

Penerapan Algoritma C4.5 untuk Klasifikasi Tingkat Kesiapan Digital Masyarakat Pedesaan di Provinsi Papua Barat

Debi Puspita Patulak^{1*}, Christian Dwi Suhendra², Julius P.P. Naibaho³

¹⁻³Program Studi Teknik Informatika, Fakultas Teknik, Universitas Papua, Indonesia

Corresponding Author's e-mail : Debipuspitapatulak2019059@gmail.com



e-ISSN: 2964-2981

ARMADA : Jurnal Penelitian Multidisiplin

<https://ejournal.45mataram.ac.id/index.php/armada>

Vol. 04, No. 07 Juli, 2026

Page: 2329-2341

DOI:

<https://doi.org/10.55681/armada.v4i7.2755>

Article History:

Received: Mei 13, 2026

Revised: Juni 02, 2026

Accepted: Juli 17, 2026

Abstract : *The technology adoption gap in rural areas remains a major challenge to Indonesia's national digital transformation. This study aims to analyse and classify the digital readiness of rural communities in West Papua Province using a computational approach. Quantitative data were collected through an online questionnaire involving 400 respondents and covering demographic characteristics and functional digital competence indicators. The C4.5 decision tree algorithm was used to develop a classification model and identify the main predictor attributes. Model performance was evaluated using 10-Fold Cross Validation. The results show that 53.50% of respondents were classified as having moderate digital readiness. The C4.5 model achieved 88.75% accuracy, 87.92% precision, 88.21% recall, and an F1-Score of 88.06%. Digital literacy was the most influential predictor, with a gain ratio of 0.421, exceeding infrastructure and demographic factors. The C4.5 algorithm proved reliable for objectively mapping community digital readiness. These findings highlight the need to shift policy from infrastructure expansion towards strengthening functional digital literacy and sustainable human resource empowerment.*

Keywords : *C4.5 Algorithm; Digital Readiness; Digital Literacy; Decision Tree*

Abstrak : Kesenjangan adopsi teknologi di wilayah pedesaan masih menjadi tantangan dalam transformasi digital Indonesia. Penelitian ini bertujuan menganalisis dan mengklasifikasikan tingkat kesiapan digital masyarakat pedesaan di Provinsi Papua Barat menggunakan pendekatan komputasional. Data kuantitatif diperoleh melalui kuesioner daring terhadap 400 responden yang mencakup karakteristik demografis dan indikator kompetensi digital fungsional. Algoritma pohon keputusan C4.5 digunakan untuk membentuk model klasifikasi dan mengidentifikasi atribut prediktor utama. Kinerja model diuji menggunakan 10-Fold Cross Validation. Hasil penelitian menunjukkan bahwa 53,50% responden berada pada kategori kesiapan digital sedang. Model C4.5 menghasilkan akurasi 88,75%, presisi 87,92%, recall 88,21%, dan F1-Score 88,06%. Literasi digital menjadi prediktor paling dominan dengan nilai gain ratio 0,421, melampaui faktor infrastruktur dan demografis. Algoritma C4.5 terbukti andal dalam memetakan kesiapan digital masyarakat secara objektif. Temuan ini menegaskan perlunya pergeseran kebijakan dari perluasan infrastruktur menuju penguatan literasi digital fungsional dan pemberdayaan sumber daya manusia secara berkelanjutan.

Kata Kunci : Algoritma C4.5; Kesiapan Digital; Literasi Digital; Pohon Keputusan

PENDAHULUAN

Transformasi digital telah menjadi pilar fundamental dalam pembangunan di berbagai sektor, mengubah paradigma operasional pemerintahan, pendidikan, layanan kesehatan, dan aktivitas ekonomi masyarakat. Adaptasi terhadap ekosistem digital ini tidak lagi dipandang sebagai pilihan, melainkan sebuah keharusan untuk mempertahankan daya saing dan produktivitas (World Economic Forum, 2024). Keberhasilan integrasi teknologi dalam masyarakat bergantung pada tingkat kesiapan digital (*digital readiness*), yang merepresentasikan kemampuan individu dalam mengakses, memahami, dan memanfaatkan teknologi informasi secara efektif, kritis, dan bertanggung jawab (United Nations Educational, Scientific and Cultural Organization, 2023). Kondisi ideal yang diharapkan dari transformasi ini adalah pemerataan akses dan kapabilitas teknologi bagi seluruh lapisan masyarakat tanpa memandang batasan geografis.

Meskipun pembangunan infrastruktur teknologi informasi dan komunikasi di Indonesia terus mengalami peningkatan, realitas empiris menunjukkan masih terdapat kesenjangan yang signifikan antara kondisi ideal dengan situasi di lapangan. Fenomena kesenjangan digital (*digital divide*) tetap menjadi tantangan krusial, khususnya bagi wilayah pedesaan di daerah tertinggal atau kawasan timur Indonesia. Provinsi Papua Barat, dengan karakteristik topografi yang kompleks dan sebaran penduduk yang tidak merata, menghadapi tantangan berat dalam mewujudkan ekuivalitas adopsi teknologi. Meskipun kepemilikan perangkat cerdas mulai meningkat, kemampuan literasi digital, pemahaman keamanan siber, serta pemanfaatan teknologi untuk aktivitas produktif pada masyarakat pedesaan masih berada pada tahap yang bervariasi (Badan Pusat Statistik Provinsi Papua Barat, 2024; World Bank, 2024). Keberadaan infrastruktur fisik terbukti tidak serta-merta berbanding lurus dengan tingkat kesiapan digital masyarakat apabila tidak diimbangi dengan analisis kapabilitas pengguna (International Telecommunication Union, 2023).

Dalam konteks penyusunan kebijakan, evaluasi terhadap kesiapan digital masyarakat pedesaan selama ini masih didominasi oleh pendekatan statistik deskriptif konvensional. Pendekatan tersebut umumnya hanya memberikan gambaran distribusi data makro tanpa memiliki kemampuan analitik yang mendalam untuk memetakan interaksi antarvariabel yang kompleks. Kelemahan identifikasi deskriptif ini berisiko menghasilkan rekomendasi kebijakan yang kurang presisi (Hidayat, Nugroho, & Prasetyo, 2024). Untuk mengatasi keterbatasan tersebut, pendekatan *data mining* dan *machine learning* menawarkan solusi analitik prediktif yang esensial. Klasifikasi berbasis *data mining* mampu memproses volume data yang besar guna mengekstraksi pola tersembunyi (*hidden patterns*) serta menghasilkan model prediksi berbasis probabilitas maupun aturan keputusan logis (Han, Pei, & Tong, 2022).

