

Penerapan Algoritma C4.5 pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara

Hariati^{*1}, Masna Wati², Bambang Cahyono³

^{1,2,3}Jurusan Teknologi Informasi dan Komunikasi, Universitas Mulawarman, Samarinda
e-mail: ^{*1}riri.hrty@gmail.com, ²masna.ssi@gmail.com, ³cbambang86@gmail.com

Abstrak

Dalam menanggulangi masalah kemiskinan, Bupati Kutai Kartanegara melalui Dinas Sosial menyediakan berbagai macam program bantuan, dimana setiap program bantuan tersebut memiliki kriteria yang berbeda-beda dalam penentuan penerimanya. Berhubungan dengan banyaknya program bantuan yang ada dengan kriteria yang berbeda-beda dalam penentuan penerimaan program bantuan, tentunya staf pemerintah daerah akan kesulitan dalam melakukan proses seleksi. Berdasarkan data histori penerima bantuan dapat dibuat rekomendasi klasifikasi penerima bantuan yang akan membantu staf pemerintah. Klasifikasi dapat dilakukan dengan menggunakan algoritma C4.5. Penelitian ini bertujuan membentuk model klasifikasi pohon keputusan untuk menentukan penerima program bantuan dan menerapkannya dalam pembangunan sistem klasifikasi penerima bantuan. Hasil klasifikasi dievaluasi dengan confusion matrix dan kurva ROC (Receiver Operating Characteristics)/AUC (Area Under Cover). Berdasarkan hasil klasifikasi menunjukkan bahwa akurasi yang diperoleh mencapai 97,14% pada data training 85% dengan data 105 dan nilai AUC 0,913 yang termasuk dalam excellent classification, hal ini menunjukkan bahwa algoritma C4.5 cocok digunakan untuk menentukan penerima program bantuan dan rule hasil klasifikasi dapat diimplementasikan dalam pembuatan sistem klasifikasi penerima bantuan.

Kata Kunci: Decision Tree, Data Mining, Algoritma C4.5, Klasifikasi

1. PENDAHULUAN

Di dalam kehidupan sehari-hari, manusia selalu dihadapkan oleh berbagai macam masalah dari berbagai macam bidang. Termasuk dalam bidang pemerintahan terutama masalah peningkatan kesejahteraan sosial masyarakat. Penyelenggaraan kesejahteraan sosial diprioritaskan kepada mereka yang memiliki kehidupan yang tidak layak secara kemanusiaan dan memiliki kriteria masalah sosial, seperti kemiskinan, ketelantaran, kecacatan, keterpencilan, ketunaan sosial dan penyimpangan perilaku, korban bencana, atau korban tindak kekerasan, eksploitasi serta diskriminasi [1]. Berdasarkan sensus penduduk tahun 2015, jumlah penduduk Kabupaten Kutai Kartanegara mencapai 717.789 jiwa. Dengan jumlah penduduk yang cukup banyak tersebut tentunya tidak lepas dari masalah peningkatan kesejahteraan masyarakat terutama masalah penduduk miskin. Hal ini dapat dilihat dari data persentase angka kemiskinan pada tahun 2015 sebesar 6,27% dari jumlah penduduk sebanyak 717.789 jiwa [2]. Untuk menanggulangi masalah kemiskinan tersebut, Bupati Kutai Kartanegara melalui Dinas Sosial menyediakan berbagai macam program bantuan, salah satunya program bantuan lanjut usia, dimana setiap program bantuan tersebut memiliki kriteria yang berbeda-beda dalam penentuan penerimanya. Berhubungan dengan banyaknya

program bantuan yang ada dengan kriteria yang berbeda-beda tersebut, tentunya staf pemerintah daerah akan kesulitan dalam melakukan proses seleksi. Berdasarkan permasalahan tersebut, maka diperlukan sebuah metode untuk mempermudah dalam penentuan penerima program bantuan agar masyarakat yang mendapatkan bantuan adalah penerima yang benar-benar membutuhkan atau tepat sasaran.

