

# Seleksi Fitur PSO untuk Klasifikasi Jenis Kekerasan dengan Algoritma C4.5

**M. Rudi Fanani<sup>1</sup>, Fari Katul Fikriah<sup>2</sup>, Umi Azizah<sup>3</sup>, M. Ghufron Aziz<sup>4</sup>.**

<sup>1,2,3,4</sup> Teknologi Informasi Fakultas Sains dan Teknologi ITSNU Pekalongan  
e-mail: \*<sup>1</sup>idurinanaf@gmail.com, <sup>2</sup>farichatulfikriyah45@gmail.com,

## **Abstrak**

*Kekerasan adalah tindakan atau ancaman terhadap individu, diri sendiri, atau masyarakat yang memungkinkan bisa menyebabkan kematian, kerugian, trauma, atau perampasan hak. Kaum yang rentan akan kejahatan adalah perempuan, anak dan kelompok kebutuhan khusus yang perlu untuk dilindungi. Anak memiliki peran yang strategis dan secara tegas dinyatakan bahwa negara menjamin hak setiap anak atas tumbuh, kelangsungan hidup, dan berkembang serta atas perlindungan dari deskriminasi dan kekerasan, oleh karena itu kepentingan terbaik bagi anak patut dihayati sebagai kepentingan terbaik bagi kelangsungan hidup umat manusia. Pada penelitian ini kami menggunakan algoritma C4.5 yang merupakan salah satu Teknik klasifikasi Data Mining yang digunakan untuk membentuk pohon keputusan (Decision Tree) yang berguna untuk mengeksplorasi data, menemukan hubungan tersembunyi antara sejumlah calon variabel input dengan variabel target. Tujuan dari penggunaan Algoritma C4.5 yaitu untuk mengetahui tingkat akurasi prediksi dari jenis kekerasan terhadap perempuan, anak dan kelompok kebutuhan khusus. Algoritma C4.5 masih mempunyai kelemahan yaitu dalam tumpang tindih data yang membuat data tidak relevan sehingga dapat mengurangi tingkat akurasi, untuk menangani hal tersebut dibutuhkan seleksi fitur yang dapat mengidentifikasi fitur yang relevan sehingga dapat meningkatkan akurasi dan lebih optimal.*

**Kata Kunci**— Data Mining, Klasifikasi, C4.5, PSO, Kekerasan.

## 1. PENDAHULUAN

Fenomena kekerasan terhadap perempuan, anak dan kebutuhan khusus merupakan fenomena universal dan global. Diperkirakan perempuan, anak dan kebutuhan khusus di dunia mengalami tindak kekerasan. Fenomena kekerasan terhadap perempuan, anak dan kebutuhan khusus telah mendorong minat para akademisi untuk melakukan penelitian tentang kekerasan tersebut [1]. Pada tahun 2018 kekerasan seksual 10 kasus tahun 2019 26 kasus dan tahun 2020 20 kasus, kekerasan fisik tahun 2018 10 kasus, tahun 2019 7 kasus dan 2020 6 kasus [2]. Kekerasan pada anak dan perempuan di Kabupaten Batang sendiri sebetulnya sudah ada Peraturan Daerah (Perda) untuk melindungi hak-hak anak yaitu Perda Nomor 11 tahun 2019. Pemerintah Kabupaten Batang meskipun dengan anggaran terbatas selalu mensosialisasikan Perda tersebut ke desa-desa dengan lisan meskipun belum bisa secara cetak [3].