Sejumlah penelitian terdahulu mengeksplorasi pemanfaatan *machine learning* untuk pengukuran kapabilitas digital masyarakat. Ahmed et.al., (2024) serta Utami et.al., (2024) mengimplementasikan berbagai teknik *machine learning* untuk menilai kesiapan digital, membuktikan bahwa algoritma klasifikasi secara signifikan mampu memprediksi level adaptasi teknologi masyarakat. Secara spesifik, Ismail et.al., (2022) menunjukkan efikasi algoritma C4.5 dalam studi klasifikasi sosio-ekonomi. Penelitian lain oleh Yusuf et.al., (2023) menerapkan pendekatan *data mining* untuk mengevaluasi inklusi digital di wilayah pedesaan dengan hasil klasifikasi yang cukup relevan. Lebih lanjut, studi komparatif oleh Alharbi et.al., (2023) menyoroti bahwa metode klasifikasi berbasis *decision tree*, khususnya perluasan seperti C4.5, memiliki keunggulan kompetitif dibandingkan algoritma *black-box* seperti *Neural Network*. Hal ini didukung oleh temuan Kumar et al. (2023) yang menegaskan bahwa meskipun algoritma ansambel (seperti *Random Forest*) memberikan margin akurasi yang sedikit lebih tinggi, model C4.5 menawarkan interpretabilitas yang jauh lebih signifikan, sebuah atribut yang vital dalam perumusan kebijakan publik.

Berdasarkan tinjauan literatur tersebut, teridentifikasi sebuah kesenjangan penelitian yang mendasar. Mayoritas studi klasifikasi kesiapan atau literasi digital terfokus pada demografi perkotaan, mahasiswa, maupun populasi umum tingkat nasional. Implementasi model analitik

yang ditujukan secara eksklusif untuk masyarakat pedesaan, khususnya di area dengan tantangan geografis seperti Papua Barat, belum dieksplorasi secara optimal. Selain itu, belum ada model klasifikasi *decision tree* yang secara spesifik dirancang untuk mengekstraksi faktor paling dominan seperti literasi digital terhadap akses infrastruktur sebagai penentu utama kategori kesiapan digital pedesaan pada kawasan tersebut. Nilai kebaruan dari penelitian ini terletak pada integrasi data primer yang representatif dengan algoritma C4.5 untuk menghasilkan *classification rules* yang transparan dan spesifik terhadap karakteristik demografis maupun teknologis kawasan Papua Barat.

Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma klasifikasi C4.5 dalam mengevaluasi dan mengkategorikan tingkat kesiapan digital masyarakat pedesaan di Provinsi Papua Barat. Melalui representasi *decision tree*, penelitian ini ditargetkan mampu mengidentifikasi atribut prediktor dominan yang paling berkontribusi terhadap status kesiapan digital masyarakat. Temuan eksperimental dari model prediksi ini diharapkan dapat diinterpretasikan menjadi rumusan rekomendasi yang berbasis bukti (*evidence-based*) bagi pemerintah daerah dan pemangku kepentingan strategis dalam mengakselerasi program inklusi serta transformasi digital yang tepat sasaran.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif komputasional dengan menerapkan teknik *data mining* untuk mengklasifikasikan tingkat kesiapan digital masyarakat pedesaan di Provinsi Papua Barat. Pendekatan ini dipilih karena *data mining* memungkinkan ekstraksi pola, hubungan, dan wawasan yang tersembunyi dari sekumpulan data berskala besar melalui algoritma *machine learning* (Han, Pei, & Tong, 2022). Secara spesifik, metode pembelajaran terarah (*supervised learning*) berupa klasifikasi digunakan untuk membangun model prediktif yang dapat mengelompokkan masyarakat ke dalam tiga kategori kesiapan digital, yakni rendah, sedang, dan tinggi. Kerangka kerja yang diadaptasi dalam penelitian ini mengacu pada metodologi *Knowledge Discovery in Databases* (KDD), yang memberikan alur sistematis mulai dari pemahaman domain masalah, akuisisi data, pra-pemrosesan, pemodelan data, hingga evaluasi dan interpretasi pengetahuan yang dihasilkan.

Populasi dalam penelitian ini mencakup masyarakat yang berdomisili di wilayah pedesaan Provinsi Papua Barat yang telah memiliki eksposur terhadap penggunaan perangkat digital maupun akses internet. Mengingat karakteristik geografis dan besarnya populasi yang tidak diketahui secara pasti, akuisisi data dilakukan menggunakan metode pengambilan sampel non-probabilitas dengan teknik *purposive sampling*. Kriteria inklusi yang ditetapkan mengharuskan responden berusia minimal 17 tahun dan memiliki pengalaman dalam menggunakan layanan digital. Berdasarkan kriteria tersebut, data primer diakuisisi secara daring dari 400 responden menggunakan instrumen kuesioner. Pengukuran melibatkan sembilan atribut independen (prediktor) dan satu atribut dependen (target), yang secara rinci dijabarkan pada Tabel 1.

Tabel 1. Operasionalisasi Variabel Penelitian

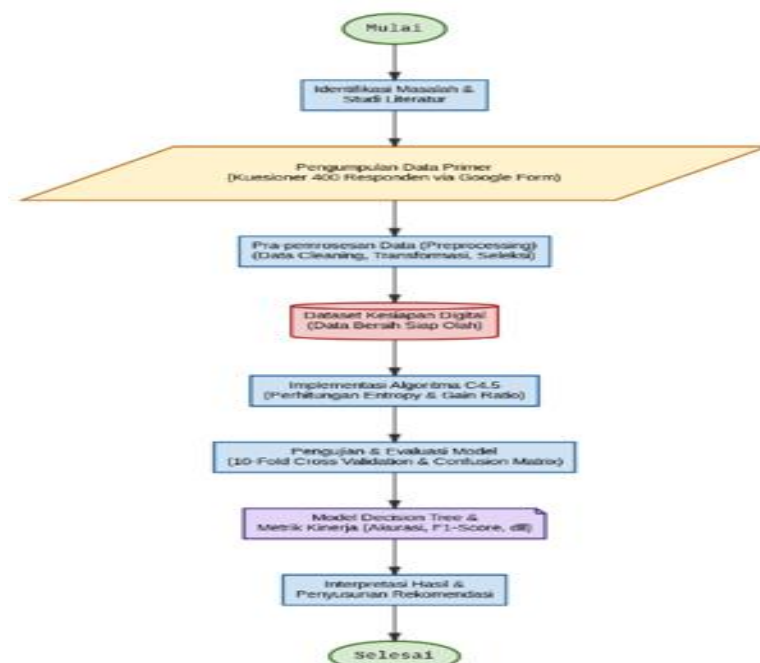
Variabel	Indikator	Skala	Peran
Usia	Kelompok usia responden	Ordinal	Prediktor
Jenis Kelamin	Laki-laki / Perempuan	Nominal	Prediktor
Pendidikan	Pendidikan terakhir	Ordinal	Prediktor
Pekerjaan	Jenis pekerjaan	Nominal	Prediktor
Pendapatan	Pendapatan bulanan	Ordinal	Prediktor
Akses Teknologi Digital	Kepemilikan perangkat dan akses internet	Likert	Prediktor
Literasi Digital	Kemampuan penggunaan teknologi digital	Likert	Prediktor
Keamanan Digital	Pemahaman keamanan dan privasi digital	Likert	Prediktor
Pemanfaatan Teknologi Digital	Tingkat penggunaan layanan digital	Likert	Prediktor

