struktur pohon [14]. Struktur tersebut terdiri atas root node sebagai titik pemilahan pertama, sejumlah decision node (internal node) yang menguji suatu atribut tertentu, serta leaf node yang menyatakan hasil akhir klasifikasi. Pohon dibangun dengan memilah data secara berulang berdasarkan atribut yang paling efektif memisahkan kelas, sampai setiap cabang berisi data yang relatif homogen [14]. Beberapa algoritma decision tree yang banyak dipakai antara lain ID3 dan C4.5 yang dikembangkan oleh Quinlan dengan kriteria pemilahan berbasis entropy dan information gain [15], serta CART yang dikembangkan oleh Breiman dkk. dengan kriteria Gini impurity [13]. Kelebihan utama decision tree terletak pada aturan keputusan yang mudah dipahami, serta kemampuannya mengolah data kategorikal maupun numerik tanpa memerlukan banyak praproses atau normalisasi [12]. Kekurangannya, pohon yang dibiarkan tumbuh terlalu dalam cenderung overfitting terhadap data latih, sehingga umumnya diperlukan pembatasan kedalaman atau pemangkasan (pruning) [14].

Classification and Regression Trees (CART) merupakan algoritma decision tree yang diperkenalkan oleh Breiman, Friedman, Olshen, dan Stone pada 1984 [13]. Berbeda dari ID3 maupun C4.5 yang dapat menghasilkan pemilahan lebih dari dua cabang pada satu node, CART selalu membentuk pohon biner sehingga setiap decision node hanya memiliki dua cabang turunan [13]. Untuk kasus klasifikasi, pemilahan terbaik pada tiap node ditentukan dengan meminimalkan nilai Gini impurity pada node anak. Gini impurity suatu node t dirumuskan sebagai berikut.

$$\text{Gini}(t) = 1 - \sum p_i^2 \quad (1)$$

dengan p_i menyatakan proporsi data kelas ke- i pada node t dan c adalah banyaknya kelas yang ada; nilai Gini yang semakin kecil menunjukkan komposisi kelas pada node tersebut semakin homogen [13]. Pemilahan diulang secara rekursif pada tiap cabang hingga node menjadi murni ($\text{Gini} = 0$) atau memenuhi kriteria penghentian lain, dan proses ini dapat dilanjutkan dengan pemangkasan berbasis cost-complexity untuk mengurangi risiko overfitting [13]. Karena mampu menangani target kategorikal (classification tree) maupun target numerik (regression tree) dalam satu kerangka algoritma, CART telah diterapkan pada berbagai konteks klasifikasi pelanggan, misalnya klasifikasi status keaktifan nasabah perbankan [12] dan klasifikasi pola konsumsi air pelanggan [16]. Karakteristik inilah yang mendasari pemilihan CART pada penelitian ini untuk mengelompokkan pelanggan PDAM Tirta Pakuan berdasarkan volume konsumsi air bulanan.

2. METODE

2.1 Waktu dan Tempat Pelaksanaan

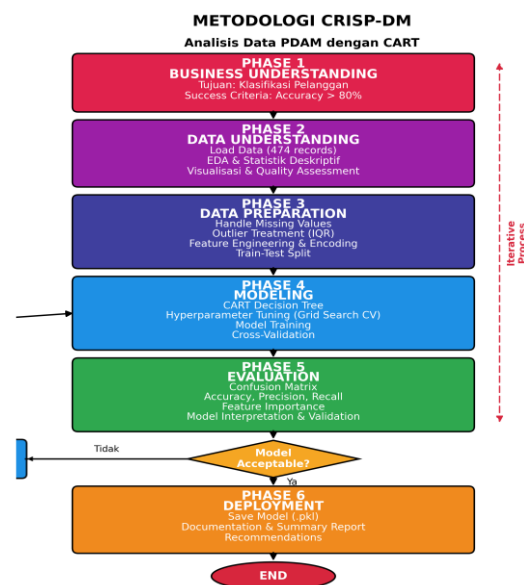
Pengumpulan dan analisis data berlangsung selama magang di Perumda Tirta Pakuan Kota Bogor, khususnya di Divisi Command Center dan Ruang Rapat, dari 1 Agustus sampai 25 Desember 2025. Kegiatan dijalankan lima hari kerja per pekan pukul 07.30–16.30 WIB secara Work From Office.

