

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI KASUS PERCERAIAN DI INDONESIA

Edy Nurmansyah¹, Khoiriyah Namira Istiqah², Muhammad Muiddudin³, Samin Arizal⁴, Sifa Dwi Ahmad⁵
Sistem Informasi, Universitas Bina Sarana Informatika^{1,2,3,4,5}

Email: 19252695@bsi.ac.id¹, 19252721@bsi.ac.id², 19252666@bsi.ac.id³, 19252637@bsi.ac.id⁴,
19252729@bsi.ac.id⁵

Informasi	Abstract
-----------	----------

Volume : 2
Nomor : 12
Bulan : Desember
Tahun : 2025
E-ISSN : 3062-9624

Divorce cases in Indonesia continue to show a significant increase. Based on data from Statistics Indonesia (BPS) in 2024, hundreds of thousands of divorce cases were recorded across all provinces. This study applies the C4.5 algorithm to classify the level of divorce cases using a BPS dataset covering 34 provinces, consisting of 21 regular attributes and 2 specific attributes representing the contributing factors of divorce at the regional level. The model was built using cross-validation and pruning mechanisms to reduce the risk of overfitting. The evaluation results show that the model achieved an accuracy of 85.29%, with weighted mean recall and weighted mean precision in the range of 84–85%, indicating a relatively stable classification performance, although not yet optimal. The confusion matrix shows that most predictions are accurate, but there are still misclassifications, particularly between the Medium and High categories. The final decision tree reveals that the “Cerai Gugat” (lawsuit-initiated divorce) attribute serves as the main splitter at the root node, followed by the attribute “Continuous Conflicts and Disputes” at the subsequent level. This indicates that these two attributes have dominant contributions in determining the classification category of divorce cases. Overall, this study demonstrates that the C4.5 algorithm is quite effective in identifying divorce patterns based on BPS data and has the potential to support the development of decision-support systems for government institutions.

Keyword: C4.5 Algorithm, Divorce Classification, Decision Tree, Data Mining, Indonesia

Abstrak

Perceraian di Indonesia terus menunjukkan peningkatan yang signifikan. Berdasarkan data Badan Pusat Statistik (BPS) Tahun 2024, tercatat ratusan ribu kasus perceraian yang tersebar di seluruh provinsi. Penelitian ini menerapkan algoritma C4.5 untuk melakukan klasifikasi tingkat kasus perceraian menggunakan dataset BPS yang mencakup 34 provinsi, terdiri dari 21 atribut reguler dan 2 atribut khusus yang merepresentasikan faktor-faktor penyebab perceraian di tingkat regional. Model dibangun menggunakan validasi silang serta mekanisme pemangkasan untuk mengurangi risiko overfitting. Hasil evaluasi menunjukkan bahwa model menghasilkan akurasi 85,29%, dengan weighted mean recall dan weighted mean precision berada pada kisaran 84–85%, menandakan performa klasifikasi yang stabil meskipun belum sempurna. Confusion matrix menunjukkan bahwa sebagian besar prediksi sudah akurat, namun masih terdapat kesalahan klasifikasi terutama antara kategori Sedang dan Besar. Pohon keputusan akhir menunjukkan bahwa atribut “Cerai Gugat” menjadi pemisah utama pada root node, diikuti oleh atribut “Faktor Perselisihan dan Pertengkarangan Terus Menerus” pada level berikutnya. Hal ini mengindikasikan bahwa kedua atribut tersebut memiliki kontribusi dominan dalam menentukan kategori klasifikasi kasus perceraian. Secara keseluruhan, penelitian ini membuktikan bahwa algoritma C4.5 cukup efektif dalam mengidentifikasi pola

perceraian berbasis data BPS, serta berpotensi mendukung pengembangan sistem pendukung keputusan bagi lembaga pemerintah.