Tingkat Kesiapan Digital

Kategori: Rendah, Sedang, Tinggi

Nominal Target
(Label)

Proses penelitian dilaksanakan melalui serangkaian tahapan terstruktur yang divisualisasikan dalam bentuk *flowchart* pada gambar 1. Alur tersebut diawali dengan tahap identifikasi masalah terkait kesenjangan digital di Papua Barat dan dilanjutkan dengan perancangan instrumen survei. Tahap selanjutnya adalah pengumpulan data mentah dari 400 responden yang kemudian dilanjutkan ke tahap pra-pemrosesan data (*preprocessing*). Tahapan pra-pemrosesan merupakan fase krusial dalam *data mining* untuk memastikan kualitas model yang dihasilkan (Kotu & Deshpande, 2023). Pada fase ini, sistem melakukan *data cleaning* untuk menghapus nilai yang tidak lengkap atau duplikat, *data transformation* untuk mengonversi nilai skor Likert menjadi data kategorikal klasifikasi, dan *data selection*. Setelah dataset bersih terbentuk, alur berlanjut pada implementasi *data mining* menggunakan algoritma C4.5 berbantuan perangkat lunak RapidMiner untuk membangun pohon keputusan (*decision tree*). Diagram alir bermuara pada tahap evaluasi performa model menggunakan *confusion matrix* dan ekstraksi aturan klasifikasi untuk perumusan rekomendasi kebijakan kesiapan digital.



Gambar 1. Kerangka Kerja (Framework) Penelitian

Algoritma C4.5 dipilih sebagai metode klasifikasi utama karena kemampuannya menangani data kategorikal serta menghasilkan struktur pohon keputusan yang transparan dan mudah diinterpretasikan oleh pemangku kebijakan (Witten et al., 2023). Algoritma ini bekerja dengan pendekatan *divide and conquer*, membangun pohon secara *top-down* rekursif. Pemilihan atribut sebagai simpul akar (*root node*) maupun simpul cabang didasarkan pada serangkaian perhitungan berbasis teori informasi. Langkah pertama adalah menghitung nilai *Entropy* untuk mengukur tingkat ketidakpastian atau heterogenitas himpunan data (S) dengan menggunakan Persamaan 1.

$$\text{Entropy}(S) = - \sum (p_i \times \log_2 p_i) \quad (1)$$

Keterangan:

- Entropy(S) = nilai entropy dataset S
- p_i = proporsi data pada kelas ke- i
- Σ = penjumlahan seluruh kelas

Di mana p_i merupakan proporsi data yang termasuk dalam kelas ke- i , dan n adalah jumlah kelas kesiapan digital. Setelah *Entropy* total diketahui, langkah kedua adalah menghitung *Information Gain* untuk setiap atribut (A) guna mengevaluasi seberapa besar atribut tersebut mampu mereduksi ketidakpastian awal. Perhitungan *Information Gain* direpresentasikan pada Persamaan 2.

$$\text{Gain}(S,A) = \text{Entropy}(S) - \sum ((|S_v| / |S|) \times \text{Entropy}(S_v)) \quad (2)$$

Keterangan:

- $\text{Gain}(S,A)$ = nilai information gain atribut A
- $\text{Entropy}(S)$ = entropy seluruh dataset
- S_v = subset data berdasarkan nilai atribut tertentu
- $|S_v|$ = jumlah data pada subset
- $|S|$ = jumlah seluruh data

Untuk menghindari bias pemilihan terhadap atribut yang memiliki cabang atau kategori terlalu banyak, algoritma C4.5 melakukan langkah ketiga, yakni menghitung *Split Information* sebagai faktor penormalisasi menggunakan Persamaan 3.

$$\text{SplitInfo}(S,A) = - \sum ((|S_v| / |S|) \times \log_2(|S_v| / |S|)) \quad (3)$$

Keterangan:

- $\text{SplitInfo}(S,A)$ = nilai split information atribut A
- S_v = subset data berdasarkan nilai atribut tertentu
- $|S_v|$ = jumlah data pada subset
- $|S|$ = jumlah seluruh data

Nilai split information digunakan untuk menghitung gain ratio. Langkah keempat adalah menentukan atribut terbaik dengan membagi nilai *Information Gain* dengan *Split Information* untuk mendapatkan *Gain Ratio* (Persamaan 4). Atribut dengan nilai *Gain Ratio* tertinggi otomatis ditetapkan sebagai simpul keputusan untuk memecah data. Langkah-langkah matematis ini diulang secara rekursif pada setiap cabang hingga membentuk daun (*leaf*) homogen. Mekanisme pemangkasan pohon (*pruning*) kemudian diterapkan untuk mereduksi cabang yang tidak signifikan guna menghindari *overfitting*.

$$\text{GainRatio}(A) = \text{Gain}(S,A) / \text{SplitInfo}(S,A) \quad (4)$$

Keterangan:

- $\text{GainRatio}(A)$ = nilai gain ratio atribut A
- $\text{Gain}(S,A)$ = nilai information gain atribut A
- $\text{SplitInfo}(S,A)$ = nilai split information atribut A

Skenario Pengujian dan Evaluasi Model

Untuk memastikan keandalan model, skenario pengujian dilakukan menggunakan teknik *10-Fold Cross Validation*. Mekanisme ini membagi dataset menjadi 10 partisi yang berukuran sama, di mana proses iterasi pelatihan dan pengujian digilir sebanyak sepuluh kali agar estimasi performa lebih stabil dan terhindar dari bias (Han, Pei, & Tong, 2022). Kinerja model dievaluasi secara komprehensif berdasarkan hasil *Confusion Matrix* multikelas. Terdapat empat metrik kuantitatif utama yang dihitung, yaitu akurasi (*Accuracy*), presisi (*Precision*), *Recall*, dan *F1-Score*, yang berturut-turut dihitung menggunakan Persamaan 5, 6, 7, dan 8.

$$\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN}) \times 100\% \quad (5)$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP}) \times 100\% \quad (6)$$

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN}) \times 100\% \quad (7)$$

$$\text{F1-Score} = 2 \times ((\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})) \quad (8)$$

Berdasarkan formulasi tersebut, TP melambangkan *True Positive*, TN adalah *True Negative*, FP adalah *False Positive*, dan FN mewakili *False Negative*. Akurasi mengukur ketepatan prediksi keseluruhan, sedangkan *F1-Score* memberikan representasi keseimbangan performa antara presisi

dan *recall*. Metrik-metrik ini secara kolektif memvalidasi kelayakan algoritma C4.5 dalam mengklasifikasikan kesiapan digital masyarakat pedesaan.

