

Klasifikasi Tingkat Kekerasan dalam Rumah Tangga Menggunakan Algoritma Decision Tree C4.5

Rizki Maulana Khair, Muhammad Ersah Setiawan, Melati Rosaensi

Universitas Bumigora, Mataram, Indonesia

Correspondence : e-mail: Melati.rn@universitasbumigora.ac.id

Abstrak

Kasus Kekerasan Dalam Rumah Tangga (KDRT) di Provinsi Nusa Tenggara Barat (NTB) terus menunjukkan tren yang mengkhawatirkan setiap tahunnya. Dampaknya tidak hanya dirasakan secara fisik dan psikis oleh korban, tetapi juga mengganggu ketahanan sosial masyarakat. Situasi ini menuntut perhatian serius dari pemerintah dan seluruh elemen masyarakat, serta dibutuhkan pendekatan berbasis data untuk mendukung kebijakan pencegahan yang lebih tepat sasaran. Penelitian ini bertujuan untuk mengklasifikasikan tingkat kasus Kekerasan Dalam Rumah Tangga (KDRT) di Provinsi Nusa Tenggara Barat (NTB) berdasarkan atribut tahun, wilayah, dan jumlah kasus. Permasalahan utama yang diangkat adalah belum tersedianya sistem klasifikasi berbasis data yang dapat mendukung kebijakan pencegahan kekerasan. Algoritma Decision Tree C4.5 digunakan untuk membangun model klasifikasi dengan data dari tahun 2020 hingga 2024, di mana jumlah kasus dikategorikan ke dalam tiga kelas: rendah, sedang, dan tinggi melalui diskretisasi manual. Model divisualisasikan dalam bentuk pohon keputusan dengan aturan klasifikasi yang mudah diinterpretasi, di mana atribut "jumlah kasus" menjadi pemisah utama. Hasil evaluasi menggunakan 10-fold cross-validation menunjukkan akurasi sebesar 94%, dengan nilai precision, recall, dan F1-score rata-rata di atas 0,94. Temuan ini menunjukkan bahwa algoritma C4.5 efektif dalam mengungkap pola distribusi kekerasan dalam rumah tangga, dan dapat dijadikan dasar penyusunan strategi intervensi serta kebijakan pencegahan yang lebih tepat sasaran di wilayah NTB.

Kata kunci: Kekerasan Dalam Rumah Tangga, klasifikasi, decision tree, C4.5, Nusa Tenggara Barat.

Abstract

Domestic Violence (DV) cases in West Nusa Tenggara (NTB) Province continue to show a worrying trend every year. The impact is not only felt physically and psychologically by the victims, but also disrupts the social resilience of the community. This situation requires serious attention from the government and all elements of society, and a data-based approach is needed to support more targeted prevention policies. This study aims to classify the level of Domestic Violence (DV) cases in West Nusa Tenggara (NTB) Province based on the attributes of year, region, and number of cases. The main problem raised is the unavailability of a data-based classification system that can support violence prevention policies. The Decision Tree C4.5 algorithm is used to build a classification model with data from 2020 to 2024, where the number of cases is categorized into three classes: low, medium, and high through manual discretization. The model is visualized in the form of a decision tree with easy-to-interpret classification rules, where the attribute "number of cases" is the main separator. The evaluation results using 10-fold cross-validation showed an accuracy of 94%, with an average precision, recall, and F1-score value above 0.94. These findings indicate that the C4.5 algorithm is effective in revealing patterns of domestic violence distribution, and can be used as a basis for formulating intervention strategies and more targeted prevention policies in the NTB region.

Keywords: Domestic Violence, classification, decision tree, C4.5, West Nusa Tenggara.

1. Pendahuluan

Kekerasan dalam rumah tangga adalah masalah sosial yang serius dan masih sering terjadi di Provinsi Nusa Tenggara Barat (NTB). Kasus kekerasan dalam rumah tangga berdampak langsung pada kondisi fisik dan psikis korban. Data dari Satu Data NTB mencatat ratusan kasus terjadi setiap tahun dari 2020 hingga 2024. Permasalahan ini tidak dapat diselesaikan dengan cara reaktif, dan membutuhkan intervensi yang terencana. Usaha untuk mengategorikan wilayah yang rentan dapat membantu dalam merancang strategi pencegahan yang efektif. Jadi, untuk membuat kebijakan berbasis bukti, metode analitik berbasis data diperlukan. Banyak orang telah menerapkan metode ini dalam penelitian mereka [1], [2].