Metode yang akan digunakan pada penelitian ini dengan penerapan *data mining*, teknik yang dapat dimanfaatkan dalam proses tersebut adalah klasifikasi. Klasifikasi dapat menentukan penerima program bantuan berdasarkan variabel-variabel yang berpengaruh dengan menggunakan algoritma tertentu dan algoritma yang digunakan adalah algoritma C4.5. Dengan menggunakan algoritma C4.5 diharapkan dapat memberikan hasil suatu model pohon keputusan yang dapat memprediksi keluarga yang layak menerima program bantuan serta dapat membantu staf pemerintah daerah dalam menentukan keputusan.

Berdasarkan uraian di atas maka rumusan masalah yang dapat diambil dari penelitian ini yaitu, 1) bagaimana menerapkan algoritma C4.5 *Decision Tree* dalam pembuatan aplikasi untuk menentukan penerima program bantuan pemerintah daerah? dan 2) seberapa besar persentase akurasi algoritma C4.5 *Decision Tree* dalam menentukan penerima program bantuan pemerintah daerah?. Tujuan dari penelitian ini yaitu: 1) menerapkan algoritma C4.5 *Decision Tree* dalam pembuatan aplikasi untuk menentukan penerima program bantuan pemerintah daerah, dan 2) mengetahui seberapa besar persentase akurasi algoritma C4.5 *Decision Tree* dalam menentukan penerima program bantuan pemerintah daerah. Manfaat dari penelitian ini yaitu: 1) bagi staf pemerintah daerah, memberikan kemudahan kepada staf pemerintah daerah dalam menentukan penerima program bantuan, 2) bagi mahasiswa, dapat memberikan tambahan referensi, pengalaman, dapat mengetahui serta membantu dalam proses penentuan penerima program bantuan, mengetahui alasan diterima atau ditolaknya sebuah program yang diberikan kepada keluarga tertentu, serta dapat pula digunakan sebagai literatur dalam pembangunan sistem selanjutnya, dan 3) bagi masyarakat, dana bantuan dari pemerintah dapat diberikan ke masyarakat yang benar-benar membutuhkan bantuan.

2. METODE PENELITIAN

2.1 Landasan Teori

2.1.1 Data Mining

Data mining merupakan sebuah proses terpadu dari analisis data yang terdiri dari serangkaian kegiatan yang berjalan berdasarkan pada pendefinisian tujuan dari apa yang akan dianalisis sampai pada interpretasi dan evaluasi hasil [3]. *Data mining*, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). KDD adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Tujuan *data mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat [4]. Alasan utama mengapa *data mining* diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan pengetahuan yang berguna. Informasi dan pengetahuan yang didapat tersebut dapat digunakan pada banyak bidang, mulai dari manajemen bisnis, kontrol produksi, kesehatan, pendidikan, dan lain-lain [5]. *Knowledge Discovery* dalam sebuah basis data melalui beberapa tahapan untuk mengolah data yang akan digunakan untuk pengetahuan baru, tahapan-tahapan tersebut yaitu, *data cleaning*, *data integration*, *data selection*, *data transformation*, *data mining*, *pattern evaluation*, dan *knowledge representation*.

2.1.2 Algoritma C4.5

Algoritma C4.5 membuat pohon keputusan dari atas ke bawah, dimana atribut paling atas merupakan akar, dan yang paling bawah dinamakan daun [6]. Tahapan dalam membuat pohon keputusan dengan algoritma C4.5 yaitu [7]:

1. Mempersiapkan data *training*, dapat diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing variabel. Sebelumnya dihitung terlebih dahulu nilai *entropy*, dengan persamaan (1).

$$\text{Entropy (S)} = \sum_{i=1}^n - p_i * \text{Log}_2 p_i \dots \dots \dots (1)$$

Dengan S adalah himpunan kasus, dan p adalah proporsi dari kelas dalam *output*.

3. Mengitung nilai *gain* dengan persamaan (2).

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropi(S)} \dots \dots \dots (2)$$

Dengan S adalah himpunan kasus, A adalah variabel kasus, |S_i| adalah jumlah kasus pada partisipasi ke i, dan |S| adalah jumlah kasus dalam himpunan kasus.

4. Mengulangi langkah ke-2 hingga semua *record* terpartisi. Proses partisi pohon akan berhenti disaat:
 - a. Semua kasus pada cabang memiliki kelas yang sama.
 - b. Tidak ada variabel dalam record yang dipartisi lagi.