HASIL DAN PEMBAHASAN

Bagian ini menyajikan hasil dengan deskripsi yang jelas. Hasil dapat dilengkapi dengan gambar dan tabel. Bagian pembahasan memaparkan hasil pengolahan data, menginterpretasikan penemuan secara logis, mengaitkan dengan sumber rujukan yang relevan, dan implikasi dari temuan. {Footlight MT Light, size 12, Spacing: before 0 pt; after 0 pt, Line spacing: 1}

Karakteristik Responden dan Profil Kesiapan Digital

Penelitian ini menggunakan data primer yang diperoleh dari penyebaran kuesioner secara daring kepada masyarakat pedesaan di Provinsi Papua Barat dengan total sampel akhir sebanyak 400 responden yang memenuhi kriteria pengujian. Profil demografis responden dikelompokkan berdasarkan jenis kelamin, usia, tingkat pendidikan terakhir, jenis pekerjaan, dan pendapatan bulanan guna memberikan konteks awal yang representatif mengenai kondisi subjek penelitian di lapangan. Data demografis ini menjadi basis variabel prediktor yang krusial dalam memetakan bagaimana faktor-faktor non-teknis berinteraksi dengan tingkat kesiapan teknologi masyarakat di wilayah pedesaan.

Komposisi responden berdasarkan jenis kelamin menunjukkan distribusi yang relatif seimbang, di mana responden laki-laki mendominasi dengan jumlah 217 orang (54,25%), sementara responden perempuan berjumlah 183 orang (45,75%). Berdasarkan kelompok usia, mayoritas responden berada pada rentang produktif, dengan kelompok usia 17 hingga 25 tahun sebagai kelompok terbesar yang mencakup 124 individu (31,00%), diikuti oleh kelompok usia 26 hingga 35 tahun sebanyak 108 individu (27,00%), kelompok usia 36 hingga 45 tahun sebanyak 82 individu (20,50%), kelompok usia 46 hingga 55 tahun sebanyak 56 individu (14,00%), serta kelompok usia di atas 55 tahun sebagai segmen terkecil dengan jumlah 30 individu (7,50%). Profil usia ini menunjukkan dominasi generasi muda dan dewasa produktif yang secara teoritis memiliki kedekatan lebih tinggi terhadap adopsi teknologi digital.

Tingkat pendidikan terakhir responden didominasi oleh lulusan sekolah menengah. Lulusan SMA/SMK/Sederajat menempati porsi terbesar dengan jumlah 186 responden (46,50%), disusul oleh lulusan SMP/Sederajat sebanyak 76 responden (19,00%), lulusan Sarjana (S1) sebanyak 58 responden (14,50%), lulusan SD/Sederajat sebanyak 42 responden (10,50%), dan Diploma sebanyak 38 responden (9,50%). Pada aspek mata pencaharian, sektor pertanian merupakan pilar utama pekerjaan responden dengan jumlah 98 orang (24,50%), diikuti oleh wiraswasta sebanyak 76 orang (19,00%), pegawai swasta sebanyak 62 orang (15,50%), pelajar atau mahasiswa sebanyak 58 orang (14,50%), nelayan sebanyak 44 orang (11,00%), profesi lainnya sebanyak 34 orang (8,50%), dan ASN sebagai kelompok terkecil dengan 28 orang (7,00%). Sejalan dengan profil pekerjaan tersebut, tingkat pendapatan bulanan responden terkonsentrasi pada kelas menengah bawah, di mana rentang pendapatan antara Rp1.500.000 hingga Rp3.000.000 mencakup porsi terbesar yaitu sebanyak 128 orang (32,00%), sedangkan pendapatan di bawah Rp1.500.000 diwakili oleh 92 orang (23,00%), pendapatan antara Rp3.000.001 hingga Rp5.000.000 sebanyak 104 orang (26,00%), dan pendapatan di atas Rp5.000.000 sebanyak 76 orang (19,00%). Distribusi demografis ini seperti pada pada Tabel 2.

Tabel 2. Distribusi Karakteristik Demografis Responden (N = 400)

Karakteristik	Kategori	Jumlah (N)	Persentase (%)
Jenis Kelamin	Laki-laki	217	54,25
	Perempuan	183	45,75
Usia	17–25 Tahun	124	31,00
	26–35 Tahun	108	27,00
	36–45 Tahun	82	20,50
	46–55 Tahun	56	14,00
	>55 Tahun	30	7,50
Pendidikan	SD/Sederajat	42	10,50
	SMP/Sederajat	76	19,00

Pekerjaan	SMA/SMK/Sederajat	186	46,50
	Diploma	38	9,50
	Sarjana (S1)	58	14,50
	Petani	98	24,50
	Nelayan	44	11,00
	Wiraswasta	76	19,00
	Pegawai Swasta	62	15,50
	ASN	28	7,00
	Pelajar/Mahasiswa	58	14,50
Pendapatan	Lainnya	34	8,50
	< Rp1.500.000	92	23,00
	Rp1.500.000–Rp3.000.000	128	32,00
	Rp3.000.001–Rp5.000.000	104	26,00
	> Rp5.000.000	76	19,00

Selain profil demografis, kepemilikan perangkat keras juga dianalisis untuk melihat kesiapan fisik responden. Mayoritas responden telah memiliki telepon pintar (*smartphone*) dengan persentase kepemilikan mencapai 95,25% (381 responden), disusul oleh laptop sebesar 42,00% (168 responden), komputer desktop sebesar 13,00% (52 responden), dan komputer tablet sebesar 9,25% (37 responden). Tingginya kepemilikan *smartphone* menunjukkan penetrasi perangkat keras yang masif di pedesaan, yang menjadi gerbang utama interaksi digital masyarakat.