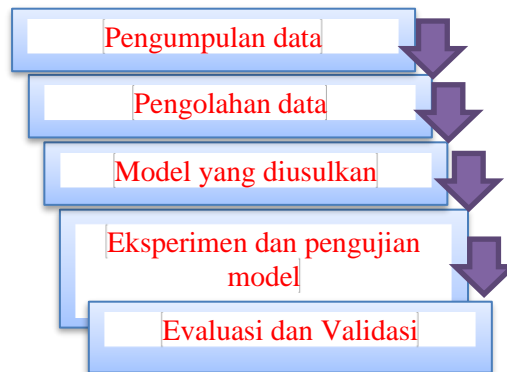
Sedikitnya laporan yang diberikan oleh para korban membuat kejadian kekerasan tersebut kerap terulang kembali sehingga masih sulit untuk meminimalisir dan menanggulangi kejadian serupa. Dengan menggunakan metode Klasifikasi Data Mining berupa Algoritma C4.5 menghasilkan beberapa aturan-aturan dan sebuah pohon keputusan dengan tujuan untuk meningkatkan keakuratan dari prediksi yang sedang dilakukan [4]. Penambahan Metode *Particle Swarm Optimization* (PSO) digunakan untuk menyeleksi atribut sehingga dapat meningkatkan hasil akurasi [5].

Peneliti akan fokus dengan data di wilayah Kecamatan Tulis yang berada di wilayah Kabupaten Batang Jawa Tengah Menurut Data BPS jumlah penduduk kecamatan Tulis 36.336 jiwa, terbagi menjadi 17 Desa/ Kelurahan antara lain: Desa Tulis, Desa Wringingintung, Desa

Posong, Desa Sembojo, Desa Beji, Desa Kaliboyo, Desa Simbangdesa, Desa Simbangjati, Desa Ponowareng, Desa Kenconorejo, Desa Kedungsegog, Desa Manggis, Desa Jolosekti, Desa Siberuk, Desa Kebumen, Desa Cluwuk dan Desa Jrahahpayung [6].

## 2. METODE PENELITIAN

Metode yang digunakan dalam penelitian yaitu: penelitian tindakan, eksperimen, studi kasus dan survei. Penelitian ini dilakukan dengan tahapan sebagaimana dalam Gambar 1.



Gambar 1. Metode Penelitian

### 2.1. Pengumpulan Data

Pengumpulan data primer pada penelitian ini menggunakan metode wawancara, survey dan observasi secara langsung kepada pihak-pihak terkait sejumlah 17 Desa/ Kelurahan Kec. Tulis Kab. Batang Jawa Tengah. Ada beberapa variabel data yang disajikan antara lain: no, desa, tempat kejadian, jenis kekerasan, usia korban, pendidikan korban, jenis kelamin pelaku, hubungan pelaku dan bentuk layanan.

### 2.2. Pengolahan Data

Pengolahan data yang dihasilkan dari data yang ada variabel diambil sebagai atribut/ variabel dalam pembentukan pohon keputusan adalah: tempat kejadian, jenis kekerasan, usia korban, pendidikan korban, jenis kelamin pelaku, dan hubungan pelaku dengan label bentuk layanan.

### 2.3. Model Yang Diusulkan

Model yang diusulkan dalam penelitian ini adalah fitur seleksi *Particle Swarm Optimization* dengan Algoritma C4.5 model ini mengolah dataset dan mengukur tingkat akurasi dalam prediksi yang digunakan untuk daya kekerasan perempuan, anak dan kelompok kebutuhan khusus di kecamatan Tulis Kabupaten Batang Jawa Tengah.

### 2.4. Algoritma C4.5

Algoritma C4.5 merupakan salah satu algoritma yang digunakan untuk melakukan klasifikasi atau segmentasi atau pengelompokan dan bersifat prediktif. Dasar algoritma adalah

pembentukan pohon keputusan (*decision tree*). Cabang-cabang pohon keputusan merupakan pertanyaan klasifikasi dan daun-daunnya merupakan kelas-kelas atau segmen-segmennya.