Pemerintah selama ini belum memiliki sistem klasifikasi wilayah rawan KDRT yang akurat dan otomatis. Akibatnya, kebijakan yang diimplementasikan sering kali tidak mencerminkan situasi di lapangan. Penggunaan metode klasifikasi seperti decision tree dapat menjadi solusi untuk menyusun peta risiko. Algoritma C4.5 adalah salah satu yang paling sering digunakan dalam klasifikasi sosial. Metode ini dapat mengolah data numerik dan kategorikal sekaligus menghasilkan model yang mudah dipahami. C4.5 juga mampu memilih atribut yang paling berpengaruh dalam proses klasifikasi. Hasil akhir dapat divisualisasikan secara jelas sebagai pohon keputusan [3], [4].

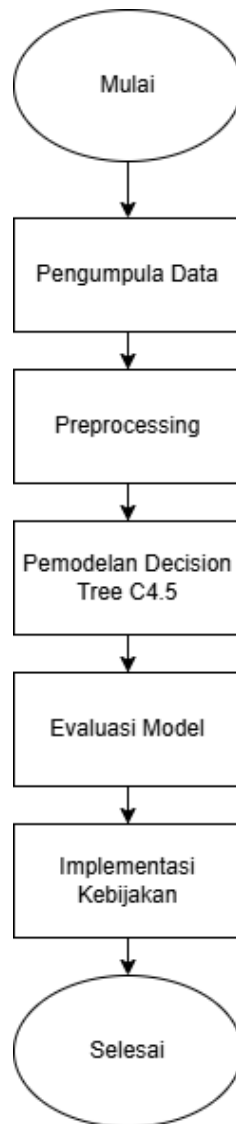
Studi oleh Eshardiansyah memanfaatkan algoritma C4.5 untuk mengklasifikasikan kekerasan terhadap anak di Karawang dan menunjukkan hasil akurasi yang memuaskan [5]. Fanani dan Katul menggunakan algoritma ini untuk mengategorikan jenis kekerasan dengan hasil yang optimal [6]. Algoritma C4.5 terbukti efektif untuk data sosial karena pohon yang dihasilkan mudah untuk dianalisis. Dalam penelitian ini, data digunakan mencakup tahun, wilayah, dan jumlah kasus yang telah dikategorikan menjadi tiga kelas: rendah, sedang, dan tinggi. Proses klasifikasi dilakukan menggunakan perangkat lunak WEKA. Hasil klasifikasi selanjutnya ditampilkan dalam bentuk pohon Keputusan. Studi lain menunjukkan bahwa pendekatan klasifikasi lebih unggul dibanding metode prediksi murni. Subroto Dkk. menggunakan Naive Bayes untuk klasifikasi kekerasan, namun akurasinya rendah [7]. Sementara itu, Yulia menggunakan regresi linear untuk prediksi kasus, namun tidak cocok untuk klasifikasi wilayah [3]. C4.5 dinilai lebih sesuai digunakan dalam konteks seperti KDRT karena kemampuannya dalam menyegmentasi data berdasarkan atribut-atribut penting. Proses klasifikasi dilakukan otomatis dan hasilnya dievaluasi secara akurasi dan interpretabilitas.

Data diperoleh dari portal Satu Data NTB mencakup jumlah kasus KDRT dari tahun 2020 sampai 2024, yang telah dikonversi ke bentuk kategorikal. Kategori dibentuk menjadi tiga kelas yaitu rendah, sedang, dan tinggi sesuai dengan distribusi data. Algoritma C4.5 digunakan dengan pendekatan entropy dan gain ratio untuk menentukan atribut paling berpengaruh. Atribut “jumlah kasus” tercatat memberikan nilai gain tertinggi dan dijadikan simpul utama dalam pohon klasifikasi [5], [6]. Penelitian ini memberikan kontribusi dalam menyajikan model klasifikasi wilayah rentan KDRT yang tepat dan praktis. Informasi yang dihasilkan dapat menjadi dasar pengambilan kebijakan yang lebih tepat sasaran. Simanjuntak juga menunjukkan penerapan C4.5 dalam proses hukum terhadap pelaku kekerasan [8]. Studi lain menunjukkan bahwa teknik klasifikasi mendukung pembuat kebijakan dalam memahami pola sebaran kasus [9]. Model klasifikasi ini dapat diperbarui secara berkala dan digunakan sebagai alat evaluasi tahunan oleh instansi terkait.

