

Algoritma C4.5 Untuk Klasifikasi Jenis Kekerasan pada Anak (Kasus DP3A Kabupaten Karawang)

Raka Putra Eshardiansyah^{*1}, Nina Sulistiyowati², Mohamad Jajuli³

^{1,2,3}Teknik Informatika, Universitas Singaperbangsa Karawang

Jl. H.S. Ronggowaluyo Kel. Puseurjaya Kec. Telukjambe Timur Kab. Karawang, Prov. Jawa Barat, 0267641177

raka.eshardiansyah17172@student.unsika.ac.id^{*1}, nina.sulistio@unsika.ac.id², mohamad.jajuli@unsika.ac.id³

Abstract

Kabupaten Karawang is an area ranked 3rd as an area that has the potential to be prone to violence against children. Every year from 2016 to 2020 always shows an increasing number every year. Classification with data mining by applying the 4.5 algorithm is expected to be a solution to reduce the number of violence against children in Karawang Regency. The data that has been obtained is validated using k-fold cross validation and then evaluated using a confusion matrix and ROC curve. The C4.5 algorithm can be used to classify data on violence against children in Karawang Regency and get good results. The results of the confusion matrix using 60:40 split data on the violence dataset in children in Karawang Regency produce an accuracy value of 82.8125%, a precision value of 96.8% and a recall value of 90.5%. While the evaluation with the ROC curve with the classification quality "Good Classification" of 0.897. The 70:30 data split method was evaluated with the confusion matrix, resulting in an accuracy value of 93.2432%, precision of 94.5% and recall of 87.5%. While the evaluation with the ROC curve with the classification quality "Excellent Classification" of 0.953. The 80:20 data split method was evaluated with the confusion matrix, resulting in an accuracy value of 92.9142%, a precision of 91.7% and a recall of 81%. While the evaluation with the ROC curve with the classification quality "Excellent Classification" of 0.941.

Keywords: Data mining, C4.5, K-fold cross validation, Violence in Children

Abstrak

Kabupaten Karawang merupakan daerah peringkat ke-3 sebagai daerah yang memiliki potensi rawan kekerasan terhadap anak. Setiap tahun mulai 2016 hingga 2020 selalu menunjukkan angka kenaikan setiap tahunnya. Klasifikasi dengan data mining dengan menerapkan algoritma C4.5 diharapkan dapat menjadi solusi untuk mengurangi angka kekerasan pada anak di Kabupaten Karawang. Data yang telah didapatkan divalidasi menggunakan k-fold cross validation kemudian dievaluasi dengan confusion matrix serta kurva ROC. Algoritma C4.5 digunakan untuk melakukan klasifikasi data kekerasan pada anak di Kabupaten Karawang dan mendapatkan hasil yang baik. Hasil dari confusion matrix menggunakan split data 60:40 pada dataset menghasilkan nilai akurasi sebesar 82.8125%, nilai precision sebesar 82.9% dan nilai recall sebesar 82.8%. Sedangkan evaluasi dengan kurva ROC dengan kualitas klasifikasi "Good Classification" sebesar 0.897. Metode split data 70:30 dievaluasi dengan confusion matrix menghasilkan nilai akurasi sebesar 93.2432%, precision sebesar 93.1% dan recall sebesar 93.2%. Sedangkan evaluasi dengan kurva ROC dengan kualitas klasifikasi "Excellent Classification" sebesar 0.953. Metode split data 80:20 dievaluasi dengan confusion matrix menghasilkan nilai akurasi sebesar 92.9142%, precision sebesar 92.8% dan recall sebesar 92.9%. Sedangkan evaluasi dengan kurva ROC dengan kualitas klasifikasi "Excellent Classification" sebesar 0.941.