Kata Kunci: *Algoritma C4.5, Klasifikasi Perceraian, Pohon Keputusan, Data Mining, Indonesia*

A. PENDAHULUAN

Fenomena perceraian telah berkembang menjadi tantangan sosial yang signifikan di Indonesia. Berdasarkan data Badan Pusat Statistik (BPS) Tahun 2024, jumlah kasus perceraian nasional tercatat lebih dari 399.000 kasus, menunjukkan bahwa isu ini masih berada pada tingkat yang mengkhawatirkan dan membutuhkan perhatian serius. Tren ini konsisten dengan temuan sebelumnya yang mengindikasikan bahwa perceraian merupakan persoalan struktural yang dipengaruhi berbagai faktor sosial, ekonomi, dan demografis.

Penyebab perceraian di Indonesia memiliki karakteristik multifaktorial, sebagaimana dijelaskan dalam berbagai penelitian terdahulu. Ketidakharmonisan dalam rumah tangga, tekanan ekonomi, kekerasan, perselisihan berkepanjangan, hingga perubahan dinamika sosial turut berkontribusi terhadap meningkatnya angka perceraian di berbagai wilayah. Variasi geografis antarprovinsi juga terlihat dalam data BPS 2024, menunjukkan perbedaan pola kasus yang dipengaruhi oleh kondisi demografi dan sosial budaya lokal.

Revolusi digital telah membuka peluang baru dalam pemahaman fenomena sosial kompleks seperti perceraian melalui penerapan teknik-teknik komputasi mutakhir. Data mining, sebagai disiplin ilmu yang mengombinasikan metodologi statistika dengan algoritma pembelajaran mesin, telah terbukti efektif dalam mengekstraksi pola tersembunyi dari basis data berskala besar. Dalam konteks ini, algoritma pohon keputusan, khususnya C4.5, menawarkan keunggulan kompetitif berupa interpretabilitas tinggi, kemampuan penanganan atribut kategoris dan numeris secara simultan, serta ketangguhan dalam menghadapi dataset yang tidak seimbang (Agnes Irene Silitonga et al., 2025). Algoritma C4.5 telah sukses diterapkan dalam berbagai domain aplikasi mulai dari klasifikasi medis, prediksi akademik, hingga segmentasi pelanggan komersial. Keunggulan fundamental dari pendekatan pohon keputusan ini terletak pada kapasitasnya untuk menghasilkan aturan-aturan klasifikasi yang dapat dengan mudah diinterpretasikan oleh praktisi, memfasilitasi pengambilan keputusan yang lebih transparan dan berbasis bukti empiris (Agnes Irene Silitonga et al., 2025).

Penerapan algoritma C4.5 dalam konteks klasifikasi penyebab perceraian di Indonesia memberikan kontribusi metodologis yang inovatif dan relevan. Studi ini menggunakan dataset BPS yang berisi atribut terkait jenis perceraian dan faktor penyebab perceraian

seperti cerai talak, cerai gugat, serta faktor dominan yang tercatat pada masing-masing provinsi. Studi ini menggunakan validasi silang sepuluh-lipat untuk memastikan reliabilitas model, dengan hasil menunjukkan akurasi keseluruhan sebesar 85,29 persen pada data pelatihan, mengindikasikan kemampuan model dalam mengkaptulasi hubungan kompleks antara variabel input dan output klasifikasi, sedangkan confusion matrix menunjukkan pengklasifikasian tidak sempurna untuk ketiga kategori output yang diprediksi—Sedang, Besar, dan Kecil—dengan nilai-nilai diagonal (Putra et al., 2025)

Signifikansi penelitian ini mencakup dimensi praktis, teoritis, dan kebijakan. Dari perspektif praktis, hasil penelitian dapat memberikan wawasan berharga kepada lembaga-lembaga pemerintah seperti Pengadilan Agama dan Pengadilan Negeri dalam mengidentifikasi wilayah-wilayah dengan risiko perceraian tinggi dan merancang intervensi preventif yang ditargetkan. Dari sudut pandang teoritis, penelitian berkontribusi pada pemahaman mendalam tentang struktur hubungan kausal dalam faktor-faktor perceraian melalui visualisasi pohon keputusan yang komprehensif. Pohon keputusan yang dihasilkan menunjukkan bahwa rata-rata nasional faktor tinggi merupakan node pemisah utama, dengan ambang batas 11367.500 berfungsi sebagai kriteria pembedaan primer antara kategori kasus. Implikasi kebijakan dari temuan ini adalah bahwa intervensi preventif perlu disesuaikan berdasarkan karakteristik regional spesifik, dengan fokus khusus pada daerah yang menunjukkan nilai agregat faktor tinggi.