Untuk variabel tingkat literasi digital, analisis awal menunjukkan bahwa mayoritas responden berada pada kategori sedang sebanyak 198 orang (49,50%), kategori tinggi sebanyak 106 orang (26,50%), dan kategori rendah sebanyak 96 orang (24,00%). Pada aspek keamanan digital, sebagian besar responden juga menempati kategori sedang sebesar 45,25% (181 orang), disusul kategori rendah sebesar 28,50% (114 orang), dan kategori tinggi sebesar 26,25% (105 orang). Pola yang serupa ditemukan pada tingkat pemanfaatan teknologi digital, di mana kategori sedang mendominasi sebesar 50,50% (202 orang), kategori tinggi sebesar 28,50% (114 orang), dan kategori rendah sebesar 21,00% (84 orang). Sintesis akhir terhadap seluruh indikator ini menghasilkan klasifikasi kelas target, yaitu tingkat kesiapan digital masyarakat pedesaan di Provinsi Papua Barat, yang didominasi oleh kategori sedang dengan total 214 responden (53,50%), kategori tinggi sebanyak 94 responden (23,50%), dan kategori rendah sebanyak 92 responden (23,00%). Distribusi profil kesiapan digital ini tersaji secara komprehensif pada Tabel 3.

Tabel 3. Profil Distribusi Variabel Digital dan Kesiapan Digital Responden

Variabel Digital	Kategori	Jumlah (N)	Persentase (%)
Literasi Digital	Rendah	96	24,00
	Sedang	198	49,50
	Tinggi	106	26,50
Keamanan Digital	Rendah	114	28,50
	Sedang	181	45,25
	Tinggi	105	26,25
Pemanfaatan Digital	Rendah	84	21,00
	Sedang	202	50,50
	Tinggi	114	28,50
Kesiapan Digital (Kelas Target)	Rendah	92	23,00
	Sedang	214	53,50
	Tinggi	94	23,50

Prakondisi Data dan Preprocessing

Sebelum data survei digunakan untuk melatih algoritma C4.5, dilakukan serangkaian tahapan preprocessing guna menjamin kualitas data input agar model klasifikasi yang dihasilkan terhindar dari bias keputusan. Tahapan prakondisi data ini mengikuti metodologi Knowledge Discovery in Databases yang meliputi pembersihan data (*data cleaning*), transformasi data, serta

seleksi atribut. Pembersihan data (*data cleaning*) difokuskan untuk mengeliminasi ketidakkonsistenan pengisian kuesioner. Dari 417 rekaman data awal yang terkumpul melalui Google Form, terdapat 9 rekaman data yang tidak lengkap, 5 data duplikat akibat pengiriman ganda oleh responden, dan 3 data dengan jawaban yang tidak konsisten. Setelah proses pembersihan ini, sebanyak 17 rekaman data dihapus secara permanen sehingga menyisakan dataset bersih berukuran 400 baris data yang siap diolah lebih lanjut.

Tahap selanjutnya adalah transformasi data, di mana data kuantitatif yang awalnya berupa skor skala Likert dikonversi menjadi kategori ordinal untuk mengakomodasi kebutuhan representasi cabang pada algoritma C4.5. Konversi variabel didasarkan pada akumulasi skor indikator pada tiap dimensi kuesioner. Sebagai contoh, skor pada variabel akses teknologi digital dikategorikan menjadi rendah untuk rentang nilai 5 hingga 11, kategori sedang untuk rentang nilai 12 hingga 18, dan kategori tinggi untuk rentang nilai 19 hingga 25. Transformasi serupa diterapkan secara konsisten pada dimensi literasi digital, keamanan digital, pemanfaatan teknologi digital, dan dimensi kesiapan digital sebagai kelas target. Melalui seleksi atribut, diputuskan bahwa kesembilan atribut demografis dan digital digunakan sebagai variabel prediktor dengan variabel kesiapan digital berperan sebagai variabel label, membentuk dataset akhir nominal yang optimal untuk analisis pohon keputusan.

Pemodelan Klasifikasi dan Evaluasi Kinerja Algoritma C4.5

Implementasi algoritma C4.5 dilakukan menggunakan perangkat lunak RapidMiner Studio dengan parameter pengujian yang diatur secara ketat guna meminimalkan risiko kejenuhan model (*overfitting*). Algoritma C4.5 bekerja secara rekursif dengan menghitung nilai entropy, split information, dan gain ratio dari seluruh atribut prediktor untuk menentukan atribut yang paling representatif sebagai pemisah data (*node* keputusan). Berdasarkan hasil pemrosesan matematis terhadap 400 data responden, nilai entropy total dataset berada pada angka 1,461. Perhitungan gain ratio untuk masing-masing atribut prediktor dirangkum dalam Tabel 4.

Tabel 4. Nilai Gain Rasio Atribut Prediktor dalam Algoritma C4.5

Peringkat	Atribut Prediktor	Gain Ratio	Peran dalam Model
1	Literasi Digital	0,421	Root Node (Akar Utama)
2	Pemanfaatan Teknologi Digital	0,387	Internal Node
3	Akses Teknologi Digital	0,346	Internal Node
4	Keamanan Digital	0,302	Internal Node
5	Pendidikan	0,191	Atribut Pendukung
6	Pendapatan	0,164	Atribut Pendukung
7	Usia	0,133	Atribut Pendukung
8	Pekerjaan	0,087	Atribut Pendukung
9	Jenis Kelamin	0,024	Atribut Pendukung

Merujuk pada hasil perhitungan Gain Ratio pada Tabel 4, atribut literasi digital terpilih sebagai *root node* karena memiliki nilai tertinggi yaitu sebesar 0,421. Hasil ini menegaskan bahwa tingkat kemampuan literasi informasi dan kognitif teknologi dari individu merupakan faktor penentu paling dominan dalam membedakan tingkat kesiapan digital masyarakat pedesaan. Di sisi lain, faktor demografis murni seperti jenis kelamin memiliki nilai gain ratio terendah (0,024), yang mengindikasikan signifikansi pengaruh yang minimal dalam struktur pohon keputusan.