Kata kunci: Data mining, C4.5, K-fold cross validation, Kekerasan pada Anak

1. PENDAHULUAN

Kabupaten Karawang merupakan salah satu daerah yang terletak di Provinsi Jawa Barat. Berdasarkan catatan data yang tertera pada karawangkab.bps.go.id, Kabupaten Karawang terdiri dari 30 Kecamatan. Pada tahun 2020, Kabupaten Karawang tercatat memiliki penduduk sebanyak 2.370.488 jiwa. Kabupaten Karawang memiliki beberapa misi untuk memajukan daerahnya, salah satunya menurut Banuara Nadeak selaku Kepala Badan Keluarga Berencana dan Pemberdayaan Perempuan ialah untuk membuat Karawang menjadi daerah yang layak akan anak. Sebagai langkah awal, pemerintah Kabupaten Karawang meresmikan Dinas Pemberdayaan Perempuan dan Perlindungan Anak pada tahun 2016. Terdapat beberapa faktor yang melatarbelakangi kasus kekerasan pada anak di antaranya ialah faktor pendidikan, faktor psikologis, faktor sosial dan ekonomi [1]. Berdasarkan data yang dilansir oleh Kementerian Pemberdayaan Perempuan dan Perlindungan Anak (Kementerian PPPA), pada tahun 2020 tercatat 4116 kasus kekerasan pada anak tersebut terdiri dari 1.111 kekerasan fisik, 979 kekerasan psikis, 2.556 kekerasan seksual, 68 eksploitasi, 73 perdagangan orang dan 346 penelantaran. Angka tersebut merupakan angka yang terbilang tinggi walaupun sedang mengalami pandemi Covid-19.

Data mining diharapkan dapat menjadi solusi untuk permasalahan yang dihadapi oleh DP3A untuk meminimalisir kejadian tersebut. Manfaat *data mining* ialah sebagai bahan atau metode untuk mendukung pengambilan keputusan serta menunjang kegiatan yang dibutuhkan [2]. DP3A memiliki data tentang kasus kekerasan pada anak yang terjadi di Kabupaten Karawang. Namun, hal tersebut belum dilakukan pengolahan data terhadap jenis kekerasan berdasarkan karakteristiknya. Klasifikasi akan dibagi menjadi 4 *class target* yaitu kekerasan ekonomi, fisik, psikologis dan seksual. Sebagai contoh dalam UU KDRT, kekerasan ekonomi adalah setiap perbuatan yang didasarkan oleh pengancaman ataupun pemaksaan terhadap kebutuhan finansial seperti eksploitasi anak dibawah umur untuk melakukan pekerjaan yang menghasilkan uang dengan cara diancam. Kekerasan Dalam Rumah Tangga (KDRT) juga dapat dikelompokkan ke dalam jenis ini. Salah satu cara yang efektif untuk mengatasi hal ini ialah dengan membuat Undang-Undang Penghapusan Kekerasan Dalam Rumah Tangga yang fungsinya ialah untuk memperoleh keadilan dalam berumah tangga [3]. Begitu pun dengan jenis kekerasan lainnya yang memiliki solusinya masing-masing. Penelitian ini menggunakan teknik *data mining* klasifikasi untuk menyelesaikan permasalahan tersebut yang bertujuan untuk mempermudah DP3A mengambil solusi terhadap jenis kekerasannya. *Knowledge Discovery in Database* (KDD) merupakan salah satu metodologi yang digunakan untuk menerapkan konsep *data mining*. KDD akan menghasilkan pengetahuan (knowledge) yang hasilnya dapat digunakan untuk mengambil keputusan [4].

Data mining algoritma C4.5 berfungsi sebagai penyusun sistem yang dapat melihat pola kejadian kekerasan pada anak untuk bisa dijadikan

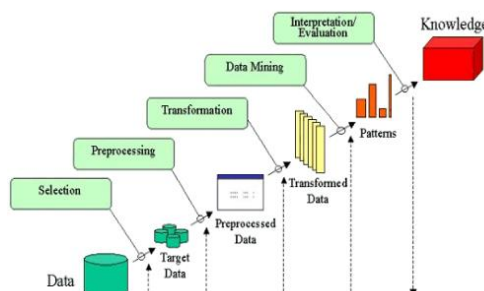
sebagai sebuah strategi dalam proses klasifikasi jenis kekerasan pada anak di DP3A [5]. Algoritma C4.5 memiliki keunggulan mudah diimplementasikan, akurasi dapat diterima dan efisien untuk menangani tipe data diskret maupun numerik kontinu [6].