2.2 Tahapan Penelitian

Penelitian ini menempuh empat tahapan pokok seperti dirangkum Tabel 1, yaitu pengumpulan data, praproses data, klasifikasi CART, dan evaluasi model. Alur keseluruhan proses digambarkan pada Gambar 1.

Tabel 1. Tahapan Analisis Data

No	Tahapan	Deskripsi
1	Pengumpulan data	Menghimpun identitas, kategori, riwayat pemakaian, dan riwayat tagihan dari 474 pelanggan melalui sistem billing dan basis data PDAM.
2	Praproses data	Membersihkan data, menangani nilai kosong, dan menyeragamkan format menggunakan Python (NumPy, Pandas).
3	Klasifikasi CART	Membangun model Decision Tree (CART) untuk mengelompokkan pelanggan ke tiga kategori konsumsi berdasarkan ambang 59,5 m3.
4	Evaluasi model	Mengukur kinerja model memakai confusion matrix dan accuracy score pada 95 data uji.



Gambar 1. Diagram alur tahapan analisis data konsumsi air dengan metode CART

2.3 Pengumpulan Data

Data bersumber dari sistem billing dan basis data pelanggan PDAM Tirta Pakuan, meliputi identitas, alamat, kategori pelanggan (domestik, komersial, industri), riwayat pembacaan meter bulanan, riwayat tagihan, serta status pembayaran. Total data yang dipakai berjumlah 474 pelanggan.

2.4 Praproses Data dan Perangkat Analisis

Praproses mencakup pembersihan data, penanganan nilai hilang, dan penyeragaman format dari berbagai sumber. Seluruh pengolahan memakai bahasa Python dengan pustaka NumPy untuk komputasi numerik serta Matplotlib dan Seaborn untuk visualisasi. Python dipilih karena sintaksnya sederhana dan ekosistem pustakanya lengkap untuk seluruh tahapan analisis, dari pengumpulan hingga visualisasi data [10], [11].

2.5 Penerapan CART pada Data Konsumsi Pelanggan

CART adalah varian Decision Tree yang dapat melakukan klasifikasi kategori sekaligus memprediksi nilai numerik dalam satu model. Pemodelan dimulai dari root node yang memilah keseluruhan dataset berdasarkan ambang konsumsi air 59,5 m3 per bulan. Gini impurity root node sebesar 0,506 menunjukkan data pada node itu masih tercampur antar tiga kategori, sehingga proses splitting dilanjutkan hingga diperoleh node yang lebih homogen. Struktur pohon yang terbentuk terdiri atas root node sebagai titik pemilahan serta leaf node yang mewakili hasil prediksi tiap cabang.

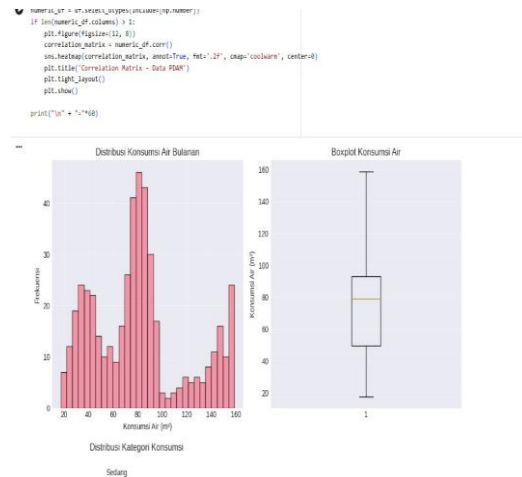
2.6 Evaluasi Model

Performa model dinilai memakai confusion matrix dan accuracy score guna mengukur ketepatan klasifikasi terhadap tiga kategori konsumsi (rendah, sedang, tinggi) pada data uji.

3. HASIL DAN PEMBAHASAN

3.1 Gambaran Umum Data Konsumsi Air

Visualisasi awal data konsumsi bulanan pelanggan (Gambar 2) memperlihatkan pola pemakaian yang beragam antar kategori. Contoh data pelanggan kawasan industri (Gambar 3) menunjukkan volume konsumsi yang cenderung tinggi dan stabil dibanding kategori domestik, sesuai karakteristik kebutuhan air yang besar di sektor tersebut.



