

Penerapan Algoritma ID3 Decision Tree Pada Penentuan Penerima Program Bantuan Pemerintah Daerah di Kabupaten Kutai Kartanegara

Dwi Kinasih Widiyati^{*1}, Masna Wati², Herman Santoso Pakpahan³

^{1,2,3}Jurusan Teknologi Informasi dan Komunikasi, Universitas Mulawarman, Samarinda
e-mail: ^{*1}dwikinasih25@gmail.com, ²masna.ssi@gmail.com, ³pakpahanherman891@yahoo.com

Abstrak

Kemiskinan dipandang sebagai ketidakmampuan seseorang dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Seseorang akan dikatakan miskin jika mereka memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan. Upaya Pemerintah Kabupaten Kutai Kartanegara untuk menanggulangi masalah kemiskinan dan kesejahteraan masyarakat yaitu dengan menyediakan berbagai macam program bantuan. Melalui Dinas Sosial program bantuan tersebut dijalankan, setiap program bantuan memiliki kriteria yang berbeda-beda dalam penentuan penerimanya, sehingga membuat proses perekapan menjadi lebih lama. Berdasarkan permasalahan tersebut, maka dibutuhkan suatu sistem yang dapat membantu dan memudahkan aparat daerah dalam proses penentuan calon penerima program bantuan sesuai dengan persyaratan yang ada, dimana sistem tersebut mampu melakukan klasifikasi berdasarkan kriteria program bantuan. Untuk melakukan pengklasifikasian, dalam penelitian yaitu dengan menerapkan Algoritma decision tree dalam penentuan penerima program bantuan daerah dengan menggunakan algoritma ID3 berdasarkan model pohon keputusan untuk menentukan calon penerima bantuan agar bantuan yang disalurkan merata dan tepat pada sasarannya.

Kata kunci— *Decision Tree, Algoritma ID3, Program Bantuan pemerintah Daerah*

1. PENDAHULUAN

Kemiskinan merupakan kejadian dari sebuah fenomena dan fakta yang terjadi di negara Indonesia. Kemiskinan dipandang sebagai ketidak mampuan seseorang dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran. Seseorang akan dikatakan miskin jika mereka memiliki rata-rata pengeluaran perkapita perbulan dibawah garis kemiskinan [1]. Upaya pemerintah kabupaten Kutai Kartanegara untuk menanggulangi masalah kemiskinan dan kesejahteraan masyarakat yaitu dengan adanya program SWTM meliputi anak terlantar, lanjut usia, janda tua, dan penyandang cacat [2]. Selama program bantuan SWTM berjalan banyak kendala yang terjadi seperti batas waktu pengumpulan berkas selalu tidak tepat pada waktu yang sudah ditentukan, banyak persyaratan yang tidak terpenuhi sehingga membuat proses perekapan menjadi lebih lama yang menyebabkan keterlambatan penyetoran data. Berdasarkan permasalahan tersebut, maka dibutuhkan suatu sistem untuk memudahkan aparat daerah dalam proses penentuan calon penerima program bantuan dengan menerapkan metode Algoritma ID3 *Decision Tree*. Penerapan Algoritma ID3 Decision Tree pada penentuan penerima program bantuan berdasarkan variabel yang ada sehingga diharapkan akan memberikan hasil yang dapat mempermudah tugas Dinas Sosial dalam penentuan calon penerima bantuan yang memang layak untuk mendapatkan bantuan.

Objek yang diteliti meliputi program bantuan lanjut usia. Data yang digunakan adalah data penerima program bantuan lanjut usia Kecamatan Kota Bangun tahun 2015. Penerapan algoritma ID3 *decision tree* untuk model klasifikasi dalam menentukan penerima program bantuan daerah lanjut usia agar bantuan yang disalurkan merata dan tepat pada sasarannya.

2. METODE PENELITIAN

2.1 Data Mining

Data mining adalah ekstraksi pola yang menarik dari kumpulan data dalam jumlah besar [3]. Penggalian data memiliki beberapa nama alternatif, meskipun definisi eksaknya berbeda, seperti *knowledge discovery in database* (KDD), analisis pola, arkeologi data, pemanenan informasi, dan intelegensia bisnis. Alasan utama mengapa data mining diperlukan adalah karena adanya sejumlah besar data yang dapat digunakan untuk menghasilkan informasi dan pengetahuan yang berguna. *Knowledge Discovery* dalam sebuah basis data melalui beberapa tahapan untuk mengolah data yang akan digunakan untuk pengetahuan baru, tahapan-tahapan tersebut adalah sebagai berikut:

1. *Data cleaning* juga dikenal sebagai pembersihan data, adalah fase di mana *noise* data dan data yang tidak relevan akan dihapus.
2. *Data integration*, pada tahap ini berbagai sumber data (biasanya data yang heterogen) dikombinasikan dalam sumber yang sama.
3. *Data selection*, pada tahap ini, data yang relevan dengan analisis diputuskan dan diambil dari kumpulan data.
4. *Data transformation*, juga dikenal sebagai konsolidasi data, adalah fase di mana data yang dipilih diubah menjadi bentuk yang sesuai untuk prosedur *mining*.
5. *Data mining* (penggalian data), adalah langkah penting yang menggunakan teknik tertentu yang diterapkan untuk mengekstrak pola berpotensi berguna.
6. *Pattern evaluation*, pada tahap ini, pola yang menarik yang mewakili pengetahuan baru diidentifikasi berdasarkan langkah-langkah yang diberikan.
7. *Knowledge representation*, adalah tahap akhir di mana pengetahuan ditemukan. Langkah menggunakan teknik visualisasi untuk membantu pengguna memahami dan menginterpretasikan hasil *data mining*.

2.2 Decision Tree

Pohon keputusan (*Decision Tree*) adalah pohon yang ada dalam analisis pemecahan masalah, pemetaan mengenai alternatif-alternatif pemecahan masalah yang dapat diambil dari masalah. Pohon Keputusan dapat juga dikatakan salah satu metode klasifikasi yang paling populer karena mudah untuk diinterpretasi oleh manusia. Konsep dasar *decision tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan/*rule* [4]. Pada *decision tree* terdapat 3 jenis *node*, yaitu:

1. *Root Node*, merupakan *node* paling atas, pada *node* ini tidak ada *input* dan bisa tidak mempunyai *output* atau mempunyai *output* lebih dari satu.
2. *Internal Node*, merupakan *node* percabangan, pada *node* ini hanya terdapat satu *input* dan mempunyai *output* minimal dua.
3. *Leaf Node* atau *Terminal Node*, merupakan *node* akhir, pada *node* ini hanya terdapat satu *input* dan tidak mempunyai *output*.

2.3 Algoritma ID3

Penelitian ini menggunakan algoritma ID3 (*Iterative Dichotomizer 3*). Algoritma ID3 adalah algoritma pembelajaran pohon keputusan yang sederhana dikembangkan oleh J. Ross Quinlan pada tahun 1993. Algoritma ID3 melakukan prosedur pencarian

secara menyeluruh pada semua kemungkinan pohon keputusan [5]. Secara ringkas, langkah kerja algoritma ID3 adalah sebagai berikut:

1. Input data training, dan variabel
2. Hitung *Entropy* dan *Information Gain* dari setiap variabel dengan persamaan (1).

$$Entropy(S) = -P_+ \log_2 P_+ - P_- \log_2 P_- \tag{1}$$

S : ruang (data) sample yang digunakan untuk training.

P_+ : jumlah yang bersolusi positif (mendukung) pada data sampel untuk kriteria tertentu.

P_- : jumlah yang bersolusi negatif (tidak mendukung) pada data sampel untuk kriteria tertentu.

$$Gain(S, A) = Entropy(S) - \sum_{ve\ Value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{2}$$

S : ruang (data) sampel yang digunakan untuk training.

A : salah satu variabel dalam S

V : Menyatakan suatu nilai yang mungkin untuk variabel A

Value(A) : himpunan nilai-nilai yang mungkin untuk variabel A.

|S_v| : jumlah sampel untuk nilai v.

|S| : jumlah seluruh sampel data.

Entropy(S_v) : entropy untuk sampel-sampel yang memiliki nilai v.

Tujuan dari pengukuran nilai *information gain* adalah untuk memilih variabel yang akan dijadikan cabang pada pembentukan pohon keputusan.

3. Pilih variabel yang memiliki nilai *information gain* terbesar.
4. Bentuk simpul yang berisi variabel tersebut.
5. Ulangi proses perhitungan *information gain* yang akan terus dilaksanakan sampai semua data telah termasuk dalam kelas yang sama. Variabel yang telah dipilih tidak diikuti lagi dalam perhitungan nilai *Information gain*[6].

2.4 Confusion matrix

Hasil evaluasi algoritma dapat ditampilkan dengan menggunakan *Confusion Matrix*. *Confusion matrix* adalah tool yang digunakan untuk evaluasi model klasifikasi untuk memperkirakan objek yang benar atau salah. Sebuah matrix dari prediksi yang akan dibandingkan dengan kelas yang asli dari inputan atau dengan kata lain berisi informasi nilai actual dan prediksi pada klasifikasi.