2. METODOLOGI PENELITIAN

Penelitian ini digunakan metodologi *data mining Knowledge Discovery in Database* (KDD). *Knowledge Discovery in Database* (KDD) dapat digunakan untuk menggambarkan garis besar proses pengekstrakan pada *database* yang besar [7].

2.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) berfungsi untuk menguraikan knowledge yang terdapat pada *database*. KDD memiliki proses yang bersifat interaktif dan iteratif yang akan membantu pengguna untuk membantu membuat keputusan dan dapat dilakukan pengulangan di antara dua buah data [8].



Gambar 1. Tahapan *Knowledge Discovery in Database*

2.2. Algoritma C4.5

Algoritma C4.5 merupakan kembangan atau evolusi dari algoritma ID3[2]. Algoritma C4.5 merupakan algoritma klasifikasi yang bentuk visualisasinya menghasilkan sebuah pohon keputusan. *Decision tree* tersebut digunakan karena memiliki keunggulan yaitu menghasilkan sebuah pohon keputusan yang mudah untuk dipahami, mempunyai taraf akurasi yang baik, efisien terhadap data yang memiliki atribut bertipe diskret serta numerik kontinyu [5]. Algoritma C4.5 bekerja dengan mengunjungi tiap *node*, memilih percabangan yang optimal, hingga tidak adalagi yang dapat dihasilkan. Algoritma C4.5 memiliki beberapa konsep seperti *Entropy* dan konsep *Gain* [9].

Secara umum, pembentukan pohon keputusan memiliki prosedur proses sebagai berikut [10] :

- Pilih atribut yang akan dijadikan sebagai akar.
- Buat cabang untuk tiap nilai.
- Bagi kasus dalam beberapa cabang.

d) Ulangi proses tersebut hingga semua kasus pada cabang memiliki kelas yang sama

Entropy merupakan jumlah *bit* yang kebutuhannya diperkirakan dengan cara mengekstrak suatu kelas (+ atau -) dari sejumlah data acak pada ruang sampel tertentu. *Entropy* digunakan untuk mengukur ketidakslisan ruang sampel. Formula dalam penentuan nilai *Entropy* ialah sebagai berikut [10] :

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

Keterangan:

S = Himpunan Kasus
 A = Fitur
 n = Jumlah Partisi S
 pi = Proporsi dari Si terhadap

Gain Rasio merupakan rumus perbaikan dari ID3 untuk diperbaharui *information gain*. Dapat dirumuskan sebagai berikut ID3 [2] :

$$GainRatio(S,A) = (GAIN(S,A))/(Splitinfo(S,A)) \quad (2)$$

Keterangan:

Gain(S,A) = *information gain* pada atribut A
Splitinfo(S,A) = *split information* pada atribut A

Gain rasio yang paling tinggi, dipilih sebagai atribut *test* untuk membentuk simpul. [2] :

$$Splitinfo(S,A) = - \sum_{i=1}^s \frac{s_i}{s} \log_2 \frac{s_i}{s} \quad (3)$$

Keterangan:

S = Ruang sampel *data training*
 A = Atribut
 Si = Jumlah sampel untuk atribut i

3. HASIL DAN PEMBAHASAN

Penelitian ini dilakukan untuk melakukan klasifikasi terhadap kekerasan pada anak menggunakan metode *k-fold cross validation* pada 3 rasio dan menggunakan algoritma C4.5. Terdapat 10 atribut pada *dataset* yang digunakan. Kemudian diseleksi menjadi 4 atribut sesuai dengan kebutuhan penelitian dimana atribut tersebut termasuk *class target*. Pemilihan dilakukan dengan bantuan *Microsoft Excel*. Berikut merupakan atribut setelah dilakukan penyeleksian pada Tabel 1.