Perhitungan Entropy

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i * \log_2 p_i$$

Keterangan:

S	= Himpunan dataset (kasus)
n	= Banyaknya partisi S
i	= jumlah partisi S
$p_i$	= proporsi dari $S_i$ terhadap S = $\frac{S_i}{S} \left( \frac{\text{Himpunan (yes)}}{\text{Total Kasus}} \right)$
$S_i$	= Partisi S ke-i

Perhitungan Gain:

$$\text{Gain (S, A)} = \text{Entropy (S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy (S}_i)$$

Keterangan:

S	= Himpunan kasus
A	= Atribut
n	= Jumlah partisi atribut A
$ S_i $	= Jumlah kasus pada partisi ke-i
$ S $	= Jumlah kasus S/ Himpunan kasus
Entropy (S <sub>i</sub> )	= Entropy untuk kasus-kasus pada partisi ke-i

## 2.5. Particle Swarm Optimization

*Particle Swarm Optimization* (PSO) merupakan algoritma optimasi berbasis populasi yang memanfaatkan individu dalam pencarian yang digunakan untuk menentukan parameter proses guna menghasilkan nilai respon yang optimum

Rumus PSO:

$$v_j(i) = v_j(i-1) + c_1 r_1 (P_{best,j} - X_j(i-1)) + c_2 r_2 (G_{best} - X_j(i-1))$$

$$x_j(i) = v_j(i) + x_j(i-1)$$

j	= 1,2,...,N representasi jumlah artikel
$P_{best,j}$	= $P_{best,1}, P_{best,2}, \dots$ , representasi personal best dari partikel ke-j
$G_{best}$	= representasi global best dari seluruh kawan
$c_1, c_2$	= learning factor
$r_1, r_2$	= konstanta berupa bilangan random yang bernilai antara 0 sampai 1

## 2.6. Eksperimen dan Pengujian Model

Eksperimen dan pengujian model akan diproses menggunakan metode yang diusulkan dengan software Rapidminer [7]. Algoritma yang digunakan adalah C4.5 selanjutnya akan diterapkan fitur seleksi menggunakan *Particle Swarm Optimization* untuk menghasilkan nilai bobot yang optimal pada setiap atribut, sehingga akan meningkatkan nilai akurasi dari dataset tersebut [8]. Model ini akan menampilkan nilai akurasi dengan tampilan *Confusion Matrix* serta diukur menggunakan *Area Under Curve* [9].

Tabel 1. *Confusion Matrix*

<i>Total Record</i>	<i>Predicted Positive</i>	<i>Predicted Negative</i>
<i>Actual Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Actual Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan:

- TP = Banyaknya data yang berhasil diprediksi dengan benar sebagai data aktual positif  
 FP = Banyaknya data yang gagal diprediksi dengan benar sebagai data aktual positif dimana sebenarnya merupakan aktual negatif  
 TN = Banyaknya data yang berhasil diprediksi dengan benar sebagai data aktual negatif  
 FN = Banyaknya data yang gagal diprediksi dengan benar sebagai data aktual negatif dimana sebenarnya merupakan aktual positif

### 2.7. Evaluasi dan Validasi

Validasi dilakukan dengan *Ten-fold Cross Validation*. *Ten-fold Cross Validation* adalah validasi yang dilakukan dengan cara membagi suatu set data menjadi sepuluh segmen yang berukuran sama besar dengan cara melakukan pengacakan data. Validasi dan pengujian dilakukan untuk mengetahui tingkat akurasi, presisi, dan recall dari hasil prediksi klasifikasi [10].

## 3. HASIL DAN PEMBAHASAN

### 3.1. Dataset

Dataset yang digunakan sebanyak 34 data dari 17 desa di Kecamatan Tulis Kabupaten Batang terlihat pada tabel 2.