Gambar 2. Visualisasi data konsumsi air pelanggan secara bulanan

```
RangeIndex: 474 entries, 0 to 473
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype  
---  -
 0   Nama        474 non-null    object  
 1   Tarif       474 non-null    object  
 2   Pakai       474 non-null    float64  
 3   Total       473 non-null    float64  
 4   Desa        473 non-null    object  
 5   Kecamatan   473 non-null    object  
 6   Sumber      473 non-null    object  
dtypes: float64(2), object(5)
memory usage: 26.1+ KB
None
```

Statistik Deskriptif:

	Pakai	Total
count	474.000000	4.730000e+02
mean	82.489451	3.000403e+05
std	41.748428	2.463034e+05
min	17.500000	4.375000e+04
25%	49.475000	1.242500e+05
50%	78.900000	2.541000e+05
75%	93.175000	3.125500e+05
max	223.100000	1.338600e+06

Sample Data (5 baris pertama):

	Nama	Tarif	Pakai	Total	Desa \
0	Pelanggan_1	R1	42.5	106250.0	Kelurahan Empang
1	Pelanggan_2	R2	68.3	204900.0	Kelurahan Bondongan
2	Pelanggan_3	R3	95.0	332500.0	Kelurahan Semplak
3	Pelanggan_4	Niaga	120.6	542700.0	Kelurahan Balumbang Jaya
4	Pelanggan_5	R1	28.4	71000.0	Kelurahan Baranangsiang

	Kecamatan	Sumber
0	Bogor Selatan	PDAM Tirta Pakuan
1	Bogor Selatan	PDAM Tirta Pakuan
2	Bogor Barat	PDAM Tirta Pakuan
3	Bogor Barat	PDAM Tirta Pakuan
4	Bogor Tinur	PDAM Tirta Pakuan

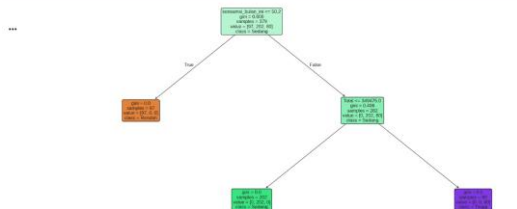
Missing Values:

Nama	0
Tarif	0
Pakai	0
Total	1

Gambar 3. Contoh data pelanggan pada kawasan industry

3.2 Klasifikasi Berdasarkan Ambang Batas Konsumsi

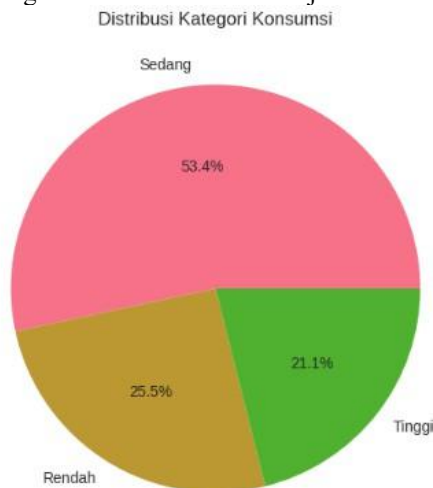
Model CART berhasil memilah 474 data pelanggan berdasarkan ambang konsumsi 59,5 m³ (Gambar 4). Pelanggan berkonsumsi $\leq 59,5$ m³ diarahkan ke cabang kiri (True), sedangkan yang di atasnya diarahkan ke cabang kanan (False). Splitting selanjutnya menghasilkan node dengan nilai Gini yang kian mengecil, menandakan homogenitas data tiap cabang terus meningkat sampai leaf node akhir.



Gambar 4. Struktur pohon keputusan (decision tree) hasil klasifikasi berdasarkan ambang batas konsumsi air