Tabel 1. Atribut Yang Digunakan

No	Atribut	Tipe Data
1	Jenis Kelamin L dan P	Nominal
2	Umur	Numerik

No	Atribut	Tipe Data
3	Kasus	Nominal
4	Bentuk Perhatian (<i>Class target</i>)	Nominal

Pada Tabel 1, Bentuk Perhatian merupakan atribut yang digunakan untuk menentukan label dari setiap *record* yang ada. Sisa dari atribut lainnya merupakan atribut pendukung yang digunakan untuk menentukan label dari *record* yang ada. Kemudian *dataset* dilakukan tahapan *preprocessing* untuk mengisi *record* yang memiliki nilai *missing value*. Atribut yang memiliki tipe data numerik diisi menggunakan *mean* sedangkan atribut yang memiliki tipe data nominal diisi dengan menggunakan modus.

Tabel 2. *Dataset* awal kekerasan pada anak

Jenis Kelamin	Umur	Kasus	Bentuk Perhatian
...	...	...	...
P	15	Pelecehan Seksual	Tinggi
P	6	Pelecehan Seksual	Tinggi
...	...	...	...

Tabel 2 merupakan tampilan *record* setelah melalui proses *preprocessing*. Semua *missing value* yang terdapat pada *record* diisi dengan menggunakan *mean* dan modus yang kemudian data tersebut perlu dilakukan transformasi untuk mengoptimalkan model yang dihasilkan terutama pada atribut umur dan kasus karena memiliki rentang yang cukup jauh sehingga hal tersebut dapat mengurangi kualitas hasil model. Selanjutnya tabel 3 menampilkan hasil *record* yang telah di transformasi.

Tabel 3. *Dataset* setelah melalui proses transformasi

Jenis Kelamin	Umur	Kasus	Bentuk Perhatian
...	...	...	...
P	Remaja	Kekerasan Seksual	Tinggi
P	Kanakkanak	Kekerasan Seksual	Tinggi
...	...	...	...

Pada tabel 3 menunjukkan *dataset* yang telah dilakukan transformasi. Pada atribut umur dibagi menjadi 3 kategori, yaitu Balita dengan rentang umur 0-5 tahun, Kanakkanak dengan rentang umur 6-11 tahun dan Remaja dengan rentang umur 12-17 tahun.

Kemudian untuk melakukan klasifikasi menggunakan algoritma C4.5 perlu menghitung nilai *entropy* dari setiap kategori dan nilai *information gain* dari setiap atributnya. Dimana proses perhitungan dilakukan sebagai berikut.

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 &= -p_1 * \log_2 p_1 - -p_2 * \log_2 p_1 \\
 &= -\left(\frac{S_1}{S}\right) * \log_2 \left(\frac{S_1}{S}\right) - \left(\frac{S_2}{S}\right) * \log_2 \left(\frac{S_2}{S}\right) - \left(\frac{S_3}{S}\right) * \log_2 \left(\frac{S_3}{S}\right) \\
 &= -\frac{70}{106} * \log_2 \left(\frac{70}{106}\right) - \left(\frac{24}{106}\right) * \log_2 \left(\frac{24}{106}\right) - \left(\frac{12}{106}\right) * \log_2 \left(\frac{12}{106}\right) \\
 &= 1.236331
 \end{aligned}$$

Proses perhitungan diatas merupakan proses perhitungan *entropy* pada jumlah total, dimana jumlah *entropy* total merupakan *entropy* penting yang digunakan untuk menghitung nilai *gain* yang terdapat pada tiap atributnya. Berikut merupakan contoh dari perhitungan *gain* atribut jenis kelamin.

$$\begin{aligned}
 Gain(S.A) &= Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \\
 &= Entropy(total) - \sum_{i=1}^2 \frac{|S_{jenis\ kelamin}|}{|S_{total}|} * Entropy(jenis\ kelamin) \\
 &= 1.236331 - \left(\frac{70}{106}\right) * 0.899756 - \left(\frac{36}{106}\right) * 1.542617 \\
 &= 0.118245
 \end{aligned}$$

Setelah melakukan perhitungan *entropy* dan *gain*, lakukan kembali perhitungan tersebut terhadap semua kategori dan atribut yang terdapat pada *dataset*. Tabel 4 menunjukkan hasil perhitungan *gain* dan *entropy* yang dihasilkan.

