



P-ISSN: 2528-5688
E-ISSN: 2528-5696

VOLT

Jurnal Ilmiah Pendidikan Teknik Elektro

Journal homepage: jurnal.untirta.ac.id/index.php/VOLT

Vol x, No. x, Oktober 2023, xx-xx



IMPLEMENTASI ALGORITMA C4-5 PADA JUMLAH DATA ANAK PENERIMA AKTA KELAHIRAN DI DISDUKCAPIL KOTA PALEMBANG

Wulan Damayanti^{1*}, Devi Udariansyah²

¹Program Studi Teknik Informatika, Fakultas Sains Teknologi
Universitas Bina Darma, Jl. Jenderal A. Yani No.12 Plaju, Palembang
E-mail: wulandamayanti793@gmail.com

²Program Studi Teknik Informatika, Fakultas Sains Teknologi
Universitas Bina Darma, Jl. Jenderal A. Yani No.12 Plaju, Palembang
Indonesia
e-mail: devi.udariansyah@binadarma.ac.id

Diterima: xx bulan xxxx. Disetujui: xx bulan xxxx. Dipublikasikan: bulan xxxx

Abstrak

Pencatatan akta lahir yang dilakukan oleh Disdukcapil Kota Palembang masih belum efektif karena masih banyaknya masyarakat yang belum mendaftarkan diri dalam akta yang temuat di Disdukcapil Kota Palembang. Tujuan dalam penelitian ini adalah untuk menerapkan data mining dengan algoritma C4.5, mengklasifikan akta anak yang terdaftar dan tidak terdaftar, menerapkan decision tree (pohon keputusan) dalam menciptakan gain pada masing-masing atribut, dan pengimplementasian algoritma C4.5 menggunakan bantuan perangkat lunak RapidMiner 9.5. Prosedur penelitian diawali dengan mengumpulkan sumber data berupa data agregat penerima akta masyarakat Kota Palembang dari Tahun 2021-2023. Kemudian, data tersebut dianalisis menggunakan bantuan Microsoft Excel dan RapidMiner 9.5 untuk menentukan nilai entropy, gain, dan decision tree. Hasil penelitian menunjukkan bahwa sebanyak 1.606 masyarakat dikategorikan sebagai True Positif (TP), dimana diklasifikasikan dengan masyarakat Kota Palembang yang telah mendaftarkan dirinya di akta kelahiran Diskdukcapil dan False Negatif (FN) sebanyak 397 record yang berarti bahwa sebanyak jumlah tersebut terdapat masyarakat yang masih belum mendaftarkan dirinya di akta kelahiran Disdukcapil Kota Palembang.

© 2023 Jurusan Pendidikan Teknik Elektro, FKIP UNTIRTA

Kata kunci: memuat kata-kata pokok, terdiri dari tiga – lima kata (ditulisurut secara alphabetic)

PENDAHULUAN

Terjadinya pertumbuhan populasi di tengah masyarakat, membuat Disdukcapil Kota Palembang menghadapi peningkatan signifikan dalam jumlah data anak penerima akta kelahiran. Pencatatan akta kelahiran adalah salah satu proses administratif penting yang dilakukan oleh Dinas Kependudukan dan Pencatatan Sipil (Disdukcapil) di Kota Palembang. Namun demikian, ternyata masih ditemukan masalah dalam pencatatan akta lahir, seperti masih adanya masyarakat yang belum mendaftarkan diri atau adanya masalah pencatatan internal di Disdukcapil.

Dengan adanya permasalahan tersebut, maka salah satu metode pencatatan akta yang efektif dapat dilakukan menggunakan analisis C4.5 dengan *RapidMiner*. *RapidMiner* adalah sebuah perangkat lunak dalam *data mining* yang prediktif yang akan menyediakan pengetahuan kepada penggunanya dalam membuat suatu keputusan [1]. Perangkat lunak ini bermanfaat dalam meningkatkan efisiensi proses pencatatan akta kelahiran dengan melakukan klasifikasi data yang efektif berdasarkan atribut-atribut yang relevan [2]. Oleh sebab itu, *RapidMiner* biasanya digunakan untuk alat pembelajaran dalam bidang *data mining* [3]

Terdapat 3 (tiga) komponen utama yang paling penting dalam *RapidMiner*, diantaranya *data set*, *attribute and criteria*, *entropy and gain*, dan *accuracy* [4]. *Data set* merupakan suatu Kumpulan data yang terdiri dari beberapa entri. Atribut merupakan variabel dan kriteria merupakan turunan dari atribut. *Entropy* ialah ukuran ketidakpastian dalam data dan *gain* ialah pemisahan data menjadi subset. Sedangkan akurasi adalah

sejauh mana model atau algoritma dalam memprediksi hasil dengan benar [5].

Akta kelahiran memiliki peran penting dalam menentukan status kependudukan seseorang sejak lahir, pernikahan, hingga kematian [6]. Dalam kehidupan masyarakat, akta kelahiran berfungsi untuk alat pangakuan yang diajukan orang-tua [7]. Selain itu, akta kelahiran juga menentukan kewarganegaraan seseorang dan merupakan pencatatan kependudukan yang penting dalam rangka menyusun program perlindungan hukum untuk memastikan kelangsungan hidup anak sebagai pewaris cita-cita bangsa dan negara [8].

Dalam rangka mengetahui jumlah masyarakat yang sudah dan belum terdaftar, maka atribut yang digunakan dalam penelitian ini adalah atribut jenis kelamin, usia, dan kecamatan. Kriteria untuk masing-masing atribut, seperti pada jenis kelamin adalah laki-laki dan perempuan, kriteria pada atribut usia adalah masyarakat yang berusia 0 - 17 tahun dan 18 - 50 tahun. Sedangkan, jumlah kecamatan sebanyak 18 (delapan belas) yang tersebar di Kota Palembang, Sumatera Selatan.

Penelitian yang dilakukan oleh Amado et al., menjelaskan bahwa implementasi algoritma *decision tree* telah memberikan gambaran singkat secara parallel menggunakan teknik seperti *taks parallelism*, *data parallelism*, dan *hybrid parallelism* [9]. Dalam menguji *data set*, maka untuk mengoptimasi pemrosesan data dapat dilakukan dengan *filtering of decision tree* (pemangkasan pohon keputusan) yang terdiri dari proses pengujian, kalibrasi, dan perbaikan [10].