Tujuan dari penelitian ini adalah menghasilkan model klasifikasi berbasis algoritma C4.5 untuk mengelompokkan wilayah berdasarkan jumlah kasus Kekerasan Dalam Rumah Tangga. Model ini dirancang agar mendukung pengambilan kebijakan pencegahan yang lebih terukur [10]. Dengan menggunakan pendekatan klasifikasi berbasis data, penelitian ini diharapkan membantu dalam mengidentifikasi wilayah berisiko secara lebih tepat. Hasil model ditampilkan dalam bentuk visualisasi pohon keputusan yang dapat dimanfaatkan oleh pembuat kebijakan untuk perencanaan jangka panjang [6], [8].

2. Metode Penelitian

Melalui pendekatan kuantitatif, penelitian ini mengklasifikasikan jumlah kasus Kekerasan Dalam Rumah Tangga di Provinsi Nusa Tenggara Barat (NTB) ke dalam tiga kategori: Rendah, Sedang, dan Tinggi. Penelitian ini bertujuan untuk mengembangkan model klasifikasi yang mampu mengenali pola wilayah dan tahun yang berkaitan dengan tingkat kekerasan yang terjadi [11], [12]. Dataset yang digunakan diambil dari portal Satu Data NTB dan mencakup informasi dari tahun 2020 sampai 2024. Data itu dikelompokkan berdasarkan tiga atribut utama: tahun, wilayah, dan jumlah kasus. Kategori ditentukan berdasarkan ambang manual, dengan batasan: ≤ 10 (Rendah), 11–30 (Sedang), dan > 30 (Tinggi), sesuai distribusi data dan kebutuhan klasifikasi. Penelitian ini dilaksanakan melalui langkah-langkah berikut [13].



Gambar 1. Tahapan Penelitian Decision Tree C4.5

2.1 Pengumpulan Data

Data diperoleh dari situs <https://satudata.ntbprov.go.id>, yang menyediakan statistik sosial resmi wilayah NTB untuk periode 2020–2024. Tiga atribut utama digunakan: tahun, wilayah, dan jumlah kasus Kekerasan Dalam Rumah Tangga (KDRT)[12].

2.2 Preprocessing

Tahap preprocessing mencakup proses pembersihan data (data cleaning), seleksi data (data sselection), serta encoding pada atribut kategorikal seperti nama wilayah dan tahun agar dapat dibaca oleh algoritma klasifikasi. Selain itu, dilakukan juga proses diskretisasi terhadap atribut “jumlah kasus”. [14].

2.3 Decision Tree C4.5

Pada tahap ini, algoritma C4.5 diterapkan untuk membentuk model klasifikasi berdasarkan data yang telah melalui proses preprocessing. Algoritma ini digunakan untuk menghasilkan pohon keputusan dengan memilih atribut yang memiliki nilai gain ratio tertinggi sebagai simpul akar (root node). Proses kemudian dilanjutkan dengan membentuk cabang berdasarkan nilai-nilai unik dari atribut tersebut, dan membagi dataset ke dalam masing-masing cabang. Tahapan ini dilakukan secara rekursif hingga seluruh data terklasifikasi ke dalam simpul dengan satu kelas akhir. Dalam penelitian ini, atribut “jumlah kasus” menghasilkan gain ratio tertinggi dan secara otomatis dipilih sebagai node awal. Model yang terbentuk divisualisasikan dalam bentuk pohon keputusan, yang selanjutnya digunakan untuk membentuk aturan klasifikasi (if-then rules). Evaluasi terhadap hasil model dilakukan dengan metode 10-fold cross-validation

untuk mengukur akurasi dan ketepatan prediksi klasifikasi [5], [6]. Prosedur umum dalam pembentukan decision tree menggunakan C4.5 adalah sebagai berikut:

1. Memilih atribut dengan nilai gain ratio tertinggi sebagai node akar.
2. Membuat cabang berdasarkan nilai-nilai unik dari atribut tersebut.
3. Membagi dataset ke dalam cabang-cabang tersebut.
4. Mengulangi proses secara rekursif hingga semua data dalam cabang termasuk ke dalam satu kelas.