Tabel 2. Dataset Jenis Kekerasan Perempuan, Anak dan Kelompok Kebutuhan Khusus

No	Desa	Tempat Kejadian	Jenis Kekerasan	Usia Korban	Pendidikan Korban	Jenis Kelamin Pelaku	Hubungan Pelaku	Bentuk Layanan
1	Manggis	Rumah	Fisik	Anak-Anak	TK	Laki-Laki	Orang Tua	Mediasi
2	Wringingintung	Fasilitas Umum	Seksual	Remaja	SMA	Laki-Laki	Tidak dikenal	Hukum
3	Beji	Sekolah	Psikis	Anak-Anak	SD	Laki-Laki	Teman	Mediasi
4	Tulis	Tempat Kerja	Psikis	Dewasa	Perguruan Tinggi	Perempuan	Rekan Kerja	Mediasi
5	Kaliboyo	Rumah	Penelantaran	Anak-Anak	PAUD	Perempuan	Orang Tua	Mediasi
6	Posong	Sekolah	Fisik	Anak-Anak	SD	Laki-Laki	Teman	Hukum
7	Simbangdesa	Tempat Kerja	Psikis	Dewasa	SMA	Laki-Laki	Rekan Kerja	Hukum
8	Kenconorejo	Fasilitas Umum	Fisik	Dewasa	SMA	Laki-Laki	Tetangga	Hukum
9	Beji	Rumah	Penelantaran	Lansia	Tidak Sekolah	Perempuan	Keluarga	Mediasi
10	Simbangdesa	Rumah	Fisik	Lansia	Tidak Sekolah	Perempuan	Tetangga	Mediasi
...	...	...	...	...	...	...	...	...
34	Sembojo	Fasilitas Umum	Fisik	Anak-Anak	SD	Laki-Laki	Teman	Mediasi

### 3.2. Perhitungan Entropy

Menghitung entropy total dilakukan dengan cara menghitung bentuk layanan “Hukum” dan “Mediasi” dari seluruh kasus yang ada

$$Entropy (total) \left( -\frac{10}{34} * \log_2\left(\frac{10}{34}\right) \right) + \left( -\frac{24}{34} * \log_2\left(\frac{24}{34}\right) \right) = 0,873981048$$

### 3.3. Perhitungan Information Gain

Hitung nilai gain untuk tiap atribut, lalu menentukan nilai gain tertinggi. nilai gain tertinggi itulah yang akan dijadikan akar dari pohon terlihat dalam tabel 3.

Tabel 3. Hasil Nilai Information Gain

No	Variabel	Nilai	Jumlah Kasus	Hukum	Mediasi	Entropy	Information Gain
1		Total	34	10	24	0,873981048	
							0,134921921
	Tempat kejadian	Rumah	13	1	12	0,391243564	
		Tempat kerja	4	2	2	1	
		Sekolah	10	5	5	1	
		Fasilitas umum	7	2	5	0,863120569	
							0,185961259
	Jenis Kekerasan	Fisik	13	5	8	0,961236605	
		Psikis/ Mental	7	2	5	0,863120569	
		Seksual	5	3	2	0,970950594	
		Eksplotasi	2	0	2	0	
		Penelantaran	7	0	7	0	
		Trafficking	0	0	0	0	
	Usia Korban	Anak-Anak	14	2	12	0,591672779	
		Remaja	8	4	4	1	
		Dewasa	8	4	4	1	
		Lansia	4	0	4	0	
	Pendidikan Korban						0,243127826
		Tidak Sekolah	2	0	2	0	
		Belum Sekolah	2	0	2	0	
		PAUD	1	0	1	0	
		TK	3	0	3	0	
		SD	9	1	8	0,503258335	
		SMP	4	2	2	1	
	SMA	9	5	4	0,99107606		
	Perguruan Tinggi	4	2	2	1		
	Jenis Kelamin Pelaku						0,285745754
		Laki-Laki	20	10	10	1	
		Perempuan	14	0	14	0	
	Hubungan Pelaku						0,273914479
		Orang Tua	9	0	9	0	
		Keluarga/ Saudara	4	1	3	0,811278124	
		Suami/ Istri	0	0	0	0	
		Tetangga	3	1	2	0,918295834	
		Teman	11	4	7	0,945660305	
		Guru/ Dosen	1	1	0	0	
		Majikan	0	0	0	0	
		Rekan Kerja	4	2	2	1	
	Tidak Dikenal	1	1	1	0		