Hasil penelitian yang dilakukan oleh Kabari dan Nwachukwu menunjukkan bahwa *accuracy* dalam pengimplementasian C4.5 sebesar 88%, dimana dapat mendukung dalam pengambilan keputusan. Dalam penelitian ini

memberikan penjelasan bagaimana seorang pelanggan akan bertindak setelah mereka membeli produk yang dijual oleh Perusahaan Kesehatan Mata [11].

Penelitian lainnya yang dilakukan oleh Gupta et al., menunjukkan bahwa *decision tree* yang diujikan kepada 100 siswa menghasilkan tingkat *accuracy* sebesar 91,68%, dimana peneliti ini dapat memeriksa keadaan pembelajaran yang paling efektif untuk siswa [12]. Penelitian yang dilakukan oleh Lugoyava menunjukkan bahwa pengimplementasian *decision tree* telah memunculkan sejumlah atribut informatif dalam menyelesaikan permasalahan data yang tidak seimbang [13].

Penelitian ini memiliki urgensi yang tinggi mengingat pertumbuhan populasi yang signifikan di masyarakat dan peningkatan jumlah data anak penerima akta kelahiran. Penelitian ini bertujuan untuk menerapkan *data mining* dengan algoritma C4.5, mengklasifikasi akta anak yang terdaftar dan tidak terdaftar, menerapkan *decision tree* (pohon keputusan) dalam menciptakan *gain* pada masing-masing atribut, dan pengimplementasian algoritma C4.5 menggunakan bantuan perangkat lunak *RapidMiner 9.5*.

METODE

Rancangan dalam penelitian ini adalah dengan melakukan analisis data agregat akta Disdukcapil Kota Palembang Tahun 2021-2023 yang berjumlah 2.500 data. Luaran yang diharapkan dalam penelitian adalah diketahuinya *gain* terbesar antara atribut jenis kelamin, usia, dan kecamatan. Jumlah data dalam penelitian ini merupakan modifikasi dari penelitian yang dilakukan oleh Naik dan Saman [14], dimana datanya berjumlah 583 data.

Teknik pengumpulan data dilakukan dengan cara wawancara dengan bagian administrasi Disdukcapil Kota Palembang untuk mendapatkan informasi seputar akta masyarakat Palembang dan mengetahui informasi awal mengenai hambatan dalam melakukan *screening* akta.

Sumber data yang dilakukan bersumber dari data primer dan sekunder, dimana data primer didapatkan dari Disdukcapil Kota Palembang berupa data agregat akta dari Tahun 2021-2023. Sementara data sekunder didapatkan dari jurnal, dokumen, dan penelitian yang bisa mendukung penelitian ini.

Analisis data dalam penelitian ini antara lain adalah untuk menghitung nilai *entropy*, *gain*, dan *decision tree* yang dilakukan secara manual menggunakan Microsoft Excel 2021 dan *RapidMiner 9.5*. Sebelum menentukan *decision tree*, maka perlu untuk diketahui terlebih dahulu *root attribute*, *information gain*, *tennis example*, dan *hypothesis space (rules)* [15]

Adapun rumus yang digunakan untuk menghitung nilai *entropy*, *gain*, dan *decision tree* [16], akan dibahas dalam rumus berikut ini:

$$\text{Entropy (S)} = \sum_i^n = 1 - p_i \quad (1)$$

* log 2

Dimana:

S = Himpunan Kasus

n = jumlah partisi S

pi = proporsi Si terhadap S

|Si| = Jumlah kasus pada partisi ke-i

|S| = Jumlah kasus dalam S

$$\text{Gain (S. A)} = \text{Entropy (S)} \quad (2)$$

$$- \sum_i^n = 1$$

* Entropy (Si)

$$\text{Entropy (seluruh kasus)} = \left(- \left(\frac{1768}{2500} \right) * \log_2 \left(\frac{1768}{2500} \right) \right) + \left(- \left(\frac{732}{2500} \right) * \log_2 \left(\frac{732}{2500} \right) \right) = 0,1653797673$$

HASIL DAN PEMBAHASAN

Data Training merujuk pada *data set* yang digunakan dalam proses pembelajaran pada kegiatan *data mining* atau saat membentuk pohon keputusan. Dalam penelitian ini, *data training* dipilih melalui pengambilan sampel yang telah ditentukan sebelumnya. *Data training* yang dilakukan secara *offline* menghasilkan *output* yang sangat lama, sehingga analisis dengan skala horizontal pada *data set* akan mempercepat proses pengujian [17]. Metode ini menghasilkan sebanyak 2.500 data akta anak yang dijadikan sebagai *data training*. Sementara itu, atribut yang dipakai dalam menentukan jumlah anak yang belum memiliki akta anak yang dibagi menjadi 18 (delapan belas) kecamatan di Kota Palembang ini menggunakan atribut jenis kelamin, umur, dan kecamatan. Berdasarkan hasil pengolahan data, ditemukan bahwa sebanyak 1768 data yang berjumlah 'true Ya' (berhasil) dan 732 data yang berjumlah 'true Tidak' (gagal). Alasan penggunaan *RapidMiner* adalah karena perangkat lunak ini dapat menganalisis kumpulan *data set* dalam jumlah yang sangat besar (*big data*) [18]. Selain itu, hasil pengujian oleh *RapidMiner* cenderung kuat dengan bisa menampilkan data perilaku (*behaviour*) dari atribut yang diteliti [19].

Berdasarkan hasil pengolahan data menggunakan rumus *entropy*, maka didapatkan hasil perhitungan sebagai berikut

Berdasarkan hasil perhitungan pada *entropy* tersebut diketahui bahwa seluruh kasus mendapat nilai sebesar 0,1653797673 serta untuk berikutnya dihitung juga nilai *entropy* pada setiap atribut yang terdiri dari 3 (tiga) atribut, yaitu usia, jenis kelamin, dan kecamatan.