3.3 Distribusi Kategori Konsumsi Pelanggan

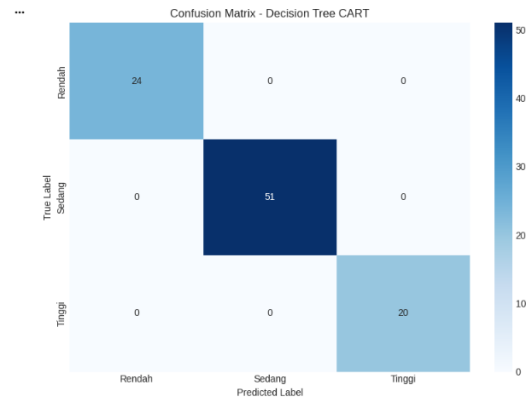
Hasil klasifikasi (Gambar 5) menunjukkan kategori Sedang mendominasi dengan 53,4%, diikuti Rendah 25,5% dan Tinggi 21,1%. Kategori Sedang lazimnya mewakili rumah tangga umum, apartemen, atau usaha kecil berpola konsumsi normal. Kategori Rendah bisa berasal dari rumah kontrakan atau hunian bertenang tunggal sehingga perlu dipantau lanjut untuk memastikan tidak ada meter rusak atau kebocoran tersembunyi. Kategori Tinggi umumnya berasal dari pelanggan berpenghuni banyak atau usaha komersial menengah-besar seperti hotel, restoran, atau industri perumahan yang membutuhkan air dalam jumlah besar.



Gambar 5. Distribusi kategori konsumsi pelanggan PDAM Tirta Pakuan Kota Bogor

3.4 Evaluasi Kinerja Model

Pengujian pada 95 data, terdiri atas 24 data kategori Rendah, 51 data kategori Sedang, dan 20 data kategori Tinggi, menghasilkan confusion matrix pada Gambar 6. Seluruh data uji terklasifikasi tepat tanpa kesalahan sehingga model mencatatkan akurasi 100% pada data tersebut.



Gambar 6. Confusion matrix hasil evaluasi model klasifikasi CART

3.5 Pembahasan

Temuan ini menegaskan bahwa volume konsumsi dan riwayat tagihan memiliki daya pembeda yang kuat terhadap kategori pelanggan PDAM Tirta Pakuan. Akurasi yang sangat tinggi pada data uji perlu disikapi hati-hati sebab ukuran dataset yang relatif kecil (474 data) berpeluang membuat model overfit terhadap pola data latih. Karena itu, validasi lanjutan lewat cross-validation serta pengujian pada dataset yang lebih besar dan rentang waktu berbeda diperlukan sebelum model ini dipakai secara operasional penuh.

Secara praktis, model klasifikasi ini berpotensi menopang manajemen PDAM dalam: (1) perencanaan distribusi dan produksi air berdasarkan segmentasi konsumsi wilayah; (2) penyusunan strategi tarif yang lebih tepat sasaran untuk tiap kategori pelanggan; (3) deteksi dini anomali konsumsi yang mengindikasikan kebocoran, pemakaian ilegal, atau kekeliruan pencatatan meter; serta (4) dukungan pengambilan keputusan bila diintegrasikan dengan Sistem Informasi Geografis (SIG) untuk pemetaan spasial segmen pelanggan berisiko tinggi. Temuan ini memperkuat kajian terdahulu yang menyoroti pentingnya kualitas dan kontinuitas layanan sebagai penentu kepuasan pelanggan [1], [5], [9], sekaligus menambahkan sudut pandang berbasis data yang melengkapi metode survei konvensional.

4. KESIMPULAN

Penerapan algoritma Decision Tree berbasis CART terhadap data konsumsi dan pelanggan PDAM Tirta Pakuan Kota Bogor berhasil mengelompokkan pelanggan ke tiga kategori konsumsi, yakni Rendah (25,5%), Sedang (53,4%), dan Tinggi (21,1%), dengan akurasi 100% pada 95 data uji. Root node model memisah data berdasarkan ambang konsumsi 59,5 m³ per bulan dengan Gini impurity awal 0,506. Meski hasil klasifikasi tergolong sangat baik, keterbatasan ukuran dataset menuntut kehati-hatian dalam menggeneralisasi temuan serta perlunya validasi lanjutan sebelum diterapkan secara penuh.

Model ini berpotensi diaplikasikan manajemen PDAM untuk mendukung perencanaan distribusi air, penyusunan strategi tarif, dan deteksi dini anomali

konsumsi. Penelitian mendatang disarankan memakai dataset yang lebih besar dan rentang waktu lebih panjang, menerapkan teknik cross-validation, membandingkan kinerja CART dengan algoritma lain seperti Random Forest atau Gradient Boosting, serta memadukan hasil klasifikasi dengan Sistem Informasi Geografis untuk mendukung pengambilan keputusan spasial yang lebih menyeluruh.