### 3.4. Decision Tree

Dari hasil Perhitungan entropy dan information gain yang didapat kemudian diolah kedalam Decision Tree



Gambar 2. *Decision Tree*

Keterangan:

Jika jenis kelamin pelaku adalah perempuan maka bentuk layanan adalah mediasi

1. Jika jenis kelamin pelaku adalah laki-laki dan tempat kejadian adalah tempat kerja maka bentuk layanan adalah hukum
2. Jika jenis kelamin pelaku adalah laki-laki dan tempat kejadian adalah sekolah dan pendidikan korban adalah SMP maka bentuk layanan adalah hukum.
3. Jika jenis kelamin pelaku adalah laki-laki dan tempat kejadian adalah sekolah dan pendidikan korban adalah SMA maka bentuk layanan adalah hukum.
4. Jika jenis kelamin pelaku adalah laki-laki dan tempat kejadian adalah sekolah dan pendidikan korban adalah SD maka bentuk layanan adalah mediasi.
5. Jika jenis kelamin pelaku adalah laki-laki dan tempat kejadian adalah rumah maka bentuk layanan adalah mediasi.
6. Jika jenis kelamin pelaku adalah laki-laki dan tempat kejadian adalah fasilitas umum dan pendidikan korban adalah SMA maka bentuk layanan adalah hukum.
7. Jika jenis kelamin pelaku adalah laki-laki dan tempat kejadian adalah fasilitas umum dan pendidikan korban adalah SD maka bentuk layanan adalah mediasi.

### 3.5. Hasil Pengujian Model dan Perbandingan

Berikut tabel perbandingan dari tingkat akurasi antara hasil klasifikasi menggunakan algoritma C4.5 dengan menggunakan algoritma C4.5 dengan PSO.

Tabel 4. Hasil Pengujian Nilai *Accuracy* C4.5

accuracy: 80.83% +/- 17.10% (mikro: 79.41%)			
	True <u>mediasi</u>	True <u>hukum</u>	Class Precision
Pred. <u>Mediasi</u>	21	4	84.00%
Pred. <u>Hukum</u>	3	6	66.67%
Class Recall	87.50%	60.00%	

Tabel 5. Hasil Pengujian Nilai *Precision* C4.5

precision: 66.67% (positive class: <u>Hukum</u> )			
	True <u>mediasi</u>	True <u>hukum</u>	Class Precision
Pred. <u>Mediasi</u>	21	4	84.00%
Pred. <u>Hukum</u>	3	6	66.67%
Class Recall	87.50%	60.00%	

Tabel 6. Hasil Pengujian Nilai *Recall* C4.5

recall: 60.00% +/- 48.99% (mikro: 60.00%) (positive class: Hukum)			
	True mediasi	True hukum	Class Precision
Pred. Mediasi	21	4	84.00%
Pred. Hukum	3	6	66.67%
Class Recall	87.50%	60.00%	

Tabel 7. Hasil Pengujian Nilai *Accuracy* C4.5 dengan *PSO*

accuracy: 82.50% +/- 14.65% (mikro: 82.35%)			
	True mediasi	True hukum	Class Precision
Pred. Mediasi	21	3	87.50%
Pred. Hukum	3	7	70.00%
Class Recall	87.50%	70.00%	

Tabel 8. Hasil Pengujian Nilai *Precision* C4.5 dengan *PSO*

precision: 70.00% (positive class: Hukum)			
	True mediasi	True hukum	Class Precision
Pred. Mediasi	21	3	87.50%
Pred. Hukum	3	7	70.00%
Class Recall	87.50%	70.00%	

Tabel 9. Hasil Pengujian Nilai *Recall* C4.5 dengan *PSO*

recall: 70.00% +/- 45.83% (mikro: 70.00%) (positive class: Hukum)			
	True mediasi	True hukum	Class Precision
Pred. Mediasi	21	3	87.50%
Pred. Hukum	3	7	70.00%
Class Recall	87.50%	70.00%	

Tabel 10. Perbandingan Nilai C4.5 dan C4.5 dengan *PSO*

Nilai	C4.5	C4.5 + PSO
Accuracy	80.83%	82.50%
Precision	66.67%	70.00%
Recall	60.00%	70.00%

Dengan menggunakan fitur seleksi *Particle Swarm Optimization* terbukti bisa meningkatkan *accuracy* 1,67% dari algoritma C4.5 yang digunakan.