Hasil pengolahan data *entropy* dilakukan menggunakan *example set (cross validation)* yang akan disajikan pada tabel berikut ini:

Tabel 1. Entropy kriteria usia

Kriteria usia	Tidak	Ya	Total	Entropy
< = 17 tahun	273	595	868	0.89832
> = 18 tahun	459	1173	1632	0.85715
Total	732	1768	2500	

Setelah melakukan perhitungan pada masing-masing *entropy* kriteria yang terbagi menjadi < = 17 tahun atau masyarakat yang berusia 0 - 17 tahun, dan > = 18 tahun atau masyarakat yang berusia 18 - 50 tahun, didapatkan masing-masing *entropy* adalah sebesar 0.89832 dan 0.85715.

Tabel 2. Entropy kriteria jenis kelamin

Kriteria usia	Tidak	Ya	Total	Entropy
---------------	-------	----	-------	---------

Laki-laki	219	1145	1364	0.63564
Perempuan	513	623	1136	0.99323
Total	732	1768	2500	

Seberang Ulu				
Satu	43	155	198	0.75497
Sematangborang		75	75	0.00000
Total	732	1768	2500	

Setelah melakukan perhitungan pada masing-masing entropy kriteria yang terbagi menjadi laki-laki dan perempuan diperoleh jika masing-masing nilai entropy adalah sebesar 0.63564 dan 0.99323

Tabel 3. Entropy kriteria kecamatan

Kriteria usia	Tidak	Ya	Total	Entropy
Alang-Alang Lebar	36	107	143	0.81404
Bukitkecil		76	76	0.00000
Gandus	29	116	145	0.72193
Iilir Barat Dua	31	102	133	0.78335
Iilir Barat Satu	131	92	223	0.97782
Iilir Timur Dua	32	105	137	0.78419
Iilir Timur Satu	13	19	32	0.97449
Iilir Timur Tiga	18	13	31	0.98115
Jakabaring	30	99	129	0.78244
Kalidoni	128	51	179	0.86206
Kemuning	34	107	141	0.79692
Kertapati	7	105	112	0.33729
Plaju	18	116	134	0.56919
Sako	31	99	130	0.79247
Seberang Ulu Dua	125	54	179	0.88333

Setelah melakukan perhitungan pada masing-masing entropy kriteria yang terbagi menjadi 18 kecamatan diperoleh nilai yang berbeda pada setiap kriteria dengan nilai entropy terendah sebesar 0.00000 dan terbesar 0.98115.

Penelitian yang dilakukan oleh Muttaqien et al., [20] menunjukkan bahwa nilai *entropy* berada di bawah 1 atau < 1 . Hal ini berarti bahwa ada kesesuaian secara teoritis dari hasil penelitian yang telah dilakukan. Hasil perhitungan *entropy* dalam penelitian tersebut, menjelaskan bahwa dari ketiga atribut yang diteliti, yaitu usia, *customer gender*, dan *product taken*, atribut dengan entropy tertinggi adalah usia dengan kriteria 'old' dengan besaran 0,994030211.

Setelah mendapatkan *entropy* pada masing-masing atribut dan kriteria, maka pengolahan dilanjutkan dengan menghitung *gain* untuk menentukan *root node* pada *decision tree*. Adapun perhitungan *gain* akan disajikan pada pembahasan berikut ini:

$$\begin{aligned}
 \text{Gain (usia)} &= 0,1653797673 \\
 &- \left(\left(\frac{868}{2500} \right) * 0,89832 \right) \\
 &+ \left(\left(\frac{1632}{2500} \right) * 0,85715 \right) \\
 &= 1,119961283
 \end{aligned}$$

Berdasarkan hasil perhitungan pada gain (usia) tersebut diperoleh bahwa nilai entropy seluruh kasus adalah sebesar 0,1653797673 dengan jumlah keseluruhan kasus usia sebesar 2500. Jumlah kasus pada nilai atribut usia dengan rentang 0 - 17 tahun atau (≤ 17 tahun) diketahui sebesar 868. Sedangkan pada nilai atribut usia dengan rentang 18 - 50 tahun atau ($\geq 18 < 50$ tahun) sebesar 1632 orang. Oleh karena itu, setelah dilakukan perhitungan didapatkan nilai gain pada atribut usia sebesar 1,119961283.

Gain (jenis kelamin)

$$\begin{aligned} &= 0,1653797673 \\ &- \left(\left(\frac{1364}{2500} \right) * 0.3468056492 \right) \\ &+ \left(\left(\frac{1136}{2500} \right) * 0.4513218323 \right) \\ &= 0.97682696 \end{aligned}$$

Berdasarkan hasil perhitungan pada gain (jenis kelamin) di atas diperoleh bahwa nilai entropy seluruh kasus adalah sebesar 0,1653797673 dengan jumlah keseluruhan kasus usia sebesar 2500. Jumlah kasus pada nilai atribut jenis kelamin Laki-laki sebesar 1364. Sedangkan pada nilai atribut jenis kelamin perempuan sebesar 1136. Oleh karena itu, setelah dilakukan perhitungan didapatkan nilai gain pada atribut jenis kelamin sebesar 0.97682696.

Gain (kecamatan) = 0,1653797673 -

$$\begin{aligned} &\left(\left(\frac{143}{2500} \right) * 0.81404 \right) + \left(\left(\frac{76}{2500} \right) * 0 \right) + \\ &\left(\left(\frac{145}{2500} \right) * 0.72193 \right) + \left(\left(\frac{133}{2500} \right) * 0.78335 \right) + \end{aligned}$$

$$\begin{aligned} &\left(\left(\frac{233}{2500} \right) * 0.97782 \right) + \left(\left(\frac{137}{2500} \right) * 0.78419 \right) = \\ &+ \left(\left(\frac{32}{2500} \right) * 0.97449 \right) + \left(\left(\frac{31}{2500} \right) * \right. \\ &0.98115 \left. \right) + \left(\left(\frac{129}{2500} \right) * 0.78244 \right) + \\ &\left(\left(\frac{179}{2500} \right) * 0.86206 \right) + \left(\left(\frac{141}{2500} \right) * 0.79692 \right) + \\ &\left(\left(\frac{112}{2500} \right) * 0.33729 \right) + \left(\left(\frac{134}{2500} \right) * 0.56919 \right) + \\ &\left(\left(\frac{130}{2500} \right) * 0.79247 \right) + \left(\left(\frac{179}{2500} \right) * 0.88333 \right) + \\ &\left(\left(\frac{198}{2500} \right) * 0.75497 \right) + \left(\left(\frac{75}{2500} \right) * 0 \right) + \\ &\left(\left(\frac{75}{2500} \right) * 0.42232 \right) = 13.83156775 \end{aligned}$$