2.2 Metode Penelitian

Metode penelitian yang diterapkan ini menggunakan tahapan data mining atau sering juga disebut sebagai *Knowledge Discovery in Database* (KDD), dengan langkah-langkah sebagai berikut:

1. *Data cleaning* dan *integration*. Dari 209 data yang dikumpulkan ditemukan 86 data yang harus dibersihkan karena data tidak konsisten atau data tidak relevan dan akan mengganggu pola aturan algoritma yang akan dibentuk.
2. *Data selection*. Data yang diperoleh masih berupa data keseluruhan program bantuan, maka dilakukan pemilihan data berupa data lansia wilayah Kota Bangun.
3. *Data transformation*, adalah fase di mana data yang dipilih diubah menjadi bentuk yang sesuai untuk prosedur *mining*. Untuk meningkatkan akurasi pada algoritma maka data dan variabel yang telah dikumpulkan akan ditransformasi.
4. *Data mining*. Algoritma yang akan digunakan dalam penelitian ini adalah algoritma C4.5 untuk pengklasifikasian data calon penerima bantuan.
5. *Pattern evaluation*. Pengukuran akurasi dalam penelitian ini akan menggunakan metode *confusion matrix*. Dalam tahapan ini akan dilakukan pengukuran keakuratan hasil dengan menggunakan persamaan (2.3).
6. *Knowledge representation*. Pada tahap ini merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna. Ada kalanya hal ini harus melibatkan orang-orang yang tidak memahami data mining. Karenanya presentasi hasil data mining dalam bentuk pengetahuan yang bias dipahami semua orang adalah satu tahapan yang diperlukan dalam proses data mining. Dalam presentasi ini, hasil berupa *decision tree* yang berisi beberapa kaidah dan *rule* yang akan diimplementasikan ke dalam bentuk aplikasi berbasis desktop.

Setelah melalui tahapan *Knowledge Discovery in Database* (KDD) di atas, dilanjutkan dengan pembentukan pohon keputusan. Secara umum alur proses algoritma C4.5 untuk membangun pohon keputusan dalam *data mining* adalah sebagai berikut:

1. Mempersiapkan data *training*, dapat diambil dari data history program bantuan pemerintah yang pernah terjadi sebelumnya dan sudah dikelompokkan dalam kelas-kelas tertentu.

2. Menentukan akar yang berupa variabel dari pohon dengan menghitung nilai *gain* yang tertinggi dari masing-masing variabel. Sebelumnya dihitung terlebih dahulu nilai *index entropy*, dengan rumus seperti tertera dalam persamaan (1).
3. Mengitung nilai *gain* dengan rumus seperti tertera dalam persamaan (2).
4. Mengulangi langkah ke-2 hingga semua *record* terpartisi. Proses partisi pohon akan berhenti disaat:
 - a. Semua kasus pada cabang memiliki kelas yang sama (Kelas status yaitu diterima atau tolak).
 - b. Tidak ada variabel dalam record yang dipartisi lagi.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Penelitian ini menggunakan data dengan jumlah 123 yang diambil dari data calon penerima bantuan tahun 2014 yang diperoleh dari bidang kesejahteraan sosial Dinas Sosial Kabupaten Kutai Kartanegara. Data yang ada kemudian akan digunakan untuk membentuk 5 kelas data *training* dan data *testing* C4.5 dengan pembagian jumlah data *training* sebesar 55%, 65%, 75%, 85%, dan 95% kemudian akan dicari tingkat akurasi tertinggi dari kelima kelas tersebut dan akan di implementasikan ke sistem.

3.2 Pembentukan Model

Pembentukan aturan pada algoritma C4.5 didapatkan dengan membentuk pohon keputusan berdasarkan nilai *gain* tertinggi dari setiap atribut variabel terhadap nilai kelas target. Dari 7 variabel yang telah ditetapkan, 6 variabel merupakan variabel kelas dengan jenis kategori (umur, jenis kelamin, status perkawinan, pendidikan terakhir, pekerjaan, keterampilan), dan 1 variabel target (status). Perhitungan nilai *gain* untuk node awal (akar) dipandang perlu untuk dilakukan secara manual untuk membuktikan bahwa alat permodelan *data mining* yang digunakan sesuai dengan aturan yang digunakan oleh peneliti. Berikut perhitungan nilai *entropy* dan *gain* untuk data training 85% dengan jumlah data 105.