Tabel 4. Hasil perhitungan *entropy* dan *gain*

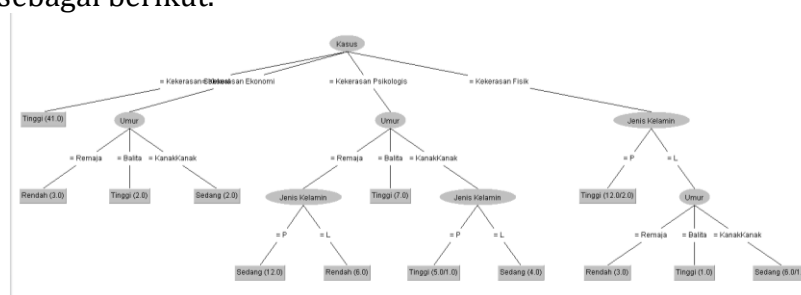
Node 1		Jumlah (s)	Tinggi (s)	Sedang (s)	Rendah (s)	Entropy	Gain
Total		106	70	24	12	1.236331	
Jenis Kelamin							0.118245
	P	70	54	14	2	0.899756	
	L	36	16	9	11	1.542617	
Umur							0.187119
	Balita	19	19	0	0	0	
	KanakKanak	39	28	10	1	0.982191	
	Remaja	48	23	13	12	1.518979	
Kasus							0.357324
	Kekerasan Fisik	20	11	5	4	1.438759	
	Kekerasan Ekonomi	8	3	2	3	1.561278	
	Kekerasan Psikologis	35	13	16	6	1.483126	
	Kekerasan Seksual	43	43	0	0	0	

Langkah selanjutnya ialah melakukan *split* pada *dataset* kekerasan pada anak menggunakan rasio 60:40, 70:30 dan 80:20. *Dataset* dibagi menjadi 2 yakni *data training* dan *data testing* untuk dilakukan proses model. Tabel 5 menunjukkan jumlah pembagian data dari masing-masing rasio.

Tabel 5. Hasil pembagian data

Rasio	<i>Data training</i>	<i>Data testing</i>
60:40	64	42
70:30	74	32
80:20	85	21

Setelah dilakukan *split data*, langkah selanjutnya ialah melakukan proses *modelling* dengan menggunakan *machine learning* WEKA serta metode *k-fold cross validation* untuk mendapatkan hasil model terbaik pada ketiga rasio tersebut. Angka *k* pada *fold* diisi dengan angka 10. Kemudian setelah dilakukan proses model, pembentukan *decision tree* juga dilakukan dengan bantuan *machine learning* WEKA sehingga menghasilkan pohon keputusan sebagai berikut.



Gambar 2. *Decision tree* dataset kekerasan pada anak

Pohon keputusan yang dibentuk oleh algoritma C4.5 menghasilkan 13 *rule* total terhadap klasifikasi jenis kekerasan pada anak dengan penjelasan *rule* sebagai berikut.

Tabel 6. *Rule decision tree*

No.	<i>Rule</i>
1	<i>If Kasus is Kekerasan Seksual, then Bentuk Perhatian is Tinggi.</i>
2	<i>If Kasus is Kekerasan Ekonomi and Umur is Balita, then Bentuk Perhatian is Tinggi.</i>
3	<i>If Kasus is Kekerasan Psikologis and Umur is Balita, then Bentuk Perhatian is Tinggi.</i>
4	<i>If Kasus is Kekerasan Psikologis and Umur is KanakKanak and Jenis Kelamin is P, then Bentuk Perhatian is Tinggi.</i>
5	<i>If Kasus is Kekerasan Fisik and Jenis Kelamin is P, then Bentuk Perhatian is Tinggi.</i>
6	<i>If Kasus is Kekerasan Fisik and Jenis Kelamin is L and Umur is Balita, then Bentuk Perhatian is Tinggi.</i>
7	<i>If Kasus is Kekerasan Ekonomi and Umur is KanakKanak, then Bentuk Perhatian is Sedang.</i>