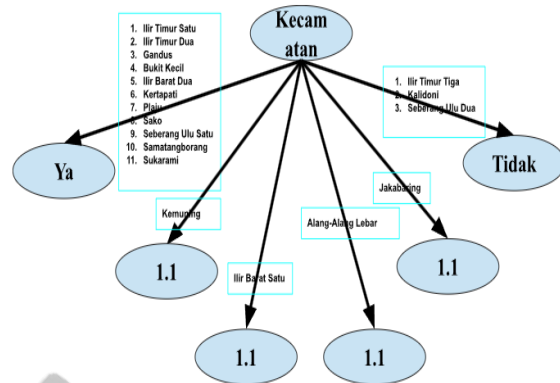
Berdasarkan hasil perhitungan pada gain (kecamatan) di atas diperoleh bahwa nilai entropy seluruh kasus adalah sebesar 0,1653797673 dengan jumlah keseluruhan kasus usia sebesar 2500. Jumlah kasus pada nilai atribut Kecamatan Alang-Alang Lebar sebesar 143. Jumlah kasus pada nilai atribut Kecamatan Bukit Kecil sebesar 76. Jumlah kasus pada nilai atribut Kecamatan Gandus sebesar 145. Jumlah kasus pada nilai atribut Kecamatan Ilir Barat Dua sebesar 133. Jumlah kasus pada nilai atribut Kecamatan Ilir Barat Satu sebesar 223. Jumlah kasus pada nilai atribut Kecamatan Ilir Timur Dua sebesar 137. Jumlah kasus pada nilai atribut Kecamatan Ilir Timur Satu sebesar 32. Jumlah kasus pada nilai atribut Kecamatan Ilir Timur Tiga sebesar 31. Jumlah kasus pada nilai atribut Kecamatan Jakabaring sebesar 129. Jumlah kasus pada nilai atribut Kecamatan Kalidoni sebesar 179. Jumlah kasus pada nilai atribut Kecamatan Kemuning sebesar 141. Jumlah kasus pada nilai atribut Kecamatan Kertapati sebesar 112. Jumlah kasus pada nilai atribut Kecamatan Plaju sebesar 134. Jumlah kasus pada nilai atribut Kecamatan Sako

sebesar 130. Jumlah kasus pada nilai atribut Kecamatan Seberang Ulu Dua sebesar 179. Jumlah kasus pada nilai atribut Kecamatan Seberang Ulu Satu sebesar 198. Jumlah kasus pada nilai atribut Kecamatan Sematangborang sebesar 75. Jumlah kasus pada nilai atribut Kecamatan Sukarami sebesar 303. Oleh karena itu, setelah dilakukan perhitungan didapatkan nilai gain pada atribut kecamatan sebesar 13,83156775.

Setelah seluruh nilai gain pada tiap atribut dihitung, maka dapat disimpulkan jika nilai gain tertinggi adalah pada atribut kecamatan dengan nilai 13.83156775, sementara nilai gain terendah pada atribut jenis kelamin dengan besaran 0.97682696.

Penelitian yang dilakukan oleh menunjukkan terjadinya peningkatan *gains (incrementally gains)* pada algoritma secara presisi dengan tingkatan 80% *accuracy* [21]. Penelitian yang dilakukan oleh Andry et al., [22] menunjukkan terdapat beberapa kriteria yang bisa dijadikan model *root node* dalam *decision tree*, dimana hasil penelitiannya menunjukkan bahwa *root node* ialah *charges*, yang mengindikasikan tiap individu memiliki informasi berbeda dalam biaya asuransi atas perilaku merokoknya. Penelitian lainnya yang dilakukan oleh Wang [23] menunjukkan jika nilai *Gain (S_j,I₁)* sebesar 0,771. Dimana nilai *gain* terbesar dijadikan sebagai *root node*. Meskipun demikian, nilai *gain* tidak terdapat kriteria tertentu, dimana data tersebut harus mencapai ambang nilai 1. Penelitian yang dilakukan oleh Wiza dan Febriadi [24] menunjukkan jika nilai *gain* sebesar 0,31128 yang merupakan *gain* tertinggi dari atribut yang diujikannya.

Sehingga atribut usia menjadi *root node* (*node akar*) yang akan digambarkan pada pohon keputusan berikut ini:



Gambar 1. *Decision Tree* tahap 1

Pada gambar *decision tree* tersebut terlihat beberapa *node* yang termasuk ke dalam kriteria 'ya' dan 'tidak' pada kecamatan-kecamatan yang digunakan dalam penelitian. Pada tabel training akan difilter kembali dengan cara mengambil kriteria laki-laki dan perempuan pada atribut jenis kelamin, serta kriteria ≤ 17 tahun dan ≥ 18 tahun pada atribut usia. Analisis lanjutan pada *node pohon* 1.1 tersebut karena *data set* hasil analisis masih menunjukkan ketidakpastian. Penelitian yang dilakukan oleh Supangat et al., [25] juga melakukan *recalculating* (perhitungan ulang) pada *node pohon* yang masih menunjukkan nilai nol.

Sehingga, akan ditentukan juga atribut yang memiliki nilai dominan untuk *node pohon* keputusan berikutnya:

Tabel 3. *Decision Tree* Tahap Dua

Kecamatan	Atribut	Kriteria	Node Pohon
		≤ 17 Tahun	Ya
Alang-Alang Lebar	Usia	≥ 18 Tahun	Ya