Dari 105 data *training*, ditemukan jumlah kelas yang diterima sebanyak 53 dan jumlah kelas yang ditolak sebanyak 52. Kemudian dicari nilai *entropy* label target menggunakan persamaan (1) dengan hasil sebagai berikut:

$$\begin{aligned} \text{Entropy (Total)} &= (-52/105 \cdot \log_2(52/105)) + (-53/105 \cdot \log_2(53/105)) \\ &= 0,999934571 \end{aligned}$$

Kemudian dihitung nilai *entropy* setiap atribut menggunakan persamaan (1), dengan contoh perhitungan *entropy* untuk atribut umur, sebagai berikut:

$$\begin{aligned} E_{\leq 60} [14,0] &= 0 \\ E_{> 60} [38,53] &= (-38/91 \cdot \log_2(38/91)) + (-53/91 \cdot \log_2(53/91)) \\ &= 0,980310798 \end{aligned}$$

Berdasarkan nilai *entropy* yang telah diketahui kemudian dapat dicari node awal (akar), dengan menggunakan persamaan (2). Contoh perhitungan *gain* untuk umur adalah sebagai berikut.

$$\begin{aligned} \text{Gain (umur)} &= 0,999934571 - ((14/105 \cdot 0) + (91/105 \cdot 0,980310798)) \\ &= 0,150331879 \end{aligned}$$

Perhitungan selanjutnya untuk variabel jenis kelamin, status perkawinan, pendidikan terakhir, pekerjaan, dan keterampilan, menggunakan langkah yang sama. Diketahui dari perhitungan yang telah dilakukan secara manual nilai *gain* tertinggi untuk node awal ditemukan pada variabel pekerjaan = "buruh pertanian tidak tetap" dengan nilai *gain* sebesar 0,542275437 dari 105 data *training* yang digunakan. Nilai *gain* dan *entropy* dengan 85% data training dapat dilihat pada tabel 1.

Tabel 1. Nilai *Gain* dan *Entropy* dengan 85% Data Training

1	Node Akar	Total	Tolak	Diterima	Entropy	Gain
	Total	105	52	53	0,999934571	
	Umur					0,150331879
	<= 60	14	14	0	0	
	> 60	91	38	53	0,980310798	
	Jenis Kelamin					0,140093994
	Laki-laki	44	33	11	0,811278124	
	Perempuan	61	19	42	0,894869231	
	Status Perkawinan					0,048597506
	Kawin	64	37	27	0,982316608	
	Belum Kawin	8	4	4	1	
	Cerai Hidup	5	3	2	0,970950594	
	Cerai Mati	28	8	20	0,863120569	
	Pendidikan Terakhir					0,012306604
	Belum/Tidak Tamat SD/Sederajat	42	18	24	0,985228136	
	SD/MI/Sederajat	60	33	27	0,992774454	
	SLTP/MTs/Sederajat	3	1	2	0,918295834	
	Pekerjaan					0,542275437
	Buruh pertanian tidak tetap	18	15	3	0,650022422	
	Buruh tidak tetap non pertanian	9	9	0	0	
	Usaha sendiri	17	17	0	0	
	Usaha dengan buruh tetap/tidak tetap	2	2	0	0	
	Tidak bekerja	59	9	50	0,616166193	
	Keterampilan					0,125849122
	Bidang pertanian/perikanan	52	35	17	0,911751759	
	Pertukangan	1	1	0	0	
	Kerajinan tangan	5	3	2	0,970950594	
	Lainnya	46	13	33	0,858981037	
	Menjahit	1	0	1	0	