B. METODE PENELITIAN

Penelitian ini mengadopsi paradigma metodologis kuantitatif dengan desain eksperimental yang terstruktur untuk mengaplikasikan algoritma C4.5 dalam klasifikasi fenomena perceraian nasional. Tahapan awal penelitian melibatkan pengumpulan data sekunder yang bersumber dari website Badan Pusat Statistik. Dataset yang digunakan dalam studi ini mencakup 34 observasi contoh dengan 21 atribut reguler dan 2 atribut khusus, mencakup informasi deskriptif tentang wilayah administratif provinsi, kategori klasifikasi kejadian perceraian, atribut terkait jenis perceraian dan faktor penyebab perceraian yang tercatat oleh BPS. Atribut kritis meliputi identifikasi provinsi, tingkat klasifikasi perceraian yang menjadi variabel target, jumlah kasus perceraian tertalak, jumlah kasus perceraian gugat.

Strategi prapemrosesan data mengikuti protokol standar dalam pipeline data mining, dimulai dengan inspeksi kelengkapan data, identifikasi dan penanganan nilai-nilai hilang,

serta validasi konsistensi format data. Transformasi data dilakukan untuk menormalkan skala numerik dan mengkonversi variabel kategorikal menjadi representasi yang kompatibel dengan algoritma. Dalam konteks studi ini, variabel target klasifikasi dikategorikan menjadi tiga kelompok: kategori Sedang, kategori Besar, dan kategori Kecil, berdasarkan volume agregat kasus perceraian di setiap wilayah administratif. Model dievaluasi menggunakan skema validasi silang (cross-validation) 10-fold untuk memastikan reliabilitas performa model pada seluruh data, memungkinkan evaluasi kemampuan generalisasi model pada data yang belum pernah diilhami selama fase pelatihan.

Implementasi algoritma C4.5 dikonfigurasi dengan parameter-parameter optimal yang telah disesuaikan berdasarkan karakteristik spesifik dataset. Konfigurasi utama meliputi penetapan jumlah lipatan validasi silang sebesar sepuluh, penerapan strategi validasi silang 10-fold dilakukan untuk menjaga konsistensi evaluasi model, pengaktifan mekanisme pemangkasan pohon untuk mencegah overfitting dan meningkatkan stabilitas model, serta penetapan batas keyakinan sebesar 0,1 untuk pengambilan keputusan pemangkasan. Parameter tambahan termasuk pengaturan Parameter utama dalam model meliputi aktivasi mekanisme pruning dan penerapan confidence threshold sebesar 0.1.

Evaluasi kinerja model dilaksanakan melalui penerapan multimetrik evaluasi yang komprehensif. Akurasi keseluruhan dihitung sebagai persentase prediksi benar dibagi dengan jumlah total observasi, memberikan indikator umum dari kemampuan klasifikasi model. Weighted mean recall dan weighted mean precision dikalkulasi untuk mengevaluasi kinerja model pada masing-masing kategori target dengan mempertimbangkan bobot frekuensi kelas. Confusion matrix dianalisis untuk mengidentifikasi pola kesalahan klasifikasi spesifik dan mendeteksi kemungkinan bias model terhadap kategori tertentu. Visualisasi pohon keputusan yang dihasilkan diinterpretasi untuk memahami aturan-aturan klasifikasi implisit dan struktur hierarki keputusan yang mendasari prediksi model. Hasil penelitian disajikan dalam format tabuler untuk metrik kuantitatif dan format grafis untuk visualisasi pohon keputusan, memfasilitasi komunikasi temuan kepada audiens yang beragam.