Kemampuan visualisasi serta struktur aturan yang jelas menjadikan algoritma ini relevan dalam studi klasifikasi data sosial seperti pada kasus KDRT di NTB [11]. Berikut rumus yang digunakan untuk menentukan akar pada pohon keputusan :

$$Entropy(s) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (1)$$

Dalam perhitungan *entropy* pada algoritma C4.5, simbol S merepresentasikan himpunan seluruh data yang sedang dianalisis, sedangkan n menunjukkan jumlah total kelas dalam data tersebut. Nilai p_i adalah proporsi data dari kelas ke- i terhadap total data, dan digunakan sebagai komponen utama dalam menghitung tingkat ketidakpastian (*entropy*) dari himpunan S .

$$Gain(S, A) = Entropy(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot Entropy(S_v) \quad (2)$$

Perhitungan *information gain* bertujuan untuk mengetahui seberapa besar pengurangan ketidakpastian saat data dibagi berdasarkan suatu atribut. Simbol A menyatakan atribut yang sedang dievaluasi, S adalah himpunan data, dan $|S|$ adalah jumlah total data. Notasi $|S_v|$ menunjukkan jumlah data dalam subset S yang memiliki nilai tertentu v pada atribut A . Semakin besar pengurangan entropy setelah pemisahan, semakin besar nilai gain yang dihasilkan.

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

$$SplitInfo(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \left(\frac{|S_i|}{|S|} \right) \quad (3)$$

Gain Ratio dihitung dengan membagi nilai *Gain* terhadap *SplitInfo*. *SplitInfo* mengukur seberapa merata atribut A membagi data ke dalam beberapa subset. Simbol $|S_i|$ menyatakan jumlah data dalam subset untuk nilai atribut ke- i , sedangkan $|S|$ tetap merepresentasikan total data. Nilai *SplitInfo* yang tinggi menandakan atribut membagi data ke banyak subset, dan digunakan untuk menormalkan nilai *Gain*. Dengan pendekatan ini, atribut yang hanya tampak dominan karena memiliki banyak nilai unik tidak secara otomatis dipilih sebagai pemisah utama.

3. Hasil dan Pembahasan

Bagian ini membahas hasil dari penelitian klasifikasi KDRT di NTB menggunakan algoritma Decision Tree C4.5. Dataset yang digunakan terdiri dari tiga atribut utama: Tahun, Wilayah, dan Jumlah kasus KDRT. Atribut Kategori digunakan sebagai target klasifikasi berdasarkan nilai dari atribut Jumlah.

Tabel 1. Atribut Yang Digunakan

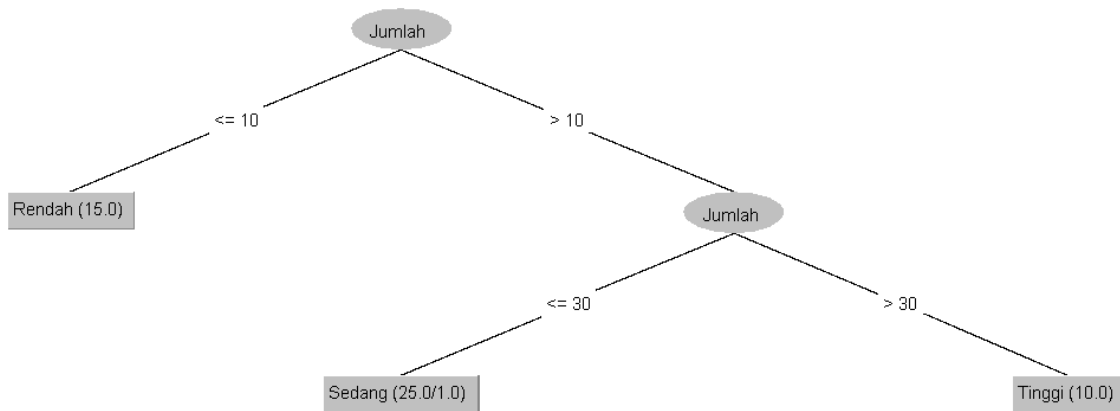
No	Atribut	Tipe
1	Tahun	Numerik
2	Wilayah	Nominal
3	Jumlah	Numerik
4	Kategori	Target