#### 4. KESIMPULAN

Berdasarkan pembahasan dan pengujian yang telah dilakukan menggunakan algoritma klasifikasi data mining C4.5 menghasilkan nilai accuracy sebesar 80.83% dengan memakai dataset jenis kekerasan perempuan, anak dan kelompok kebutuhan khusus di Kecamatan Tulis Kabupaten Batang Jawa Tengah. Kemudian dilakukan penambahan fitur seleksi pembobotan menggunakan Particle Swarm Optimization menghasilkan nilai accuracy sebesar 82.50% mengalami peningkatan sebesar 1,67% dibandingkan menggunakan algoritma C4.5 saja.

#### DAFTAR PUSTAKA

- [1] Y. Fendrita dan G. Pawitan, "Persepsi Masyarakat Terkait Kekerasan terhadap Perempuan di Indonesia," *Open Journal System Kementerian Sosial RI*, pp. 157-170, 2021.
- [2] D. Indriani, "Pantura Tribunnews," 31 Agustus 2021. [Online]. Available: <https://pantura.tribunnews.com/2021/08/31/kasus-kdrt-di-batang-meningkat-banyak-kekerasan-seksual-terhadap-anak-ini-langkah-dp3ap2kb>.
- [3] Jumadi, "Berita Batang," 23 Juli 2020. [Online]. Available: <https://berita.batangkab.go.id/?p=1&id=5157>.
- [4] A. Purwanto and A. Primajaya, "Penerapan Algoritma C4.5 dalam Prediksi Potensi Tingkat Kasus Pneumonia di Kabupaten Karawang," *Jurnal Sistem dan Teknologi Informasi*, pp. 390-396, 2020.
- [5] I. Yulianti, R. A. Saputra, M. S. Mardiyanto and A. Rahmawati, "Optimasi Akurasi Algoritma C4.5 berbasis Particle Swarm Optimization dengan Teknik Bagging Pada Prediksi Penyakit Ginjal Kronis," *Techno Com*, vol. 19, pp. 411-421, 2020.
- [6] B. P. S. K. Batang, Kecamatan Tulis dalam Angka, Batang: CV. Niaga, 2019.
- [7] W. R. S. O. Ningse, "Klasifikasi Algoritma C4.5 untuk Penentuan Penerima Program Indonesia Pintar pada MIS Al-Khoirot," *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, pp. 65 - 76, 2021.
- [8] N. F. Istighfarin, "Penerapan Metode Particle Swarm Optimization (PSO) dan Genetic Algorithm (GA) pada Sistem Optimasi Visible Light Communication (VLC) Untuk menentukan Posisi Robot," *Simetris*, pp. 279-286, 2020.
- [9] A. Rakhman and M. R. Tsani, "Analisis Sentimen Review Media Massa Menggunakan Metode C4.5 Berbasis Forward Selection," *Smart Comp*, vol. 8, no. 2089-676X, pp. 78 - 82, 2019.

- [10] M. F. Arifin and D. Fitiranah, "Penerapan Algoritma Klasifikasi C4.5 dalam Rekomendasi Penerimaan Mitra Penjualan Studi Kasus : PT Atria Artha Persada," *Jurnal Telekomunikasi dan Komputer*, vol. 8, no. 2085-4811, pp. 87-102, 2018.
- [11] K. F. Irnanda, "Analisa Klasifikasi C4.5 Terhadap Faktor Penyebab Menurunnya Prestasi Belajar Mahasiswa Pada Masa Pandemi," *Jurnal Media Informatika Budidarma*, pp. 327-331, 2021.