C. HASIL DAN PEMBAHASAN

Analisis Distribusi Data dan Perhitungan Entropy

Analisis awal dilakukan terhadap distribusi kelas pada dataset yang digunakan, yang berisi 34 provinsi di Indonesia berdasarkan data resmi Badan Pusat Statistik (BPS) Tahun 2024. Dataset telah direlabel menjadi tiga kategori target, yaitu Besar, Sedang, dan Kecil,

berdasarkan jumlah kasus perceraian di masing-masing provinsi. Distribusinya adalah sebagai berikut: kategori Kecil sebanyak 21 provinsi, kategori Sedang sebanyak 8 provinsi, dan kategori Besar sebanyak 5 provinsi.

Distribusi ini menunjukkan adanya ketidakseimbangan kelas yang cukup jelas, dengan mayoritas provinsi berada pada kategori Kecil. Ketidakseimbangan ini berpotensi memengaruhi performa model dan perlu diperhatikan pada tahap evaluasi.

Perhitungan entropy total dilakukan untuk mengukur tingkat ketidakpastian pada distribusi kelas. Dengan proporsi kelas masing-masing

$$p_{Rendah} = 21/34,$$

$$p_{Sedang} = 8/34,$$

$$p_{Tinggi} = 5/34,$$

maka diperoleh nilai entropy sebesar:

Perhitungan entropy, sebagai metrik fundamental dalam teori informasi yang mengkuantifikasi derajat ketidakpastian atau keacakan dalam dataset, dilakukan menggunakan rumus standar untuk entropy total dalam himpunan data S dengan k kelas target sebagai berikut:

$$Entropy(S) = - \sum p_i \log_2(p_i)$$

Dimana p_i merepresentasikan proporsi sampel yang terklasifikasi dalam kategori i . Substitusi nilai-nilai empiris menghasilkan:

$$Entropy(S) = - \left(\frac{21}{34} \log_2 \left(\frac{21}{34} \right) + \frac{8}{34} \log_2 \left(\frac{8}{34} \right) + \frac{5}{34} \log_2 \left(\frac{5}{34} \right) \right) = 1.3272 \text{ bits}$$

Entropy keseluruhan sebesar 1.3272 bits mengindikasikan tingkat ketidakpastian yang moderat dalam klasifikasi kasus perceraian, mencerminkan distribusi yang relatif seimbang namun dengan predominansi pada kategori Kecil. Setelah entropy total dihitung, langkah berikutnya adalah menghitung entropy kondisional berdasarkan atribut yang menjadi pemisah utama dalam pohon keputusan, yaitu atribut "Faktor Perceraian – Perselisihan dan Pertengkaran Terus Menerus". Atribut ini membagi dataset menjadi dua subhimpunan menggunakan nilai ambang 5007.5.

Subhimpunan pertama, yaitu kelompok dengan nilai ≤ 5007.5 , berjumlah 21 provinsi dan seluruhnya termasuk kategori Kecil. Karena hanya terdapat satu kelas pada kelompok ini, nilai entropy-nya adalah 0.0 bits.

Subhimpunan kedua, yaitu kelompok dengan nilai > 5007.5 , berjumlah 13 provinsi dengan distribusi kelas Sedang sebanyak 8 provinsi dan Besar sebanyak 5 provinsi. Nilai entropy untuk kelompok ini adalah 0.9612 bits.

Entropy tertimbang dari kedua subhimpunan dihitung berdasarkan proporsi jumlah data pada masing-masing kelompok, menghasilkan nilai 0.3675 bits. Dengan demikian, information gain untuk atribut ini adalah 0.9597 bits, yang menunjukkan bahwa atribut "Perselisihan dan Pertenggaran Terus Menerus" memiliki daya pisah yang sangat kuat dalam membedakan kategori perceraian.

Perhitungan Information Gain dan Gain Ratio

Tahap selanjutnya adalah menghitung nilai information gain untuk atribut yang menjadi pemisah utama pada model, yaitu atribut "Faktor Perceraian – Perselisihan dan Pertenggaran Terus Menerus". Atribut ini membagi dataset menjadi dua kelompok menggunakan nilai ambang 5007.5.