Pada Tabel 1, Kategori merupakan atribut target yang ditentukan berdasarkan nilai dari atribut "Jumlah". Atribut Tahun dan Wilayah berfungsi sebagai atribut pendukung untuk analisis spasial dan temporal. Nilai-nilai yang hilang tidak ditemukan dalam dataset, sehingga tidak dilakukan proses imputasi.

Tabel 2. Tampilan data awal

Tahun	Wilayah	Jumlah	Kategori
2020	Kabupaten Bima	19	Sedang
2020	Kabupaten Lombok Timur	42	Tinggi
2020	Kabupaten Sumbawa Barat	6	Rendah

Pada Tabel 2, Menunjukkan contoh data awal yang digunakan dalam penelitian. Ini mencakup tahun kejadian, wilayah laporan, tipe data jumlah kasus (numerik atau nominal), dan kategori tingkat KDRT (Rendah, Sedang, dan Tinggi). Selanjutnya, algoritma Decision Tree C4.5 digunakan untuk memproses data untuk digunakan dalam klasifikasi [8]. Setelah pengolahan dan analisis data awal selesai, langkah berikutnya adalah pemodelan data menggunakan metode klasifikasi dalam perangkat lunak WEKA.



Gambar 2. Struktur Pohon Keputusan Hasil Klasifikasi Kategori KDRT Berdasarkan jumlah

Pohon keputusan yang dibentuk menghasilkan tiga aturan (rule) utama dalam mengklasifikasikan jumlah kasus Kekerasan Dalam Rumah Tangga (KDRT) berdasarkan kategorinya. Setiap aturan menggambarkan hubungan antara jumlah kasus dengan tingkat klasifikasi kekerasan dalam rumah tangga yang terjadi. Adapun penjelasan dari masing-masing rule adalah sebagai berikut.

Tabel 3. Rule decision tree

No.	Rule
1.	IF Wilayah Jumlah ≤ 10 THEN Kategori KDRT = Rendah
2.	IF Wilayah $10 < \text{Jumlah} \leq 30$ THEN Kategori KDRT = Sedang
3.	IF Wilayah Jumlah > 30 THEN Kategori KDRT = Tinggi

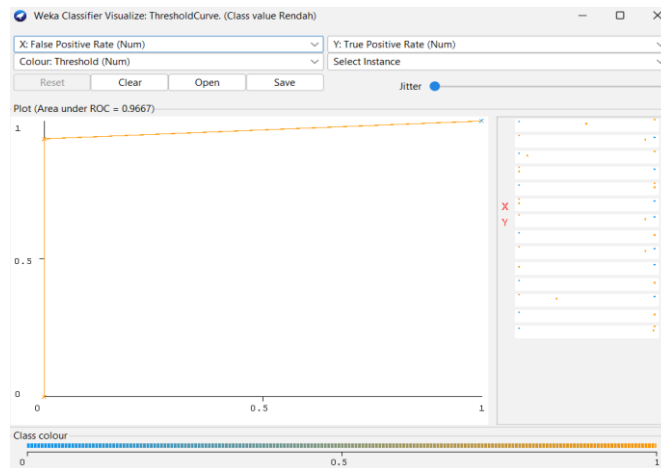
Dalam Tabel 3, aturan-aturan ini menunjukkan bahwa jumlah kasus KDRT berpengaruh signifikan terhadap klasifikasi tingkat kekerasan. Selain itu, model ini menunjukkan akurasi yang tinggi pada tiap node akhir. Dengan demikian, model ini dapat berfungsi sebagai dasar sistem pendukung keputusan untuk menangani atau meramalkan wilayah yang rawan terhadap KDRT. Evaluasi model klasifikasi menggunakan metrik per kelas menunjukkan bahwa model memiliki kinerja yang sangat unggul dalam mengenali setiap kategori tingkat KDRT. Evaluasi dilakukan menggunakan metode 10-fold cross-validation, yang membagi data menjadi sepuluh subset seimbang berdasarkan distribusi kelas [14]. Setiap subset secara bergiliran digunakan sebagai data uji, sementara sisanya sebagai data latih. Metode ini dipilih

karena menghasilkan evaluasi yang stabil dan mengurangi risiko overfitting, terutama saat jumlah data relatif terbatas [15].