3.3 Evaluasi dan Validasi Hasil

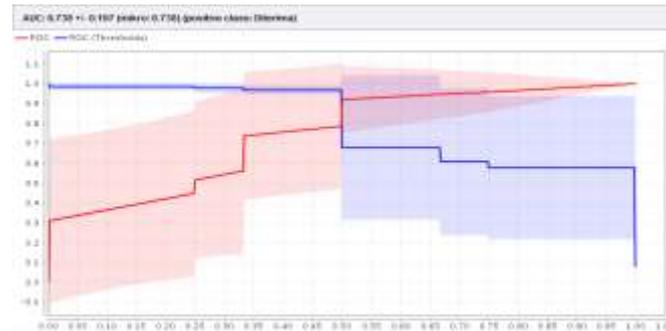
Setelah dilakukan beberapa percobaan, maka akan dilakukan pengujian tingkat akurasi dari tiap-tiap model untuk melihat kinerja dari metode Algoritma C4.5 dan kemudian akan dipilih salah satu model dengan akurasi tertinggi untuk diimplementasikan kedalam aplikasi. Evaluasi tingkat akurasi Algoritma dilakukan dengan menggunakan *confussion matrix* dan kurva ROC (*Receiver Operating Characteristics*) /AUC (*Area Under Cover*). Berikut adalah pengujian tingkat akurasi algoritma C4.5 dengan *RapidMiner*.

1. Akurasi Algoritma C4.5 dengan 55% data training

Tabel 2. *Confusion matrix* algoritma C4.5

accuracy: 79.29% +/- 17.37% (mikro: 79.41%)			
	true Tolak	true Diterima	class precision
pred. Tolak	26	7	78.79%
pred. Diterima	7	28	80.00%
class recall	78.79%	80.00%	

Dari tabel 2 diatas, hasil *confusion matrix* dari 68 data *training* yang diberikan pada Algoritma C4.5 menunjukkan bahwa persentase akurasi algoritma C4.5 dengan 55% data *training* adalah sebesar 79,41%. Tingkat akurasi Algoritma juga dapat dilihat pada kurva ROC/AUC berikut ini.



Gambar 1. Kurva ROC/AUC dengan 55% data training

Kurva ROC dibuat berdasarkan nilai yang telah didapatkan dari *confusion matrix* yaitu antara *False Positive Rate* dengan *True Positive Rate*. Selain itu, kinerjanya juga dapat dilihat dari luas dibawah kurva atau disebut *Area Under Cover* (AUC) yang jika luasnya lebih besar maka menunjukkan kinerjanya bagus. ROC memiliki tingkat nilai diagnosa yaitu:

- Akurasi bernilai 0,90 - 1,00 = *excellent classification*
- Akurasi bernilai 0,80 - 0,90 = *good classification*
- Akurasi bernilai 0,70 - 0,80 = *fair classification*
- Akurasi bernilai 0,60 - 0,70 = *poor classification*
- Akurasi bernilai 0,50 - 0,60 = *failure classification*

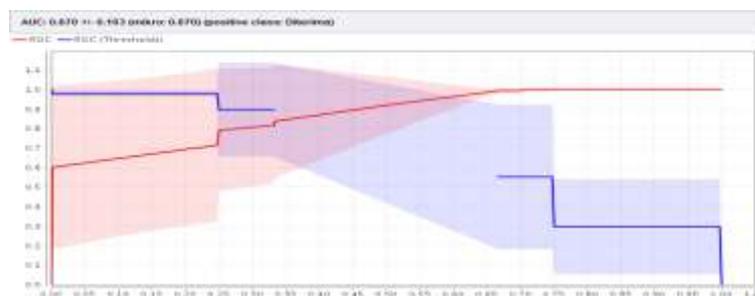
Berdasarkan klasifikasi tersebut maka dapat disimpulkan bahwa algoritma C4.5 dengan 55% data *training* atau sebanyak 68 data yang digunakan memiliki nilai keakurasian AUC sebesar 0,738 dan termasuk dalam *fair classification* untuk melakukan klasifikasi calon penerima bantuan lansia.

2. Akurasi Algoritma C4.5 dengan 65% data training

Tabel 3. *Confusion matrix* algoritma C4.5

accuracy: 83,75% +/- 13,75% (mikro: 83,75%)			
	true Total	true Ditema	class precision
pred. Tolak	33	6	80,48%
pred. Ditema	5	34	87,18%
class recall	85,84%	80,95%	

Hasil *confusion matrix* pada tabel 6 dari 80 data *training* yang diberikan pada algoritma C4.5 menunjukkan bahwa persentase akurasi algoritma C4.5 dengan 65% data *training* adalah sebesar 83,75%. Tingkat akurasi Algoritma juga dapat dilihat pada kurva ROC/AUC berikut ini.