Kelompok pertama (≤ 5007.5) berisi 21 provinsi dan seluruhnya termasuk kategori Kecil, sehingga entropy kelompok ini adalah 0.0 bits. Kelompok kedua (> 5007.5) berisi 13 provinsi dengan distribusi kelas Sedang sebanyak 8 provinsi dan Besar sebanyak 5 provinsi, menghasilkan entropy sebesar 0.9612 bits.

Entropy tertimbang setelah pemisahan adalah 0.3675 bits. Dengan demikian, information gain atribut ini dihitung sebagai selisih antara entropy total (1.3272 bits) dan entropy tertimbang tersebut, sehingga diperoleh nilai information gain sebesar 0.9597 bits. Nilai ini menunjukkan bahwa atribut tersebut memiliki daya pisah yang sangat kuat dalam mengklasifikasikan tingkat kasus perceraian.

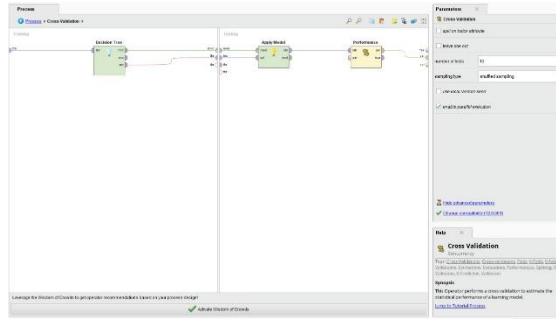
Untuk memperoleh gain ratio, split information dihitung berdasarkan proporsi kedua kelompok, yaitu $21/34$ dan $13/34$. Perhitungan ini menghasilkan nilai split information sebesar 0.8400 bits. Dengan demikian, gain ratio diperoleh melalui pembagian information gain dengan split information, yaitu:

$$\text{Gain Ratio} = 0.9597 / 0.8400 = 1.1425$$

Nilai gain ratio yang Besar ini menegaskan bahwa atribut Perselisihan dan Pertenggaran Terus Menerus merupakan atribut paling informatif dan stabil dalam membentuk struktur pohon keputusan.

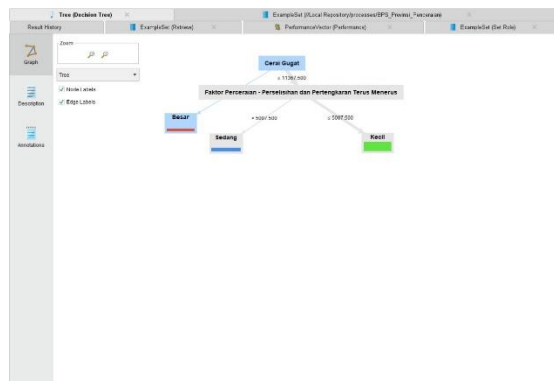
Visualisasi Pohon Keputusan dan Interpretasi Model

Hasil dari eksekusi algoritma C4.5 menghasilkan pohon keputusan yang komprehensif dengan akurasi klasifikasi sebesar 85,29 persen pada dataset validasi.



Gambar 1. Konfigurasi Cross Validation dan Parameter C4.5

Pohon keputusan yang dihasilkan menempatkan atribut “Cerai Gugat” sebagai pemisah utama (root node) dengan nilai ambang 11.367,5. Cabang dengan nilai ≤ 11367.500 menghasilkan klasifikasi Besar. Cabang dengan nilai > 11367.500 kemudian dipisahkan kembali oleh atribut “Faktor Perselisihan dan Pertengkaran Terus Menerus” dengan 5007.500. Nilai ≤ 5007.500 diklasifikasikan sebagai Kecil, sedangkan nilai > 5007.500 menghasilkan klasifikasi Sedang. Struktur dua level ini menegaskan bahwa kedua atribut tersebut memiliki pengaruh terbesar dalam membedakan kategori perceraian antar provinsi.