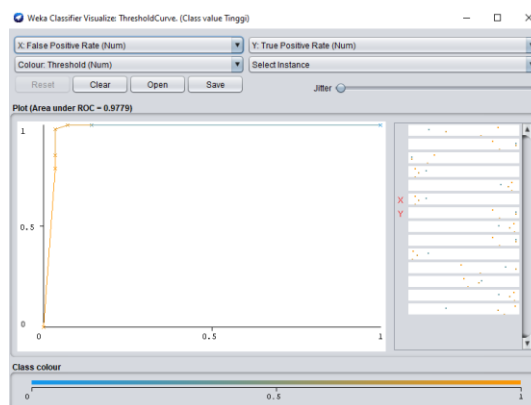
No.	Rule
8	If Kasus is Kekerasan Psikologis and Umur is Remaja and Jenis Kelamin is P, then Bentuk Perhatian is Sedang.
9	If Kasus is Kekerasan Psikologis and Umur is KanakKanak and Jenis Kelamin is L, then Bentuk Perhatian is Sedang.
10	If Kasus is Kekerasan Fisik and Jenis Kelamin is L and Umur is KanakKanak, then Bentuk Perhatian is Sedang.
11	If Kasus is Kekerasan Ekonomi and Umur is Remaja, then Bentuk Perhatian is Rendah.
12	If Kasus is Kekerasan Psikologis and Umur is Remaja and Jenis Kelamin is L, then Bentuk Perhatian is Rendah.
13	If Kasus is Kekerasan Fisik and Jenis Kelamin is L and Umur is Remaja, then Bentuk Perhatian is Rendah.

Kemudian untuk melakukan perhitungan validasi akurasi, *precision* dan *recall* yang dihasilkan oleh WEKA dapat menggunakan *confusion matrix*. Berikut merupakan proses perhitungan *confusion matrix* pada *split* 70:30.

Tabel 7. Tabel perhitungan *confusion matrix* *split* 70:30

Confusion matrix		Target				
		Tinggi	Sedang	Rendah		
Model	Tinggi	46	0	0	Recall tinggi	46/46 = 100%
	Sedang	1	14	1	Recall sedang	14/16 = 87.5%
	Rendah	2	1	9	Recall rendah	9/12 = 75%
		Precision tinggi	Precision sedang	Precision rendah	Akurasi = 69/74 = 93.2432%	
		46/49 = 93.9%	14/15 = 93.3%	9/10 = 90%		

Kemudian hasil dari model yang dihasilkan oleh *split* 70:30 dilakukan pengukuran kualitas klasifikasi dengan menggunakan grafik ROC. Grafik ROC menghasilkan nilai AUC yang digunakan untuk menentukan hasil kualitas klasifikasi dari model yang dihasilkan. Grafik ROC dilakukan dengan WEKA. Berikut merupakan grafik ROC yang dihasilkan.



Gambar 3. Grafik ROC *Split* 70:30

Setelah dilakukan proses *modelling* pada ketiga rasio tersebut, kemudian dibentuk sebuah tabel untuk membandingkan nilai akurasi, *precision* dan *recall* yang dihasilkan. Dari ketiga metode *split* tersebut, metode *split* 70:30 dengan *k-fold cross validation* menghasilkan nilai akurasi, *precision*, *recall* serta AUC paling tinggi dibandingkan dengan yang lainnya.

Metode	Akurasi	Precision	Recall	AUC	Kualitas
60:40	82.8125%	82.9%	82.8%	0.897	Good Classification
70:30	93.2432%	93.1%	93.2%	0.953	Excellent Classification
80:20	92.9412%	92.8%	92.9%	0.941	Excellent Classification