Gambar 2. Kurva ROC/AUC dengan 65% data training

Dari Gambar 2 dapat disimpulkan bahwa algoritma C4.5 dengan 65% data *training* atau sebanyak 80 data yang digunakan memiliki nilai keakurasian AUC sebesar 0,870 dan termasuk dalam *good classification* untuk melakukan klasifikasi calon penerima bantuan lansia.

3. Akurasi Algoritma C4.5 dengan 75% data training

Tabel 4. Tabel *Confusion matrix* algoritma C4.5

accuracy: 81.67% +/- 10.72% (mikro: 81.52%)			
	true Tolak	true Diterima	class precision
pred. Tolak	38	10	79.37%
pred. Diterima	7	37	84.09%
class recall	84.44%	78.72%	

Dari tabel 4 di atas, hasil *confusion matrix* dari 92 data *training* yang diberikan pada algoritma C4.5 menunjukkan bahwa persentase akurasi algoritma C4.5 dengan 75% data *training* adalah sebesar 81,75%. Tingkat akurasi Algoritma juga dapat dilihat pada kurva ROC/AUC berikut ini.



Gambar 3. Kurva ROC/AUC dengan 75% data training

Dari Gambar 3 dapat disimpulkan bahwa algoritma C4.5 dengan 75% data *training* atau sebanyak 92 data yang digunakan memiliki nilai keakurasian AUC sebesar 0,853 dan termasuk dalam *good classification* untuk melakukan klasifikasi calon penerima bantuan lansia.

4. Akurasi Algoritma C4.5 dengan 85% data training

Tabel 5. *Confusion matrix* algoritma C4.5

accuracy: 97.27% +/- 5.82% (mikro: 97.14%)			
	true Tolak	true Diterima	class precision
pred. Tolak	49	0	100.00%
pred. Diterima	3	53	94.64%
class recall	94.23%	100.00%	

Dari tabel 5 di atas, hasil *confusion matrix* dari 105 data *training* yang diberikan pada algoritma C4.5 menunjukkan bahwa persentase akurasi algoritma C4.5 dengan 85% data *training* adalah sebesar 97,14%. Tingkat akurasi Algoritma juga dapat dilihat pada kurva ROC/AUC berikut ini.



Gambar 4. Kurva ROC/AUC dengan 85% data training

Dari Gambar 4 dapat disimpulkan bahwa algoritma C4.5 dengan 85% data *training* atau sebanyak 105 data yang digunakan memiliki nilai keakurasian AUC sebesar 0,913 dan termasuk dalam *exellent classification* untuk melakukan klasifikasi calon penerima bantuan lansia.

5. Akurasi Algoritma C4.5 dengan 95% data training

Tabel 6. *Confusion matrix* algoritma C4.5

accuracy: 91.29% +/- 3.81% (mikro: 91.45%)			
	Isu Tulus	Isu Ditarma	class precision
pred. Tulus	52	4	92.86%
pred. Ditarma	5	55	90.91%
class recall	89.66%	93.22%	

Dari tabel 6 di atas, hasil *confusion matrix* dari 117 data *training* yang diberikan pada algoritma C4.5 menunjukkan bahwa persentase akurasi algoritma C4.5 dengan 95% data *training* adalah sebesar 91,45%. Tingkat akurasi Algoritma juga dapat dilihat pada kurva ROC/AUC berikut ini.



Gambar 5. Kurva ROC/AUC dengan 95% data training

Dari Gambar 5 dapat disimpulkan bahwa algoritma C4.5 dengan 95% data *training* atau sebanyak 117 data yang digunakan memiliki nilai keakurasian AUC sebesar 0,922 dan termasuk dalam *exellent classification* untuk melakukan klasifikasi calon penerima bantuan lansia.

Dari beberapa percobaan yang telah dilakukan, diperoleh tingkat akurasi dari tiap-tiap model seperti yang terlihat pada tabel 7.