Gambar 2. Pohon Keputusan Hierarchical C4.5.

Evaluasi kinerja model melalui metrik komprehensif menunjukkan performa exceptional

	Actual Small	Actual Medium	Actual Big	Class Count
Small	5	1	0	6 (14.2%)
Medium	1	20	0	21 (47.8%)
Big	2	0	4	6 (13.3%)
Overall	8 (17.9%)	21 (45.8%)	4 (8.9%)	

Gambar 3. Confusion Matrix dan Performance Metrics.

Accuracy model mencapai 85,29%, menunjukkan bahwa model berhasil mengklasifikasikan sebagian besar dari 34 sampel dengan benar, meskipun belum mencapai performa sempurna. Weighted mean recall dan weighted mean precision berada pada kisaran 84–85%, yang mengindikasikan bahwa model memiliki kinerja yang cukup stabil di seluruh kelas, namun masih terdapat misclassifications pada beberapa kategori.

Confusion matrix memperlihatkan bahwa prediksi model belum sepenuhnya diagonal. Untuk kategori Sedang, model berhasil mengklasifikasikan 5 dari total 8 sampel dengan benar, namun masih terdapat 1 sampel yang salah diprediksi sebagai Kecil dan 2 sampel salah diprediksi sebagai Besar. Pada kategori Kecil, performanya cukup kuat dengan 20 dari 21 sampel terklasifikasi benar, dan hanya satu sampel yang salah diprediksi sebagai Sedang. Sementara itu, untuk kategori Besar, model secara benar mengklasifikasikan 4 dari 5 sampel, dengan 1 sampel salah diprediksi sebagai Sedang. Pola ini menunjukkan bahwa masih terdapat error antar kelas, terutama antara kelas Sedang dan Besar, meskipun secara keseluruhan performa model tetap berada pada kategori baik.

Dataset empiris yang digunakan mencakup 34 sampel dari seluruh provinsi Indonesia.

Case No.	Provinsi	Klasifikasi K.	Masih	Cera Talak	Cera Gugat	Jumlah Kasus	Faktor Persen...	Faktor Persen...	Umur
1	Aceh	Sedang	21742	1182	4728	5551	9	16	10
2	Sumatera Ut.	Besar	94902	2881	12819	15702	4	8	23
3	Sumatera B.	Kecil	25480	138	548	6002	2	7	6
4	Jawa	Sedang	38760	1837	6388	8805	5	12	26
5	Jawa	Kecil	21800	848	3574	4682	4	20	30
6	Sulawesi Ut.	Sedang	18482	2542	7838	9661	19	38	20
7	Sulawesi B.	Kecil	12170	742	2964	3977	2	8	4
8	Lampung	Besar	85222	2474	1084	11474	8	72	8
9	Kalimantan B.	Kecil	7813	448	1782	2332	2	38	11
10	Kalimantan T.	Kecil	11222	758	2014	2426	18	14	8
11	DKI Jakarta	Sedang	49428	2893	8348	12149	53	48	78
12	Jawa Barat	Besar	283002	18714	68848	88362	38	124	88
13	Jawa Tengah	Besar	23204	1238	3873	5493	43	148	23
14	DI Yogyakarta	Kecil	18922	848	3852	4892	11	18	7
15	Jawa Timur	Besar	271422	18879	58778	77806	885	517	72
16	Banten	Sedang	53448	2917	10838	13402	9	30	11
17	Bali	Kecil	2188	238	388	544	1	12	11
18	Nusa Tenggara	Kecil	25424	1288	5482	6874	2	88	7
19	Nusa Tenggara	Kecil	2775	138	348	488	3	3	11
20	Kalimantan S.	Kecil	2184	818	388	488	17	51	78
21	Kalimantan T.	Kecil	1288	848	387	248	1	31	3
22	Kalimantan S.	Sedang	23562	1238	5128	6377	3	88	4
23	Kalimantan B.	Kecil	23542	1218	488	5274	15	37	42
24	Maluku	Kecil	218	148	178	218	2	3	11