Struktur pohon keputusan yang berhasil diekstraksi dari algoritma C4.5 menghasilkan serangkaian aturan keputusan (*classification rules*). Aturan keputusan pertama menetapkan bahwa apabila tingkat literasi digital dikategorikan tinggi dan pemanfaatan teknologi digital berada pada kategori tinggi, maka kesiapan digital diprediksi berada pada tingkat tinggi dengan dukungan data sebanyak 78 responden dan tingkat kepercayaan mencapai 96,15%. Aturan keputusan kedua menyatakan bahwa apabila literasi digital berkategori tinggi, namun pemanfaatan teknologi digital berada pada level sedang, keputusan klasifikasi akan beralih ke variabel keamanan digital; jika keamanan digital dikategorikan tinggi, maka kesiapan digital bernilai tinggi dengan tingkat kepercayaan sebesar 92,86%. Aturan keputusan ketiga merumuskan bahwa apabila tingkat literasi digital berada pada kategori sedang dan didukung oleh tingkat akses

teknologi digital yang tinggi, maka kesiapan digital diprediksi sedang dengan tingkat kepercayaan sebesar 89,77%. Sebaliknya, aturan keputusan keempat menetapkan bahwa jika literasi digital bernilai sedang namun pemanfaatan teknologi digital rendah, maka tingkat kesiapan digital bernilai rendah dengan tingkat kepercayaan sebesar 86,11%. Aturan keputusan kelima menegaskan bahwa apabila literasi digital dikategorikan rendah sejak awal, maka tingkat kesiapan digital responden secara mutlak diprediksi rendah dengan tingkat kepercayaan mencapai 95,06%.

Evaluasi kinerja model klasifikasi dilakukan dengan menggunakan metode 10-Fold Cross Validation untuk menjamin stabilitas hasil pengujian. Dataset dibagi menjadi sepuluh bagian secara acak, di mana secara bergantian sembilan bagian bertindak sebagai data latih dan satu bagian bertindak sebagai data uji hingga seluruh bagian teruji. Hasil evaluasi disajikan dalam bentuk matriks kebingungan (*confusion matrix*) multiclass yang membandingkan nilai aktual dengan nilai prediksi model, seperti yang disajikan pada Tabel 5.

Tabel 5. Confusion Matrix Multiclass Model C4.5

Aktual \ Prediksi	Rendah (Prediksi)	Sedang (Prediksi)	Tinggi (Prediksi)	Recall Kelas (%)
Rendah (Aktual)	80	10	2	86,96
Sedang (Aktual)	12	189	13	88,32
Tinggi (Aktual)	1	7	86	91,49
Precision Kelas (%)	86,02	91,75	85,15	Akurasi: 88,75%

Berdasarkan data pada Tabel 5, dari total 400 sampel, model klasifikasi C4.5 berhasil memprediksi secara akurat sebanyak 355 data responden (80 kelas rendah, 189 kelas sedang, dan 86 kelas tinggi), sedangkan sisa 45 data lainnya mengalami salah klasifikasi (*misclassification*). Kinerja keseluruhan model dinilai baik dengan capaian nilai akurasi (*accuracy*) rata-rata sebesar 88,75%. Keseimbangan prediksi antarkelas juga terjaga dengan baik, ditandai dengan nilai presisi (*precision*) rata-rata sebesar 87,92%, sensitivitas (*recall*) rata-rata sebesar 88,21%, dan nilai F1-Score sebesar 88,06%. Hasil ini membuktikan secara ilmiah keandalan algoritma C4.5 dalam mengklasifikasikan kesiapan digital masyarakat pedesaan.

Pembahasan

Analisis terhadap nilai gain ratio dan struktur pohon keputusan yang terbentuk menunjukkan secara konsisten bahwa dimensi kompetensi individu yang direpresentasikan oleh variabel literasi digital (0,421) dan pemanfaatan teknologi digital (0,387) memiliki dampak pengaruh yang jauh lebih signifikan dibandingkan dengan ketersediaan fisik teknologi yang direpresentasikan oleh akses teknologi digital (0,346). Hasil penelitian ini sejalan dengan kerangka konseptual yang dikemukakan oleh *UNESCO (2023)*, yang menegaskan bahwa literasi digital bukan sekadar kemampuan mekanis untuk mengoperasikan gawai pintar, melainkan kapasitas kognitif untuk memproses, menganalisis, dan memanfaatkan informasi secara cerdas di ruang siber. Fenomena dominasi literasi digital ini mengonfirmasi temuan dari penelitian terdahulu oleh *Purwanto et al. (2024)* yang menyatakan bahwa literasi digital bertindak sebagai determinan utama yang menggerakkan kesiapan individu untuk beradaptasi dengan perubahan ekosistem digital.

Ketika literasi digital masyarakat pedesaan berada pada level rendah, algoritma C4.5 langsung mengklasifikasikan mereka ke dalam kategori kesiapan digital rendah dengan tingkat kepercayaan yang tinggi (95,06%). Pola linier ini menunjukkan bahwa tanpa adanya pembekalan keterampilan digital yang memadai, penyediaan infrastruktur internet di pedesaan Papua Barat akan berujung pada kondisi pemanfaatan yang tidak optimal atau bahkan berisiko memicu dampak siber negatif akibat kurangnya pemahaman keamanan digital. Hal ini relevan dengan kajian *Ahmed et al. (2024)* yang menekankan pentingnya intervensi program pelatihan literasi secara terstruktur untuk menjembatani jurang pemisah antara kepemilikan gawai dengan kesiapan fungsional penggunaannya.

Atribut pemanfaatan teknologi digital menempati posisi strategis kedua sebagai faktor dominan dengan nilai gain ratio sebesar 0,387. Pola klasifikasi membuktikan bahwa masyarakat

pedesaan yang aktif menggunakan internet untuk kebutuhan produktif, seperti transaksi keuangan digital, layanan niaga-el (*e-commerce*), dan pencarian materi edukasi fungsional, memiliki kecenderungan masuk ke dalam kelas kesiapan digital tinggi sebesar 96,15% jika ditunjang literasi digital yang mumpuni. Sebaliknya, apabila pemanfaatan internet sebatas untuk sarana hiburan pasif atau komunikasi dasar, tingkat kesiapan digital responden tertahan pada kategori sedang atau bahkan jatuh ke kategori rendah. Pola ini memperkuat argumen *Lestari et al. (2024)* bahwa pergeseran dari konsumsi konten hiburan menuju pemanfaatan teknologi yang bernilai guna ekonomi merupakan indikator matangnya kesiapan digital di tingkat komunitas perdesaan.