			Analisis decision tree berikutnya. Sehingga berdasarkan Lanjutan hasil filter data training akta anak Kota Palembang akan disajikan pada pohon keputusan berikut:
Jenis Kelamin	Laki-Laki	Analisis Lanjutan	
	Perempuan	Analisis Lanjutan	
Usia	<= 17 Tahun	Ya	
	>= 18 Tahun	Ya	
Jenis Kelamin	Laki-Laki	Analisis Lanjutan	
	Perempuan	Analisis Lanjutan	
Usia	<= 17 Tahun	Analisis Lanjutan	
	>= 18 Tahun	Analisis Lanjutan	
Jenis Kelamin	Laki-Laki	Analisis Lanjutan	
	Perempuan	Analisis Lanjutan	
Jenis Kelamin	Laki-Laki	Analisis Lanjutan	Berdasarkan hasil pengolahan pada Analisis node pohon ke-2, diperoleh jika terdapat 4 Lanjutan kriteria, dimana setiap kriteria memiliki decision tree yang berbeda satu dengan yang Analisis lainnya. Pada decision tree tersebut, dapat Lanjutan dikatakan masyarakat yang berada di Analisis Kecamatan Ilir Barat Satu dengan kriteria usia < Lanjutan = 17 tahun dan >= 18 tahun memiliki akta Analisis kelahiran secara menyeluruh. Namun pada Lanjutan jenis kelamin perlu dilakukan analisis lebih Analisis lanjut dengan cara melakukan filter pada data Lanjutan training. Begitu juga pada Kecamatan Alang-Alang Lebar perlu melakukan analisis lebih Lanjutan lanjut pada kriteria jenis kelamin. Pada kriteria Analisis lainnya dari Kecamatan Kemuning dan Lanjutan Jakabaring secara menyeluruh perlu dilakukan Analisis analisis lebih lanjut yang akan disajikan pada Lanjutan tabel berikut ini.

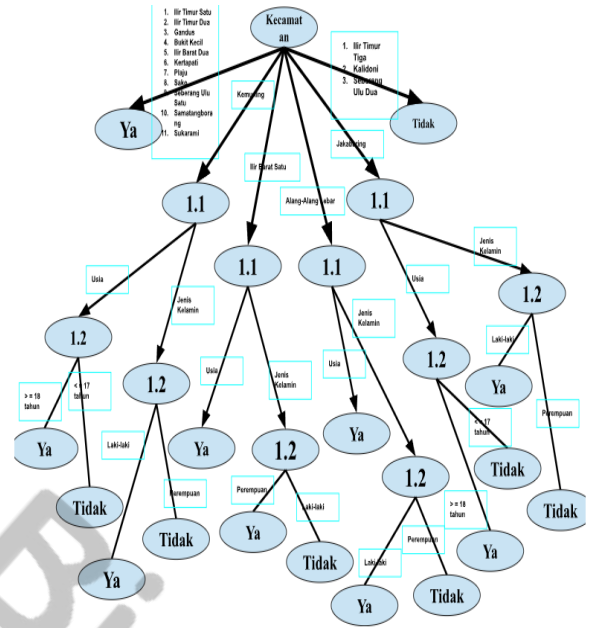
Gambar 2. Decision Tree tahap 2

Berdasarkan tabel perhitungan 'sum ya' dan 'sum tidak' pada tiap kriteria dalam atribut tersebut, maka didapatkan hasil berupa keputusan node pohon yang akan dijadikan

Tabel 3. Decision Tree Tahap Dua

Kecamatan	Kriteria	Sum Ya	Sum Tidak
Alang-Alang Lebar	Laki-Laki	74	0
	Perempuan	0	0
Ilir Barat Satu	Laki-Laki	0	72
	Perempuan	92	59
	<= 17 Tahun	74	0
	>= 18 Tahun	0	0
Jakabaring	Laki-Laki	8	14
	Perempuan	17	16
	<= 17 Tahun	72	0
	>= 18 Tahun	28	19
Kemuning	Laki-Laki	7	15
	Perempuan	28	19

Setelah melakukan filter pada data training, maka dapat diputuskan decision tree pada penelitian ini adalah sebagai berikut:



Gambar 3. Decision Tree tahap 3

Berdasarkan *decision tree* diatas maka *rules* yang didapatkan antara lain adalah: IF (Kecamatan = Ilir Timur Tiga, Kalidoni, Seberang Ulu Dua) THEN Tidak. IF (Kecamatan = Ilir Timur Satu, Ilir Timur Dua, Gandus, Bukit Kecil, Ilir Barat Dua, Kertapati, Plaju, Seko, Seberang Ulu Satu, Samatangborang, Sukarami) THEN Tidak, IF (Kecamatan = Kemuning AND Usia = Lebih atau sama dengan 18 Tahun) THEN Ya, IF (Kecamatan = Kemuning AND Usia = Kurang atau sama dengan 17 Tahun) THEN Tidak, IF (Kecamatan = Kemuning AND Jenis Kelamin = Laki-Laki) THEN Ya, IF (Kecamatan = Kemuning AND Jenis Kelamin = Perempuan) THEN Ya, IF (Kecamatan = Ilir Barat Satu AND Usia = Lebih atau sama dengan 18 Tahun dan Kurang atau sama dengan 17 Tahun) THEN Ya, IF (Kecamatan = Ilir Barat Satu AND Jenis Kelamin = Perempuan) THEN Ya. IF (Kecamatan = Ilir Barat Satu AND Jenis Kelamin = Laki-Laki) THEN Tidak, IF (Kecamatan = Alang-Alang Lebar AND Usia = Lebih atau sama

dengan 18 Tahun dan Kurang atau sama dengan 17 Tahun) THEN Ya, IF (Kecamatan = Alang-Alang Lebar AND Jenis Kelamin = Perempuan) THEN Ya, IF (Kecamatan = Alang-Alang Lebar AND Jenis Kelamin = Laki-Laki) THEN Ya, IF (Kecamatan = Alang-Alang Lebar AND Jenis Kelamin = Perempuan) THEN Tidak, IF (Kecamatan = Jakabaring AND Usia = Lebih atau sama dengan 18 Tahun) THEN Ya, IF (Kecamatan = Jakabaring AND Usia = Kurang atau sama dengan 17 Tahun) THEN Tidak, IF (Kecamatan = Jakabaring AND Jenis Kelamin = Laki-Laki) THEN Ya, IF (Kecamatan = Jakabaring AND Jenis Kelamin = Perempuan) THEN Ya

Penelitian yang dilakukan oleh Kan Lo et al., [26] menunjukkan terdapat tahapan *filtering* dan *cleaning* data dari satu atribut dengan 4 (empat) kriteria. Hasil penelitiannya menunjukkan terdapat perbedaan yang signifikan pada *data set* setelah dilakukannya pemangkasan data.