Tabel 7. Evaluasi algoritma

Evaluasi Algoritma		Jumlah Data Training	Tingkat Akurasi
Rasio Data Training	55 %	68	74,41%
	65 %	80	83,75%
	75 %	92	81,52%
	85 %	105	97,14%
	95 %	117	91,45%

Tabel 7 menunjukkan akurasi dari algoritma C4.5 untuk klasifikasi penerima program bantuan lansia berdasarkan rasio data *training* dengan interval 55%, 65%, 75%, 85%, dan 95%. Dari tabel tersebut diketahui bahwa algoritma C4.5 memiliki akurasi terbaik pada rasio data *training* sebesar 85% menggunakan 105 data dengan nilai akurasi sebesar 97,14%. Dengan nilai ini maka di pilihlah rasio data *training* 85% sebagai rule yang akan di gunakan untuk implementasi program.

4. KESIMPULAN

Berdasarkan hasil pembuatan sistem penentuan penerima program bantuan pemerintah daerah Kabupaten Kutai Kartanegara menggunakan algoritma C4.5, maka kesimpulan yang diperoleh adalah Algoritma C4.5 dapat diterapkan pada aplikasi penentuan penerima program bantuan pemerintah daerah dengan menggunakan rule yang terbentuk dari hasil akurasi tertinggi pada data training 85% dengan jumlah data 105. Dalam melakukan klasifikasi penentuan penerima program bantuan pemerintah daerah, algoritma C4.5 memiliki nilai akurasi sebesar 97,14% pada data training 85% dengan data 105 dan nilai AUC 0,913 yang termasuk dalam Exellent Classification. Dengan demikian dapat disimpulkan bahwa metode ini akurat dalam melakukan klasifikasi penentuan penerima program bantuan.

5. SARAN

Penelitian ini dapat dikembangkan dengan melakukan komparasi beberapa algoritma untuk mengetahui algoritma yang paling akurat. Karena belum tentu algoritma yang digunakan merupakan algoritma yang paling akurat. Penelitian ini menggunakan satu program bantuan, untuk penelitian selanjutnya bisa dilakukan dengan menambahkan program bantuan lainnya seperti penyandang cacat dan anak terlantar.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Tuhan Yang Maha Esa yang telah memberi rahmat dan hidayahnya sehingga memberikan kemudahan dalam penelitian ini. Terima kasih kepada kedua orang tua, kakak-kakakku tercinta yang selalu memberi dukungan do'a, moril maupun materil kepada penulis. Terima kasih kepada kedua pembimbing, yang telah membimbing penelitian ini sehingga dapat terlaksana dengan baik.

DAFTAR PUSTAKA

- [1] Sekretariat Negara Republik Indonesia. *Undang-Undang Republik Indonesia Nomor 11 Tahun 2009 tentang Kesejahteraan Sosial*. Jakarta: Sekretariat Negara Republik Indonesia.
- [2] Seksi Statistik Sosial Badan Pusat Statistik Kabupaten Kutai Kartanegara. 2015. *Statistik Kesejahteraan Rakyat Kabupaten Kutai Kartanegara*. Tenggarong: Badan Pusat Statistik Kabupaten Kutai Kartanegara.
- [3] Andriani, Anik. 2013. *Sistem Prediksi Penyakit Diabetes Berbasis Decision Tree*. Jakarta Selatan: Jurnal Bianglala Informatika Vol. I No. 1.
- [4] Kusriani, & Emha Taufik Luthfi. 2009. *Algoritma Data Mining*. Yogyakarta: Andi.
- [5] Prasetyo, E. 2012. *Data Mining dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.
- [6] Achmad, Budanis Dwi Meilani, & Slamet, Fauzi. 2012. *Klasifikasi Data Karyawan untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree*. Jurnal IPTEK Vol 16 No.1.
- [7] Gorunescu, Florin. 2011. *Data Mining: Concepts and Techniques*. Verlag berlin Heidelberg: Springer.
- [8] Manlian Ronald, dkk (2012) " Jurnal Identifikasi Penyebab Risiko Kecelakaan Kerja Pada Kegiatan Kontruksi Bangunan Gedung Di DKI Jakarta". Jakarta:Universitas Pelita Harapan.
- [9] Jefri, 2013, Implementasi Algoritma C4.5 Dalam Aplikasi untuk Memprediksi Jumlah Mahasiswa yang Mengulang Mata Kuliah STMIK Yogyakarta, STMIK AMIKOM, Yogyakarta.
- [10] Mengkepe, E. ,2004 Sistem Pendukung Keputusan Pemberian Kredit Mobil PT. Astra International tbk., Universitas Widyatama, Bandung.