Tabel 4. *Detailed Accuracy By Class*

Kelas	Precision	Recall	F1-Score	TP Rate	FP Rate	ROC Area	PRC Area
Tinggi	0.909	0.909	0.909	0.909	0.026	0.917	0.846
Sedang	0.920	0.958	0.939	0.958	0.077	0.921	0.862
Rendah	1.000	0.933	0.966	0.933	0.000	0.967	0.953
Rata-rata Tertimbang	0.942	0.940	0.940	0.940	0.043	0.934	0.886

Tabel 4 menampilkan hasil evaluasi performa model klasifikasi untuk masing-masing kategori tingkat Kekerasan Dalam Rumah Tangga, yaitu Tinggi, Sedang, dan Rendah. Berdasarkan hasil evaluasi, kelas "Rendah" memiliki presisi sempurna 1.000, yang menunjukkan bahwa semua perkiraan untuk kategori ini akurat tanpa kesalahan. Recall tertinggi terdapat pada kelas "Sedang" dengan nilai 0.958, yang berarti sebagian besar data kelas Sedang berhasil dikenali dengan baik oleh model. F1-score untuk ketiga kelas juga menunjukkan nilai yang tinggi dan merata, menunjukkan bahwa model memiliki keseimbangan yang baik antara akurasi dan kemampuan dalam mengenali data. Selain itu, nilai ROC Area untuk seluruh kelas berada di atas 0.91, menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam membedakan antar kelas. FP Rate yang rendah, terutama pada kelas Rendah yang bernilai 0.000, menandakan bahwa hampir tidak ada kesalahan klasifikasi terhadap kelas lain. Nilai rata-rata tertimbang untuk precision, recall, dan F1-score masing-masing adalah 0.942, 0.940, dan 0.940. Secara keseluruhan, hasil ini menunjukkan bahwa model klasifikasi Decision Tree C4.5 mampu mengenali dan memisahkan kategori KDRT dengan sangat baik dan konsisten [12].



Gambar 3. Kurva ROC untuk Kelas "Rendah"

Gambar 3 menunjukkan kurva ROC untuk kelas "Rendah" dengan nilai AUC sebesar **0.9667**. Kurva ini menunjukkan kemampuan model yang sangat baik dalam membedakan kelas "Rendah" terhadap kelas lainnya. Hal ini juga didukung oleh nilai precision sempurna 1.000 dan recall 0.933 pada kelas tersebut. Semakin dekat kurva ke sudut kiri atas grafik, semakin baik kinerja model dalam klasifikasi. Dengan nilai AUC mendekati 1.0, model ini menunjukkan stabilitas dan performa tinggi dalam mengenali pola data kelas rendah secara konsisten.

Tabel 5. *confusion matrix (matriks klasifikasi)*

Confusion matrix		Target		
		Tinggi	Sedang	Rendah
Model	Tinggi	10	1	0
	Sedang	1	23	0
	Rendah	0	1	14