Berdasarkan decision tree tersebut, diketahui bahwa terdapat beberapa node pohon yang dapat dijadikan pertimbangan oleh Diskdukcapil Kota Palembang dalam menentukan kecamatan mana yang masyarakatnya masih belum mencatatkan akta lahir di Diskdukcapil. Dari hasil temuan diketahui bahwa dari Kecamatan Kemuning yang berusia 0 – 17 tahun dengan jenis kelamin perempuan masih ada sebagian yang belum tercatat. Kemudian, kecamatan lainnya adalah Kecamatan Ilir Barat Satu yang ditemukan bahwa baik masyarakat yang berusia 0 – 17 tahun dan 18 – 50 tahun, dengan jenis kelamin pria masih ada sebagian yang belum mencatatkan dirinya di akta kelahiran Diskdukcapil Kota Palembang. Berbeda dengan Kecamatan Alang-Alang Lebar dimana sebagian masyarakat yang berjenis kelamin perempuan

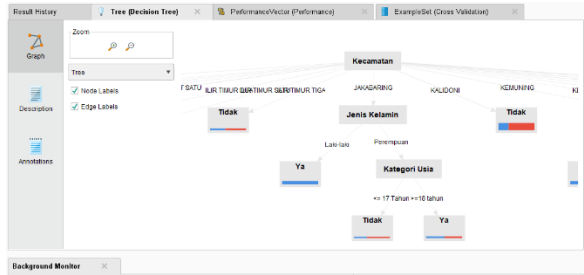
dengan usia 0 – 17 tahun dan 18 – 50 tahun, masih belum ada yang mencatatkan dirinya di Diskdukcapil. Sedangkan pada Kecamatan Jakabaring, masyarakat yang berusia 0 – 17 tahun dengan berjenis kelamin perempuan masih belum banyak yang mencatatkan dirinya di akta kelahiran Diskdukcapil Kota Palembang.

Setelah melakukan perhitungan secara manual, maka langkah berikutnya adalah dengan menguji *data set* menggunakan *RapidMiner 9.5*. Berikut adalah penjelasan mengenai langkah-langkah yang diambil untuk menentukan pohon keputusan menggunakan *RapidMiner*.



Gambar 4. Proses Decision Tree *RapidMiner*

Model algoritma C4.5 yang tergambar pada Gambar di atas digunakan dalam kombinasi dengan proses validasi silang (*cross validation*) untuk menguji decision tree melalui penerapan model dan mengukur kinerjanya (*performance*) menggunakan perangkat *RapidMiner*. Terdapat beberapa langkah dalam *main process* seperti melakukan validasi silang pada *read excel* terhadap *apply model* dan *performance*. Sehingga ketiga proses ini *output* nya adalah *decision tree model* [27].



Gambar 4. *Decision Tree RapidMiner*

Setelah bagian proses tersebut selesai, hasil keluaran dapat ditemukan dalam tampilan hasil (*views result*) pada halaman menu decision tree, di mana pohon keputusan akan ditampilkan dalam bentuk grafik, dan pada menu deskripsi (*menu description*) pada halaman *decision tree*, aturan-aturan yang dihasilkan akan ditampilkan, serupa dengan yang terlihat pada Gambar di atas

Table View Plot View

accuracy: 80.12% +/- 39.92% (micro average: 80.12%)

	true Ya	true Tidak	class precision
pred. Ya	1606	335	82.74%
pred. Tidak	182	397	71.02%
class recall	90.84%	54.23%	

Gambar 4. *Accucary RapidMiner*

Berdasarkan hasil tingkat akurasi (*accuracy*) pada halaman *performance* gambar diatas tersebut diperoleh sebesar 80,12%. Diketahui bahwa tingkat kinerja *accuracy* dengan rentang tersebut (80,12%) menunjukkan kriteria *good and on average* [28]. Kemudian, dihasilkan juga True Positif (TP) yaitu sebanyak 1606 *record* yang diklasifikasikan dengan masyarakat Kota Palembang yang telah mendaftarkan dirinya di akta kelahiran Diskdukcapil dan False Negatif (FN) sebanyak 397 *record* yang berarti bahwa sebanyak jumlah tersebut terdapat masyarakat yang masih belum mendaftarkan dirinya di akta

kelahiran Diskdukcapil Kota Palembang. Selanjutnya, dari hasil penelitian diperoleh jika True Negatif (TN) sebesar 162 *record* yang berarti masyarakat Kota Palembang kemungkinan telah mendaftarkan diri di Diskdukcapil, namun tidak di *acc* oleh Diskdukcapil. Sementara itu sebanyak 335 diidentifikasi sebagai False Positive (FP) yang dikategorikan sebagai masyarakat yang telah mendaftarkan diri di Diskdukcapil, namun masih belum diproses atau masih dalam proses validasi oleh pihak Diskdukcapil Kota Palembang, sehingga memiliki risiko kegagalan.

Model algoritma yang dianalisis akan menghasilkan sebuah *output* FP dan TP, dimana keduanya adalah hasil pengujian C4.5 pada *data set* yang diterima. Hal ini berlaku juga pada penelitian yang dilakukan oleh Tembhrne et al., [29] yang menunjukkan bahwa False Positive (FP) lebih besar dibandingkan dengan True Positive (TP), sehingga hasil penelitiannya menunjukkan bahwa terdapat individu yang memiliki pengingatan lebih rendah dibandingkan dengan individu yang memiliki pengingatan yang tinggi. Penelitian yang dilakukan oleh Gosh et al., [30] menunjukkan terdapat 72 *record* dengan TP, 30 *record* dengan TN, yang mengindikasikan jika *data set* yang diolahnya menghasilkan data yang sesuai dengan asumsinya. Penelitian yang dilakukan oleh Vyas dan Uma [31] menunjukkan terdapat 82,61 % *precision*, 40,23% *recall*, dan 79,08% *accuracy*, dimana tingkatan akurasi dari penelitiannya cukup serupa dengan akurasi hasil yang dilakukan oleh peneliti.

Hasil penelitian yang dilakukan oleh Wang et al., [32] menunjukkan bahwa interval yang dinyatakan TP hanya sebesar 0,93%, dimana *data set* yang diteliti menunjukkan tingkat akurasi yang begitu tinggi. Penelitian dari Joshuva et al., [33], menunjukkan jika

tingkatan akurasi dari 3 atribut yang diturunkan menjadi 12 kriteria berada pada rentang 44% - 60%, dimana kriteria *stochastic gradient descent* merupakan kriteria dengan nilai akurasi terbesar dengan nilai 60,67%.