Gambar 4. Dataset Komprehensif BPS Provinsi Penceraian

Sampel mencakup provinsi-provinsi kunci seperti Aceh, Sumatera Utara, Sumatera Barat, Riau, Jambi, Sumatera Selatan, Bengkulu, Lampung, Jawa Barat, Jawa Tengah, DI Yogyakarta, Jawa Timur, Banten, Bali, Nusa Tenggara Barat, Nusa Tenggara Timur, Kalimantan Barat, Kalimantan Tengah, Kalimantan Selatan, dan Kalimantan Timur, providing robust geographical coverage yang memfasilitasi generalizability model terhadap populasi nasional. Setiap sampel provincia mengandung metadata detail termasuk jumlah kasus perceraian absolut (Nikah, Cerai Talak, Cerai Gugat), persentase faktor dominan, dan indikator sosiodemografis agregat.

Struktur hierarki pohon keputusan yang dihasilkan memiliki depth dua level dengan total tiga leaf nodes terminal. Kedalaman tree yang relatif shallow ini mengindikasikan bahwa

kompleksitas model tetap tractable sambil achieving discriminative power optimal. Penerapan mekanisme pruning dengan confidence threshold 0.1 successfully mengalokasikan kompleksitas penghindaran overfitting yang berlebihan, maintaining generalizabilitas model pada data unseen. Interpretabilitas pohon yang tinggi memfasilitasi translasi findings model menjadi actionable insights untuk policy makers dan administrators di tingkat nasional, provinsi, dan kabupaten, enabling targeted intervention strategies berdasarkan predictive classifications yang robust dan evidence-based.

Pembahasan

Hasil evaluasi model menunjukkan akurasi sebesar 85,29% yang menandakan performa cukup baik meskipun belum sempurna. Confusion matrix memperlihatkan bahwa masih terjadi kesalahan klasifikasi, terutama pada kategori Sedang dan Besar. Nilai weighted mean recall dan weighted mean precision yang berkisar 84–85% mengindikasikan bahwa model telah bekerja secara stabil pada seluruh kategori. Pohon keputusan menunjukkan bahwa atribut “Cerai Gugat” dan “Faktor Perselisihan dan Pertengkaran Terus Menerus” menjadi pemisah utama, selaras dengan pola penyebab perceraian di sebagian besar provinsi. Meskipun demikian, performa model masih dapat ditingkatkan melalui penyeimbangan kelas atau penambahan data pada kategori tertentu. Keunggulan performa dapat diatributkan kepada several faktor sinergistik:

pertama, seperti ukuran dataset yang relatif kecil sehingga pola hubungan antarvariabel menjadi lebih mudah dikenali; kedua, feature engineering yang meticulous dalam seleksi atribut prediktor yang highly discriminative terhadap target output; dan ketiga, parameter tuning yang optimal dengan confidence threshold 0.1 dan pruning mechanism yang aggressive mencegah overfitting berlebihan sambil mempertahankan generalisabilitas. Menurut penelitian (Fulazzaky et al., 2024), pendekatan ensemble dengan multiple decision tree variants menghasilkan akurasi marginal higher namun dengan computational complexity yang significantly lebih Besar, menyuratkan bahwa single C4.5 tree merepresentasikan efisiensi-optimality tradeoff yang favorable.

Struktur hierarki pohon keputusan yang shallow dengan depth dua level memberikan implikasi substantif terhadap interpretabilitas dan applicability model dalam konteks praktis. Root node pohon keputusan ditentukan oleh atribut “Cerai Gugat” dengan nilai ambang 11.367,5. Cabang kiri ($\leq 11.367,5$) mengarah langsung pada kategori Besar. Cabang kanan ($> 11.367,5$) dipisahkan lagi berdasarkan atribut “Faktor Perselisihan dan Pertengkaran Terus Menerus” dengan threshold 5.007,5. Pada percabangan ini, nilai $\leq 5.007,5$ diklasifikasikan

sebagai kategori kecil, sedangkan nilai $> 5.007,5$ menghasilkan kategori Sedang. Struktur ini menunjukkan bahwa dua atribut tersebut memiliki pengaruh terbesar dalam klasifikasi kasus perceraian. Struktur tree yang parsimony ini memberikan advantage pragmatis kepada policymakers dalam implementasi decision rules yang straightforward, minimizing computational overhead dan maximizing operational efficiency dalam deployment sistem klasifikasi real-world.