Model ini sangat andal dan akurat untuk mengklasifikasikan kategori tingkat kekerasan dalam rumah tangga berdasarkan jumlah kasus, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan atau kebijakan intervensi sosial di wilayah NTB. Hasil penelitian menunjukkan bahwa model klasifikasi yang dibangun menggunakan algoritma C4.5 sangat akurat, skor precision, recall, dan F1 rata-rata lebih dari 0.94. Hasil ini sejalan dengan penelitian yang dilakukan oleh peneliti terdahulu, yang menggunakan C4.5 untuk mengklasifikasikan jenis kekerasan terhadap anak dan menemukan bahwa modelnya berhasil dengan skenario data 70:30 [5]. Dengan performa model yang sangat baik, hasil klasifikasi ini dapat dijadikan dasar dalam penyusunan strategi pencegahan KDRT yang lebih tepat sasaran. Model ini memiliki potensi sebagai sistem pendukung keputusan untuk membantu implementasi kebijakan oleh instansi terkait. Selain menerapkan validasi 10-fold, penelitian ini juga menguji kinerja model dengan variasi 3-fold dan 5-fold cross-validation. Hasil dari ketiga metode tersebut menunjukkan tingkat akurasi yang stabil, yaitu 94%, dengan rata-rata precision, recall, dan F1-score di atas 0,94. Walaupun ada perbedaan nilai antara fold yang tergolong kecil, validasi 10-fold dipilih sebagai konfigurasi paling efektif karena memberikan evaluasi yang paling konsisten, proporsi pelatihan yang ideal (90% latih, 10% uji), serta merupakan metode yang paling sering digunakan dalam penelitian klasifikasi. Hal ini menegaskan pangakuan bahwa model yang dikembangkan memiliki kinerja yang unggul, stabil, dan cocok untuk mendukung pengambilan keputusan yang berbasis data.

4. Kesimpulan

Penelitian ini berhasil menciptakan model klasifikasi tingkat Kekerasan Dalam Rumah Tangga (KDRT) menggunakan algoritma Decision Tree C4.5, dengan mengelompokkan jumlah kasus ke dalam tiga kategori: rendah, sedang, dan tinggi. Data yang dimanfaatkan meliputi atribut tahun, wilayah, dan jumlah kasus dari tahun 2020 hingga 2024 yang diperoleh melalui portal Satu Data NTB. Proses klasifikasi dilakukan melalui langkah-langkah preprocessing, diskretisasi secara manual, dan pelatihan model dengan menggunakan algoritma C4.5. Hasil klasifikasi menunjukkan bahwa atribut “jumlah kasus” merupakan faktor terpenting dalam pembentukan pohon keputusan, yang menghasilkan tiga aturan klasifikasi utama. Penilaian model menggunakan metode 10-fold cross-validation menunjukkan hasil klasifikasi yang sangat memuaskan, dengan rata-rata precision, recall, dan F1-score melebihi 0,94. Dengan demikian, model yang dibangun dapat dimanfaatkan sebagai dasar sistem pendukung keputusan dalam merancang strategi intervensi dan kebijakan pencegahan kekerasan yang lebih tepat sasaran. Penelitian ini juga menegaskan bahwa algoritma C4.5 efektif diterapkan pada data sosial untuk analisis klasifikasi dan perencanaan yang berbasis data.

Daftar Pustaka

- [1] L. Awaliyah, N. Rahaningsih, and R. Danar Dana, “Implementasi Algoritma K-Means Dalam Analisis Cluster Korban Kekerasan Di Provinsi Jawa Barat,” *JATI (Jurnal Mhs. Tek. Inform., vol. 8, no. 1, pp. 188–195, 2024, doi: 10.36040/jati.v8i1.8332.*
- [2] J. M. Polgan *et al.*, “Klasterisasi Wilayah Rawan Kekerasan Anak Menggunakan Algoritma k-Medoids di Kota Padang,” vol. 13, pp. 2309–2319, 2025, [Online]. Available: <https://doi.org/10.33395/jmp.v13i2.14447>
- [3] Y. Yulia *et al.*, “Optimalisasi Parameter Feature Selection Pada Prediksi Kasus,” vol. 7, no. 6, pp. 3708–3713, 2023, [Online]. Available: <https://ejournal.itn.ac.id/plugins/generic/pdfJsViewer/pdf.js/web/viewer.html?file=https%3A%2F%2Fejournal.itn.ac.id%2Findex.php%2Fjati%2Farticle%2Fdownload%2F8285%2F4890%2F>
- [4] A. Klasifikasi *et al.*, “Algoritma Nearest Neighbour 373,” vol. 12, no. 2, 2021, [Online]. Available: <https://doi.org/10.33859/dksm.v12i2.743>
- [5] R. Putra Eshardiansyah, N. Sulistiyowati, M. Jajuli, T. Informatika, U. H. Singaperbangsa Karawang Jl Ronggowaluyo Kel Puseurjaya Kec Telukjambe Timur Kab Karawang, and P. Jawa, “Algoritma C4.5 Untuk Klasifikasi Jenis Kekerasan pada Anak (Kasus DP3A Kabupaten Karawang),” *J. Sains Komput. Inform. (J-SAKTI, vol. 5, no. September, pp. 687–696, 2021, [Online]. Available: https://tunasbangsa.ac.id/ejurnal/index.php/jsakti/article/view/368*
- [6] M. Fanani Rudi and F. Fikriah Katul, “Seleksi Fitur PSO untuk Klasifikasi Jenis Kekerasan dengan Algoritma C4.5,” *Smart Comp Jurnalnya Orang Pint. Komput., vol. 12, no. 1, 2023, doi: 10.30591/smartcomp.v12i1.4407.*
- [7] G. Subroto, N. Sulistiyowati, and A. A. Ridha, “Klasifikasi Jenis Kekerasan Pada Perempuan Dan Anak Dengan Algoritma Multinomial Naïve Bayes,” *INTECOMS J. Inf. Technol. Comput. Sci., vol. 5, no. 1, pp. 104–113, 2022, doi: 10.31539/intecomsv5i1.3598.*