Namun demikian, hasil penelitian yang telah dilakukan menunjukkan hasil yang berbeda dengan penelitian yang dilakukan oleh Mantik et al., [34] yang menunjukkan bahwa kurang optimalnya klasifikasi data yang dilakukan, dimana persentase akurasi (*accuracy*) menunjukkan kurang dari 50%. Tingkat akurasi yang lemah ini kemungkinan besar karena *data set* masih belum di *cleaning* dan *filtering* sehingga tingkat akurasi yang didapatkan cenderung lemah. Meskipun sebenarnya terdapat cara yang bisa dilakukan untuk meningkatkan performa analisis C4.5 dengan cara mengonfigurasi *data set* melalui Hadoop sehingga akan mengurangi kesenjangan data yang didapat [35].

PENUTUP

Penerapan data mining pada objek Disdukcapil yang terdiri dari atribut usia dengan kriteria laki-laki dan perempuan, atribut jenis kelamin dengan kriteria 0 - 17 tahun dan 18 - 50 tahun, dan atribut kecamatan yang terbagi menjadi 18 (delapan belas) kecamatan yang ada di Kota Palembang.

Klasifikasi akta anak yang telah terdaftar berdasarkan uji rapidminer pada halaman *accuracy* menunjukkan bahwa sebanyak 1.606 masyarakat dikategorikan sebagai True Positif (TP), dimana diklasifikasikan dengan masyarakat Kota Palembang yang telah mendaftarkan dirinya di akta kelahiran Diskdukcapil dan False Negatif (FN) sebanyak 397 record yang berarti bahwa sebanyak jumlah tersebut terdapat masyarakat

yang masih belum mendaftarkan dirinya di akta kelahiran Disdukcapil Kota Palembang.

Berdasarkan penerapan decision tree diketahui bahwa gain terbesar adalah kecamatan dengan besaran 13.83156775. Kemudian, root node tersebut dianalisis kembali hingga menghasilkan 3 (tiga) kecamatan yang memiliki keputusan 'tidak', 4 (empat) kecamatan yang perlu dilanjutkan dengan analisis pohon keputusan, dan 10 (sepuluh) kecamatan yang memiliki keputusan 'ya'. 4. Data testing yang digunakan kemudian diuji menggunakan aplikasi RapidMiner dan menunjukkan adanya kesesuaian antara perhitungan manual dengan perhitungan menggunakan software RapidMiner 9.5. Diketahui bahwa tingkat *accuracy* menunjukkan besaran 80,12% yang mengindikasikan bahwa sebagian besar data menunjukkan tidak adanya error.

Pada saran teoritis, peneliti berikutnya dapat menambahkan atribut lain, seperti data keluarga yang bercerai. Sehingga tingkat *accuracy* dari pengolahan data menggunakan RapidMiner dapat bertambah. Hal ini karena asumsi dari peneliti jika adanya sampling yang masih error karena adanya data yang masih belum diproses akibat adanya gangguan dari sistem informasi Disdukcapil tersebut. Sementara itu, saran praktis untuk Disdukcapil adalah melakukan sosialisasi dan pencatatan pada kecamatan yang dianggap masih memiliki gain yang rendah dan termasuk ke dalam keputusan 'tidak', seperti Kecamatan Ilir Timur Tiga, Kalidoni, Seberang Ulu Dua.

DAFTAR PUSTAKA

- [1] E. D. Madyatmadja, D. J. M. Sembiring, S. M. Br Perangin Angin, D. Ferdy, dan

- J. F. Andry, "Big Data in Educational Institutions using RapidMiner to Predict Learning Effectiveness," *Journal of Computer Science*, vol. 17, no. 4, hlm. 403–413, 2021, doi: 10.3844/jcssp.2021.403.413.
- [2] M. R. Nahjan, N. Heryana, dan A. Voutama, "IMPLEMENTASI RAPIDMINER DENGAN METODE CLUSTERING K-MEANS UNTUK ANALISA PENJUALAN PADA TOKO OJ CELL," 2023.
- [3] V. R. Prasetyo, H. Lazuardi, A. A. Mulyono, dan C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Linear Regression," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 7, no. 1, hlm. 8–17, Mei 2021, doi: 10.25077/teknosi.v7i1.2021.8-17.
- [4] P. S. Chauha, "Opinion Mining and Sentiment Analysis using Rapidminer," Vienna, 2016.
- [5] S. Marzukhi, N. Awang, S. N. Alsagoff, dan H. Mohamed, "RapidMiner and Machine Learning Techniques for Classifying Aircraft Data," dalam *Journal of Physics: Conference Series*, IOP Publishing Ltd, Agu 2021. doi: 10.1088/1742-6596/1997/1/012012.
- [6] F. S. Napitupulu, I. S. Damanik, S. Saragih, dan A. Wanto, "Algoritma K-Means untuk Pengelompokan Dokumen Akta Kelahiran pada Tiap Kecamatan di Kabupaten Simalungun," *Technology and Science (BITS)*, vol. 2, no. 1, hlm. 55–63, 2020.
- [7] E. L. Nasution, "Perlindungan Hukum Melalui Akte Kelahiran Terhadap Anak Yang Tidak Diketahui Asal-Usulnya," 2017.
- [8] S. Indrawati, "Pembuatan Akta Kelahiran Anak yang Orang-tuanya Belum Memiliki Akta Perkawinan di Kota Denpasar," *Prodi Hukum Keperdataan*, 2016.
- [9] N. Amado, J. Gama, dan F. Silva, "Parallel Implementation of Decision Tree Learning Algorithms."
- [10] Y. Hou, "Decision Tree Algorithm for Big Data Analysis," 2018.
- [11] L. G. Kabari dan E. O. Nwachukwu, "Decision Support System Using Decision Tree and Neural Networks," 2013. [Daring]. Tersedia pada: www.iiste.org
- [12] S. Gupta, P. Kumar, dan R. Tekchandani, "A machine learning-based decision support system for temporal human cognitive state estimation during online education using wearable physiological monitoring devices," *Decision Analytics Journal*, hlm. 100280, Sep 2023, doi: 10.1016/j.dajour.2023.100280.