Meskipun akses teknologi digital (0,346) berada di bawah pengaruh literasi dan pemanfaatan, ketersediaan gawai dan stabilitas koneksi internet tetap menjadi prasyarat fisik yang mutlak diperlukan. Berdasarkan data empiris kepemilikan gawai pintar yang tinggi (95,25%), kendala utama kesenjangan digital di pedesaan Provinsi Papua Barat saat ini bukan lagi terletak pada keterbatasan fisik kepemilikan perangkat keras pribadi, melainkan pada keandalan infrastruktur jaringan telekomunikasi dan jaminan ketersediaan sinyal internet yang stabil di pelosok desa. Hasil klasifikasi memperlihatkan bahwa akses teknologi digital yang memadai berandil besar dalam meningkatkan posisi responden dari kategori kesiapan digital rendah menuju sedang, yang selaras dengan laporan dari *Badan Pusat Statistik (2024)* mengenai dinamika penetrasi telekomunikasi pedesaan di wilayah timur Indonesia.

Variabel keamanan digital (0,302) teridentifikasi sebagai atribut penunjang yang penting dalam memperkuat tingkat kesiapan digital masyarakat. Sebagaimana dimodelkan dalam aturan keputusan pohon klasifikasi, responden dengan tingkat literasi tinggi yang tidak dibarengi dengan pemahaman proteksi keamanan informasi pribadi yang baik akan diturunkan prediksinya menjadi tingkat kesiapan digital sedang. Temuan ini mengindikasikan bahwa kewaspadaan terhadap ancaman kejahatan siber, penipuan daring, dan kecakapan mengamankan kredensial akun digital merupakan parameter krusial untuk menciptakan lingkungan interaksi digital yang aman dan berkelanjutan, sebuah argumen yang didukung penuh oleh riset *Sari et al. (2023)* tentang pentingnya aspek keamanan siber pada masyarakat berkembang.

Kontras dengan faktor-faktor kompetensi digital fungsional tersebut, faktor demografis murni seperti tingkat pendidikan terakhir (0,191), pendapatan (0,164), usia (0,133), pekerjaan (0,087), dan jenis kelamin (0,024) menempati posisi gain ratio terbawah. Hal ini mengimplikasikan bahwa kendatipun tingkat pendidikan formal dan taraf pendapatan memiliki korelasi linear yang positif terhadap kemudahan adopsi teknologi, faktor demografis eksternal ini tidak lagi menjadi sekat mutlak bagi masyarakat pedesaan untuk mencapai status kesiapan digital yang tinggi. Seseorang dengan latar belakang pendidikan menengah atau pendapatan terbatas tetap memiliki peluang yang sama untuk mencapai kesiapan digital yang optimal asalkan memiliki akses latihan mandiri, literasi digital yang baik, serta kemauan memanfaatkan teknologi untuk hal produktif. Temuan ini menyajikan perspektif optimis bagi percepatan inklusi digital pedesaan, sejalan dengan visi *Kementerian Komunikasi dan Digital RI (2024)* dalam mempromosikan pemerataan ekosistem digital nasional.

Secara penelitian ini membuktikan efektivitas pemanfaatan algoritma C4.5 dalam studi kemasyarakatan digital. Model klasifikasi yang divalidasi dengan metode 10-Fold Cross Validation mampu menyajikan akurasi klasifikasi sebesar 88,75%. Capaian akurasi ini tidak hanya membuktikan keandalan algoritma C4.5 yang bersaing dengan algoritma *supervised learning* modern lainnya, tetapi juga menegaskan aspek transparansi interpretasi model dalam bentuk pohon keputusan (*decision tree*) dan aturan keputusan (*decision rules*). Sebagaimana dijelaskan oleh *Witten et al. (2023)*, keuntungan utama pohon keputusan adalah kemampuannya menyajikan logika klasifikasi yang transparan dan akuntabel, berbeda dengan model klasifikasi berbasis kotak hitam (*black box*) yang sulit dipahami oleh pengambil kebijakan non-teknis di pemerintahan daerah.

Implikasi kebijakan yang lahir dari temuan penelitian ini menegaskan bahwa strategi pembangunan ekosistem digital di wilayah pedesaan Provinsi Papua Barat tidak boleh lagi berfokus secara eksklusif pada pendekatan fisik pembangunan infrastruktur jaringan telekomunikasi (*supply-side approach*). Pemerintah daerah dan pemangku kepentingan harus mengimbangi

pembangunan fisik tersebut dengan program pemberdayaan kapasitas sumber daya manusia secara intensif (*demand-side approach*). Rekomendasi program nyata yang mendesak untuk diimplementasikan meliputi penyelenggaraan pelatihan literasi digital fungsional tingkat desa yang fokus pada literasi informasi, pelatihan pemanfaatan aplikasi produktif berbasis ekonomi lokal, edukasi keamanan data pribadi, serta sosialisasi literasi keuangan digital untuk transaksi aman. Peta klasifikasi kesiapan digital yang didominasi oleh kelas sedang (53,50%) ini memberikan sinyal kuat bahwa masyarakat pedesaan Papua Barat berada pada fase transisi kritis; intervensi kebijakan pelatihan literasi digital yang tepat waktu dan terarah akan mendorong kelompok mayoritas sedang ini menuju tingkat kesiapan digital tinggi, sekaligus mencegah mereka jatuh kembali ke dalam jebakan kesenjangan digital yang lebih dalam.

KESIMPULAN

Penelitian ini telah berhasil menjawab rumusan masalah dengan membangun model klasifikasi tingkat kesiapan digital masyarakat pedesaan di Provinsi Papua Barat menggunakan algoritma C4.5 berdasarkan data survei terhadap 400 responden. Hasil evaluasi menggunakan metode *10-Fold Cross Validation* membuktikan bahwa model pohon keputusan (*decision tree*) yang dihasilkan andal dan efektif dengan capaian akurasi sebesar 88,75%, presisi 87,92%, *recall* 88,21%, dan *F1-Score* 88,06%. Berdasarkan ekstraksi pola data, pemetaan kondisi di lapangan menunjukkan bahwa mayoritas masyarakat pedesaan saat ini berada pada fase transisi dengan tingkat kesiapan digital kategori sedang (53,50%), disusul oleh kategori tinggi (23,50%) dan rendah (23,00%). Lebih lanjut, penelitian ini mengonfirmasi bahwa kompetensi fungsional individu merupakan determinan utama; di mana atribut literasi digital teridentifikasi sebagai faktor dominan tertinggi yang paling memengaruhi kesiapan digital masyarakat (nilai *gain ratio* 0,421), diikuti secara berurutan oleh pemanfaatan teknologi digital (0,387), akses teknologi (0,346), dan keamanan digital (0,302). Temuan ini secara empiris menegaskan bahwa pemerataan infrastruktur fisik dan faktor demografis tidak lagi menjadi batasan mutlak, karena kesiapan digital yang optimal jauh lebih bergantung pada kapasitas kognitif masyarakat dalam memproses informasi serta memanfaatkan teknologi secara produktif, aman, dan bertanggung jawab.