- [8] L. P. Simanjuntak, "Penerapan Data Mining Dalam Proses Hukum Pidana Bagi Pelaku Kekerasan Pada Wanita Menggunakan Algoritma C4. 5," *ADA J. Inf. Syst. ...*, vol. 1, no. 2, pp. 79–92, 2024, [Online]. Available: <https://journals.adaresearch.or.id/adajisr/article/view/41>
- [9] I. Ayu Made Supartini, I. Komang Gde Sukarsa, and I. Gusti Ayu Made Srinadi, "Analisis Diskriminan Pada Klasifikasi Desa Di Kabupaten Tabanan Menggunakan Metode K-Fold Cross Validation," *E-Jurnal Mat.*, vol. 6, no. 2, pp. 106–115, 2017, [Online]. Available: <https://doi.org/10.24843/MTK.2017.v06.i02.p154>
- [10] M. A. Sundari, R. Pane, and R. Rohani, "Data Mining Clustering Korban Kejahatan Pelecehan Seksual dengan Kekerasan Berdasarkan Provinsi Menggunakan Metode AHC," *Build. Informatics, Technol. Sci.*, vol. 5, no. 1, 2023, doi: 10.47065/bits.v5i1.3499.
- [11] F. Ihsan, I. Iskandar, N. S. Harahap, and S. Agustian, "Decision tree algorithm for multi-label hate speech and abusive language detection in Indonesian Twitter," *J. Teknol. dan Sist. Komput.*, vol. 9, no. 4, pp. 199–204, 2021, doi: 10.14710/jtsiskom.2021.13907.
- [12] U. M. Pratiwi and M. Ibad, "Klasifikasi Faktor Yang Berpengaruh Dalam Kehamilan Tidak Diinginkan Menggunakan Metode Algoritma Decision Tree," *J. Lebesgue J. Ilm. Pendidik. Mat. Mat. dan Stat.*, vol. 3, no. 2, pp. 406–416, 2022, doi: 10.46306/lb.v3i2.129.
- [13] M. H. Lubis and S. Sumijan, "Prediksi Tingkat Kriminalitas Menggunakan Metode Single Moving Average (Studi Kasus Polres Asahan Sumatera Utara)," *J. Sistim Inf. dan Teknol.*, vol. 3, pp. 183–188, 2021, doi: 10.37034/jsisfotek.v3i4.63.
- [14] A. E. Wibowo, A. Khairi, H. Humairoh, M. I. Fadillah, and M. J. D. Hartawan, "Text Mining: Sistem Prediksi Cyberbullying pada Platform Twitter menggunakan Logistic Regression, KNN, dan Naive Bayes," *J. Rekayasa Elektro Sriwij.*, vol. 4, no. 1, pp. 17–23, 2023, doi: 10.36706/jres.v4i1.56.
- [15] A. D. W. Sumari, R. Y. A. Pratama, and O. D. Triswidrananta, "Sistem Prediksi Tingkat Kriminalitas Menggunakan Metode Triple Exponential Smoothing: Studi Kasus Pada Polres Kabupaten Probolinggo," *J. Tek. Inform.*, vol. 13, no. 2, pp. 171–178, 2021, doi: 10.15408/jti.v13i2.18128.