- [13] N. Lugovaya, "Application of Decision Trees to Detect Process Disruptions in Aluminum Production *," 2020.
- [14] A. Naik dan L. Samant, "Correlation Review of Classification Algorithm Using Data Mining Tool: WEKA, Rapidminer, Tanagra, Orange and Knime," dalam *Procedia Computer Science*, Elsevier B.V., 2016, hlm. 662–668. doi: 10.1016/j.procs.2016.05.251.
- [15] D. Roth, B. Zhou, dan C. Cervantes, "Decision Trees," 2016.
- [16] N. Yang, T. Li, dan J. Song, "Construction of Decision Trees based Entropy and Rough Sets under Tolerance Relation," Atlantis Press, 2007. doi: 10.2991/iske.2007.258.
- [17] D. Singh dan C. K. Reddy, "A survey on platforms for big data analytics," *J Big Data*, vol. 2, no. 1, Des 2015, doi: 10.1186/s40537-014-0008-6.
- [18] V. S. Agneeswaran, P. Tonpay, dan J. Tiwary, "Paradigms for realizing machine learning algorithms," *Big Data*, vol. 1, no. 4. Mary Ann Liebert Inc., hlm. 207–214, 1 Desember 2013. doi: 10.1089/big.2013.0006.
- [19] K. Chen, F. Abtahi, J.-J. Carrero, C. Fernandez-Llatas, dan F. Seoane, "Process mining and data mining applications in the domain of chronic diseases: A systematic review," *Artif Intell Med*, hlm. 102645, Agu 2023, doi: 10.1016/j.artmed.2023.102645.
- [20] R. Muttaqien, M. Galih, dan A. Pramuntadi, "Implementation of Data Mining Using C4.5 Algorithm for Predicting Customer Loyalty of PT. Pegadaian (Persero) Pati Area Office," 2021. [Daring]. Tersedia pada: <https://ijcis.net/index.php/ijcis/index>JournalIJCIHomepage-<https://ijcis.net/index.php/ijcis/index>
- [21] T. M. Fonseka, V. Bhat, dan S. H. Kennedy, "The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors," *Australian and New Zealand Journal of Psychiatry*, vol. 53, no. 10. SAGE Publications Inc., hlm. 954–964, 1 Oktober 2019. doi: 10.1177/0004867419864428.
- [22] J. F. Andry, H. Hartono, Honni, A. Chakir, dan Rafael, "Data Set Analysis Using Rapid Miner to Predict Cost Insurance Forecast with Data Mining Methods," *Journal of Hunan University Natural Sciences*, vol. 49, no. 6, hlm. 167–175, Jun 2022, doi: 10.55463/issn.1674-2974.49.6.17.
- [23] J. Wang, "Application of C4.5 Decision Tree Algorithm for Evaluating the College Music Education," *Mobile Information Systems*, vol. 2022, 2022, doi: 10.1155/2022/7442352.
- [24] F. Wiza dan B. Febriadi, "Classification Analysis Using C4.5 Algorithm To Predict The Level of Graduation of Nurul Falah Pekanbaru

- High School Students,” *International Journal of Information System & Technology*, vol. 2, no. 2, hlm. 43–52, 2019.
- [25] A. Pratama dan T. Rahmawati, “Implementation of C4.5 Algorithm for Analysis of Service Quality in Companies of PT. XYZ,” 2021.
- [26] C.-K. Lo, H.-C. Chen, P.-Y. Lee, M.-C. Ku, L. Ogiela, dan C.-H. Chuang, “Smart Dynamic Resource Allocation Model for Patient-Driven Mobile Medical Information System Using C4.5 Algorithm,” 2019. [Daring]. Tersedia pada: <http://www.journal.uestc.edu.cn>.
- [27] H. Amalia ; Yunita, ; Ari Puspitasari, ; Ade, F. Lestari, dan C. Author, “STUDENT PERFORMANCE ANALYSIS USING C4.5 ALGORITHM TO OPTIMIZE SELECTION.”
- [28] E. D. Madyatmadja, S. I. Jordan, dan J. F. Andry, “Big data analysis using rapidminer studio to predict suicide rate in several countries,” *ICIC Express Letters, Part B: Applications*, vol. 12, no. 8, hlm. 757–764, Agu 2021, doi: 10.24507/icicelb.12.08.757.
- [29] D. Sadanand Tembhurne, J. Adhikari, dan P. Babu, “ Implementation of Data Mining Techniques in CRM of Pharmaceutical Industry,” *IOSR Journal of Engineering*, [Daring]. Tersedia pada: www.iosrjen.org
- [30] S. Ghosh, S. M. Thamke, dan U. R. S. Kalyani, “International Journal on Recent and Innovation Trends in Computing and Communication Sentiment Analysis using Rapid Miner for Polarity Dataset”, [Daring]. Tersedia pada: <http://www.ijritcc.org>
- [31] V. Vyas dan V. Uma, “An Extensive study of Sentiment Analysis tools and Binary Classification of tweets using Rapid Miner,” dalam *Procedia Computer Science*, Elsevier B.V., 2018, hlm. 329–335. doi: 10.1016/j.procs.2017.12.044.
- [32] X. Wang, C. Zhou, dan X. Xu, “Application of C4.5 decision tree for scholarship evaluations,” dalam *Procedia Computer Science*, Elsevier B.V., 2019, hlm. 179–184. doi: 10.1016/j.procs.2019.04.027.
- [33] A. Joshuva, R. Sathish Kumar, S. Sivakumar, G. Deenadayalan, dan R. Vishnuvardhan, “An insight on VMD for diagnosing wind turbine blade faults using C4.5 as feature selection and discriminating through multilayer perceptron,” *Alexandria Engineering Journal*, vol. 59, no. 5, hlm. 3863–3879, 2020.
- [34] N. Abdillah dan M. Ihksan, “APPLICATION OF THE C4.5 ALGORITHM FOR CLASSIFICATION OF MEDICAL RECORD DATA AT M.DJAMIL HOSPITAL BASED ON THE

INTERNATIONAL DISEASE CODE,”
Jurnal Mantik, vol. 6, no. 1, 2022.

- [35] H. Bin Wang dan Y. J. Gao, “Research on C4.5 algorithm improvement strategy based on MapReduce,” dalam *Procedia Computer Science*, Elsevier B.V., 2021, hlm. 160–165. doi: 10.1016/j.procs.2021.02.045.