Menindaklanjuti temuan analitik tersebut, prospek aplikasi strategis yang direkomendasikan adalah perlunya pergeseran fokus kebijakan pemerintah daerah dari pendekatan yang semata-mata berpusat pada ketersediaan infrastruktur jaringan menuju pemberdayaan kapasitas sumber daya manusia secara intensif. Intervensi ini dapat diaplikasikan melalui penyelenggaraan program pelatihan literasi digital fungsional, edukasi privasi dan keamanan siber, serta penciptaan layanan publik digital yang inklusif untuk mendorong masyarakat pedesaan agar lebih berdaya guna dalam sektor pendidikan dan ekonomi lokal. Sementara itu, untuk prospek pengembangan penelitian selanjutnya, pemodelan ini potensial untuk diekspansi menggunakan skala responden yang lebih masif dan cakupan demografis yang lebih luas guna memfasilitasi analisis komparatif antarprovinsi. Peneliti di masa mendatang juga direkomendasikan untuk mengintegrasikan variabel prediktor makro-struktural, seperti kualitas dukungan regulasi pemerintah dan stabilitas telekomunikasi, serta membandingkan arsitektur algoritma C4.5 dengan model *machine learning* tingkat lanjut atau pendekatan ansambel seperti *Random Forest*, *Support Vector Machine* (SVM), dan *Neural Network* guna mengeksplorasi optimasi akurasi prediksi secara lebih holistik.

DAFTAR PUSTAKA

- Ahmed, S., Hossain, M., & Rahman, M. (2024). Digital readiness assessment and classification using machine learning approaches. *International Journal of Advanced Computer Science and Applications*, 15(2), 112–121.
- Alharbi, A., Aljuaid, H., & Alotaibi, R. (2023). Decision tree algorithms for classification problems: A comparative study. *International Journal of Computer Science and Network Security*, 23(5), 87–96.
- Badan Pusat Statistik. (2024). *Statistik telekomunikasi Indonesia 2024*. Jakarta: BPS.
- Badan Pusat Statistik Provinsi Papua Barat. (2024). *Papua Barat dalam angka 2024*. Manokwari: BPS Provinsi Papua Barat.

- Chen, Y., Liu, H., & Wang, X. (2023). Machine learning approaches for digital inclusion assessment in rural communities. *Journal of Information Technology Research*, 16(3), 44–59.
- Direktorat Jenderal Aplikasi Informatika. (2023). *Status literasi digital Indonesia 2023*. Jakarta: Kementerian Komunikasi dan Informatika Republik Indonesia.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A primer on partial least squares structural equation modeling (PLS-SEM)* (3rd ed.). Thousand Oaks: Sage Publications.
- Han, J., Pei, J., & Tong, H. (2022). *Data mining: Concepts and techniques* (4th ed.). Burlington: Morgan Kaufmann.
- Hidayat, R., Nugroho, A., & Prasetyo, D. (2024). Analysis of digital literacy and digital readiness among rural communities in Indonesia. *Journal of Information Systems Engineering and Business Intelligence*, 10(1), 25–36.
- International Telecommunication Union. (2023). *Measuring digital development: Facts and figures 2023*. Geneva: ITU.
- International Telecommunication Union. (2024). *Digital development dashboard 2024*. Geneva: ITU.
- Ismail, M., Rahman, F., & Abdullah, N. (2022). Application of C4.5 algorithm in socio-economic classification studies. *International Journal of Data Science and Analytics*, 8(4), 211–220.
- Kementerian Komunikasi dan Digital Republik Indonesia. (2024). *Indonesia digital outlook 2024*. Jakarta: Komdigi.
- Kotu, V., & Deshpande, B. (2023). *Data science: Concepts and practice* (3rd ed.). Burlington: Morgan Kaufmann.
- Kurniawan, B., Saputra, E., & Wijaya, H. (2023). Classification performance analysis using C4.5 algorithm on social survey data. *Journal of Computer Science and Informatics Engineering*, 7(2), 102–111.
- Lestari, N., Putri, A., & Hidayat, T. (2024). Digital transformation readiness in Indonesian rural communities. *International Journal of Social Informatics*, 12(1), 15–28.
- Marr, B. (2023). *Artificial intelligence in practice* (2nd ed.). Hoboken: Wiley.
- Mulyani, S., Nugraha, D., & Firmansyah, A. (2023). Measuring digital readiness using machine learning techniques: Evidence from developing countries. *Journal of Digital Society Studies*, 5(2), 76–89.
- Organisation for Economic Co-operation and Development. (2023). *OECD digital economy outlook 2023*. Paris: OECD Publishing.
- Organisation for Economic Co-operation and Development. (2024). *Bridging digital divides in rural areas*. Paris: OECD Publishing.
- Prasetyo, Y., Sari, D., & Kurniawati, E. (2022). Decision tree implementation using C4.5 algorithm for classification of community development indicators. *International Journal of Information Systems and Technology*, 6(1), 55–63.
- Purwanto, A., Rahmawati, I., & Setiawan, D. (2024). Digital literacy as a determinant of digital readiness among rural populations. *Journal of Rural Development and Technology*, 9(1), 31–43.
- Rokach, L., & Maimon, O. (2022). *Data mining with decision trees: Theory and applications* (3rd ed.). Singapore: World Scientific.
- Sari, N., Hidayah, R., & Prabowo, M. (2023). Evaluation of classification performance using confusion matrix and cross validation techniques. *Journal of Computer Applications and Data Analytics*, 4(2), 88–99.
- Statista. (2024). *Digital population and internet usage worldwide 2024*. Hamburg: Statista Research Department.
- United Nations Development Programme. (2023). *Human development report 2023/2024*. New York: UNDP.
- United Nations Educational, Scientific and Cultural Organization. (2023). *Digital literacy global framework 2023*. Paris: UNESCO.

- Utami, F., Rahman, A., & Kholid, M. (2024). Predicting digital readiness levels using decision tree algorithms. *International Journal of Information Technology and Decision Making*, 23(2), 301–318.
- Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2023). *Data mining: Practical machine learning tools and techniques* (5th ed.). Cambridge: Morgan Kaufmann.
- World Bank. (2024). *Digital development overview 2024*. Washington DC: World Bank.
- World Economic Forum. (2024). *Future of jobs report 2024*. Geneva: World Economic Forum.
- Yusuf, M., Rahmat, A., & Hidayatullah, F. (2023). Evaluating digital inclusion in rural communities through data mining approaches. *Journal of Information and Communication Technology Research*, 11(3), 145–156.