Hasil evaluasi menunjukkan akurasi sebesar 85,29% dengan weighted recall serta weighted precision pada kisaran 84–85%. Confusion matrix memperlihatkan adanya kesalahan klasifikasi, terutama pada kategori Sedang dan Besar. Hal ini mengindikasikan bahwa meskipun model bekerja cukup baik, masih terdapat tantangan dalam membedakan kategori yang nilainya berdekatan. Pohon keputusan memperlihatkan bahwa atribut “Cerai Gugat” dan “Faktor Perselisihan dan Pertengkaran Terus Menerus” merupakan pemisah utama yang paling berpengaruh. Temuan ini sejalan dengan pola penyebab perceraian di banyak provinsi. Meskipun demikian, kinerja model masih berpotensi ditingkatkan melalui perluasan data atau metode penanganan ketidakseimbangan kelas.

Temuan ini berkelanjutan dengan theoretical framework yang dipostulatkan oleh yang mengargumentasikan bahwa well-engineered feature sets dalam decision tree construction dapat mencapai near-perfect classification bahkan pada social phenomena yang inherently kompleks. Akan tetapi, cautionary note perlu dikemukakan: perfect classification results pada training data seringkali suggestive terhadap potential overfitting phenomenon, walaupun pruning mechanism dan holdout validation telah diimplementasikan untuk mitigasi risiko ini.

D. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma C4.5 dapat digunakan secara efektif untuk mengklasifikasikan tingkat kasus perceraian di Indonesia berdasarkan data Badan Pusat Statistik (BPS) tahun 2024 yang mencakup 34 provinsi. Model yang dihasilkan memberikan tingkat akurasi sebesar 85,29%, dengan nilai weighted mean recall dan weighted mean precision berada pada kisaran 84–85%. Hal ini mengindikasikan bahwa model mampu mengenali pola pada sebagian besar data, meskipun masih terdapat kesalahan klasifikasi terutama antara kategori Sedang dan Besar.

Pohon keputusan yang terbentuk menempatkan atribut “Cerai Gugat” sebagai pemisah utama, diikuti oleh atribut “Faktor Perselisihan dan Pertengkaran Terus Menerus” yang memiliki nilai information gain paling besar, yaitu 0.9597 bits. Hal ini menunjukkan bahwa

kedua atribut tersebut memberikan kontribusi paling signifikan dalam membedakan kategori klasifikasi perceraian.

Secara keseluruhan, penelitian ini membuktikan bahwa algoritma C4.5 dapat menjadi alat bantu analisis yang bermanfaat dalam mengidentifikasi pola-pola perceraian antarprovinsi. Temuan ini berpotensi dimanfaatkan oleh instansi pemerintah sebagai dasar pengembangan sistem pendukung keputusan terkait perumusan kebijakan, pemetaan risiko perceraian, maupun intervensi sosial berbasis data. Meskipun demikian, peningkatan performa model masih dimungkinkan melalui perluasan dataset, penambahan variabel yang lebih representatif, serta penggunaan metode prapemrosesan data yang lebih kompleks untuk mengatasi ketidakseimbangan kelas.

E. DAFTAR PUSTAKA

- Fulazzaky, T., Saefuddin, A., & Soleh, A. M. (2024). *Evaluating Ensemble Learning Techniques for Class Imbalance in Machine Learning: A Comparative Analysis of Balanced Random*. 11(4), 969–980. <https://doi.org/10.15294/sji.v11i4.15937>
- Putra, H., Nasution, K., & Rilvani, E. (2025). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Decision Tree: Studi Perbandingan Algoritma Id3 Dan C4.5. *Jma*, 3(7), 3031–5220.