4. SIMPULAN

Setelah dilakukan proses evaluasi dengan menggunakan *confusion matrix* untuk menghitung akurasi, *precision* serta *recall* pada ketiga rasio tersebut. *Confusion matrix* menunjukkan pada metode *k-fold cross validation* rasio 60:40 menghasilkan nilai akurasi sebesar 82.8125%, *precision* sebesar 82.9% dan *recall* sebesar 82.8%. Pada metode *k-fold cross validation* rasio 70:30 menghasilkan nilai akurasi sebesar 93.2432%, *precision* sebesar 93.1% dan *recall* sebesar 93.2%. Pada metode *k-fold cross validation* rasio 80:20 menghasilkan nilai akurasi sebesar 92.9412%, *precision* sebesar 92.8% dan *recall* sebesar 92.9%. Sementara itu, penelitian sebelumnya dengan kasus klasifikasi pola kelayakan kredit menggunakan algoritma C4.5 yang dilakukan oleh [11] menyimpulkan bahwa *k-fold cross validation* dapat membuat hasil akurasi, *precision* dan *recall* menjadi lebih akurat serta menghasilkan nilai bias yang kecil dan grafik ROC dapat menentukan hasil kualitas dari klasifikasi yang dihasilkan. Tetapi, hal ini bergantung juga kepada jumlah dan kualitas *dataset* yang digunakan sehingga dapat mendapatkan hasil yang terbaik.

Berdasarkan dari penelitian yang dilakukan, dapat disimpulkan bahwa penggunaan *k-fold cross validation* dapat meningkatkan akurasi, *precision* serta *recall* dari ketiga *split* rasio yang digunakan. Metode 70:30 mendapatkan hasil model yang terbaik dibandingkan dengan metode *split* lainnya. Hal ini ditunjukkan nilai akurasi, presisi, *recall* dan AUC yang dihasilkan memiliki nilai yang paling tinggi dibandingkan dengan metode *split* lainnya dengan mendapatkan nilai akurasi sebesar 93.2432%, *precision* sebesar 93.1%, *recall* sebesar 93.2% dan AUC sebesar 0.953 dengan kualitas klasifikasi "Excellent Classification".

DAFTAR PUSTAKA

- [1] Bariah, O., & Marlina, R., "Tinjauan Psikologi dan Agama Terhadap Tindak Kekerasan pada Anak", Jurnal Studia Insania, 7(2), 92., (2019).
- [2] Romli, I., & Zy, A. T., "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4. 5", J-SAKTI (Jurnal Sains Komputer dan Informatika), 4(2), 694-702., (2020).

- [3] Iskandar, D., "Upaya Penanggulangan Terjadinya Kekerasan Dalam Rumah Tangga". *Jurnal Yustisi*, 3(2), 13-22., (2016).
- [4] Gustientiedina, G., Adiya, M. H., & Desnelita, Y., "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan", *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17-24., (2019).
- [5] Gultom, J. P., & Rikki, A., "Implementasi *Data mining* menggunakan Algoritma C-45 pada Data Masyarakat Kecamatan Garoga untuk Menentukan Pola Penerima Beras Raskin", *KAKIFIKOM* 02(01), 11-19., (2020).
- [6] Yusup, M., Panjaitan, E. S., & Yunis, R., "Analisis Kinerja dalam Mendeteksi Student Loses Berdasarkan Nilai *Gain* dengan *Splite* Feature Reduction Model pada Algoritma C4.5", *CESS (Journal of Computer Engineering, System and Science)*, 5(2), 267., (2020).
- [7] Hendra, H., "Penerapan *Data mining* untuk Prediksi Penjualan Readymix Menggunakan Metode Algoritma C4. 5 pada PT Remicon Widyaprima", (Doctoral dissertation, Prodi Sistem Informasi), (2020).
- [8] Apriliana, Natalis Ransi, J. N., "Implementasi Text Mining Klasifikasi Skripsi Menggunakan Metode Naïve Bayes Classifier", *SemanTIK*, Vol.3, No.2, Jul-Des 2017, 3(2), 187-194., (2017).
- [9] Oktavia, R., Sutardi, & Nangi, J., "Implementasi *Data mining* Untuk Menentukan Daerah Rawan Kecelakaan Menggunakan Algoritma C4.5 Rezki", *SemanTIK*, 3(2), 95-104., (2017).
- [10] Merawati, D., & Rino., "Penerapan *data mining* penentu minat Dan bakat siswa Smk dengan metode C4.5", *Jurnal Algor*, 1(1), 28-37., (2019).
- [11] Abdussomad., "Penerapan Algoritma Klasifikasi C4.5 Untuk Menghasilkan Pola Kelayakan Kredit", *Lnovasi Lnfornnalika*, II(1), 19-27., (2017).