UCAPAN TERIMA KASIH

Penulis menyampaikan terima kasih kepada Perumda Tirta Pakuan Kota Bogor atas kesempatan pelaksanaan magang, khususnya kepada tim Command Center dan Ruang Rapat atas bimbingan teknis selama kegiatan berlangsung. Terima kasih juga disampaikan kepada Ibu Gina Purnama Insany, S.ST., M.Kom. selaku dosen pembimbing, serta Program Studi Teknik Informatika, Fakultas Teknik Komputer dan Desain, Universitas Nusa Putra atas dukungan yang diberikan.

REFERENSI

- [1] A. A. Nugroho, "Analisis Tingkat Kepuasan Pelanggan Perusahaan Daerah Air Minum (PDAM) Tirta Pakuan Kota Bogor," *Holistic Journal of Management Research*, vol. 3, no. 2, 2015.
- [2] A. S. Sulistiono and Yuniany, "Pengaruh Distribusi dan Kualitas Produk Terhadap Kepuasan Pelanggan: Studi Kasus Produk Air Minum pada PDAM Tirta Pakuan Kota Bogor," *Jurnal Ilmiah Ranggagading (JIR)*, vol. 9, no. 2, 2009.
- [3] "Profil Perusahaan - PERUMDA Tirta Pakuan Kota Bogor," 2024.
- [4] Palahudin, Ismartaya, and J. D. Wetik, "Kompetensi Sumber Daya Manusia dan Iklim Organisasi terhadap Kinerja Pegawai PDAM Tirta Pakuan Kota Bogor," *Management Studies and Entrepreneurship Journal*, vol. 2, no. 1, 2021.
- [5] D. W. Retna and A. Sulistiono, "Pengaruh Kualitas Produk dan Kualitas Pelayanan Terhadap Kepuasan Pelanggan pada PDAM Tirta Pakuan Kota Bogor," *Jurnal Ilmiah Manajemen Kesatuan*, vol. 2, no. 3, 2014, doi: 10.37641/jimkes.v2i3.805.
- [6] W. Pradiani, H. Andriana, and A. Y. Suningrat, "Efektivitas Koagulan Terhadap Penurunan Energi Potensial pada Pengolahan Air di PDAM Tirta Pakuan Kota Bogor," *Integrasi: Jurnal Ilmiah Teknik Industri*, vol. 7, no. 1, 2022, doi: 10.32502/js.v7i1.4464.
- [7] A. Lasmana and N. Nofiyati, "Analisis Kinerja Manajerial Berdasarkan Penerapan Akuntansi Pertanggungjawaban, Partisipasi Anggaran dan Implementasi Strategi pada PDAM Tirta Pakuan Kota Bogor," *Jurnal Akunida*, vol. 3, no. 1, 2017, doi: 10.30997/jakd.v3i1.982.
- [8] D. Armadi, A. Hidayat, and S. M. Simanjuntak, "Analisis Pengelolaan Air Bersih Berkelanjutan di Kota Bogor (Studi Kasus: PDAM Tirta Pakuan)," *Journal of Agriculture, Resource and Environmental Economics*, vol. 2, no. 1, 2019, doi: 10.29244/jaree.v2i1.25928.
- [9] W. S. Aisah, "Analisis Kualitas Pelayanan Terhadap Kepuasan Pelanggan pada PDAM Tirta Pakuan Kota Bogor," *Jurnal Online Mahasiswa (JOM) Bidang Manajemen*, vol. 1, no. 2, 2017.
- [10] W. McKinney, *Python for Data Analysis*, 2nd ed. Sebastopol, CA, USA: O'Reilly Media, 2017.
- [11] J. VanderPlas, *Python Data Science Handbook: Essential Tools for Working with Data*. Sebastopol, CA, USA: O'Reilly Media, 2016.
- [12] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, and S. S. Febrihartina, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024.
- [13] L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone, *Classification and Regression Trees*. Belmont, CA, USA: Wadsworth International Group, 1984.
- [14] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2012.
- [15] J. R. Quinlan, "Induction of decision trees," *Machine Learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [16] S. N. Khofiyah, A. Hamdi, and K. N. Isnaini, "Analisis Pola Konsumsi Air Menggunakan Algoritma Random Forest Classifier pada Distribusi Air Bersih Desa Rempoah Baturraden," *Jurnal Algoritma*, vol. 22, no. 2, pp. 1559–1571, Nov. 2025.