Tabel 1. Confusion Matrix 2 Kelas

		Predicted Class	
		Negative	Positive
Actual Class	Negative	a	b
	Positive	c	d

Dengan,

- a. a, adalah jumlah prediksi yang benar bahwa *instance* bersifat negatif
- b. b, adalah jumlah prediksi yang salah bahwa *instance* bersifat negatif
- c. c, adalah jumlah prediksi yang salah bahwa *instance* bersifat positif
- d. d, adalah jumlah prediksi yang benar bahwa *instance* bersifat positif

Beberapa persyaratan standar yang telah ditetapkan untuk matriks klasifikasi dua kelas:

- (1) Akurasi (AC), adalah proporsi jumlah prediksi yang benar. Hal ini ditentukan dengan menggunakan persamaan (3).

$$Akurasi(AC) = \frac{a+d}{a+b+c+d} \tag{3}$$

ROC (*Receiver Operating Characteristic*) Curve adalah grafik antara sensitifitas (*true positive rate*) pada sumbu Y dengan 1- spesifisitas pada sumbu X

(*false positive rate*), seakan - akan menggambarkan tawar menawar antara sensitivitas dan spesifisitas, yang tujuannya adalah untuk menentukan *cut off point* pada uji diagnostic yang bersifat kontinyu. Untuk klasifikasi data mining , nilai AUC dapat dibagi menjadi beberapa kelompok [7].

- a. 0.90 - 1.00 = *Exellent classification*
- b. 0.80 - 0.90 = *Good Classification*
- c. 0.70 - 0.80 = *Fair Classification*
- d. 0.60 - 0.70 = *Poor Classification*
- e. 0.50 - 0.60 = *Failure Classification*

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penelitian

Penelitian ini mengklasifikasikan penerima program bantuan pemeintah daerah berdasarkan variabel yang telah ditentukan. Dalam penelitian ini data diperoleh dari Dinas Sosial Kabupaten Kutai Kartanegara. Data yang digunakan dalam penelitian ini berupa data yang berkaitan dengan keterangan diri calon penerima bantuan. Data yang digunakan berjumlah 123 yang diambil dari data calon penerima bantuan tahun 2014.

Data yang ada kemudian akan digunakan untuk membentuk 5 kelas data *training* dan data *testing* 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.

3.1. Pembentukan Model

Pembentukan aturan pada algoritma ID3 didapatkan dengan membentuk pohon keputusan berdasarkan nilai *gain* tertinggi dari setiap atribut variabel. Untuk membentuk model pohon keputusan maka akan dilakukan proses perhitungan dengan algoritma ID3. Berikut contoh perhitungan nilai *entropy* dan *gain* menggunakan data *training* 55 % dengan jumlah data 68.

Diketahui jumlah data 68 , 36 data bestatus= "Diterima" dan 32 data berstatus "Tolak". Kemudian hitung nilai entropy keseluruhan menggunakan persamaan (1):

$$\begin{aligned} \text{Entropy (S)} &= -\left(\frac{36}{68} \times \log_2 \left(\frac{36}{68}\right)\right) - \left(\frac{32}{68} \times \log_2 \left(\frac{32}{68}\right)\right) \\ &= 0,997502546 \end{aligned}$$

kemudian hitung nilai entropy pada masing-masing variabel menggunakan persamaan (1). Contoh perhitungan *entropy* untuk variabel umur, sebagai berikut:

Values (Umur) = < 60, > 60

$S_{<= 60} = [0+, 7-]$, $S_{< 60} = 7$

$S_{> 60} = [36+, 25-]$, $S_{> 60} = 61$

$$\text{Entropy (S <= 60)} = \left(-\frac{0}{7} \times \log_2 \left(\frac{0}{7}\right)\right) - \left(\frac{7}{61} \times \log_2 \left(\frac{7}{61}\right)\right) = 0$$

$$\text{Entropy (S > 60)} = \left(-\frac{36}{61} \times \log_2 \left(\frac{36}{61}\right)\right) - \left(\frac{25}{61} \times \log_2 \left(\frac{25}{61}\right)\right) = 0,976414308$$

Berdasarkan nilai entropy yang telah diketahui kemudian menghitung gain untuk mendapatkan *node* awal (akar) dengan menggunakan persamaan (2).

Contoh perhitungan *gain* untuk variabel umur, sebagai berikut:

$$\begin{aligned} \text{Gain (S, Umur)} &= 0,997502546 - 0 - \left(\frac{61}{68} \times 0,976414308 \right) \\ &= 0,997502546 - 0 - 0,875901070 = 0,1216014756 \end{aligned}$$

Kemudian untuk mencari nilai *entropy* dan *gain* dari variabel jenis kelamin, status perkawinan, pendidikan, pekerjaan, dan keterampilan yaitu dengan cara yang sama. Diketahui dari perhitungan manual nilai *gain* tertinggi yang menjadi akar dari pohon keputusan adalah variabel pekerjaan. Nilai *entropy* dan *gain* dengan 55% data *training* dapat dilihat pada tabel 2.

Tabel 2. Nilai *entropy* dan *gain* menggunakan data *training* 55 %

NO	Variabel	Values	Jumlah kasus	diterima (+)	Entropi ditolak (-)	Hasil	gain
1	Umur	<= 60	7	0	7	0,997502546	0,121601476
		> 60	61	36	25	0,976414308	
	Jenis Kelamin	Laki-Laki	28	6	22	0,749595257	0,211623215
		Perempuan	40	30	10	0,811278124	
	Status Perkawinan	Belum Kawin	5	3	2	0,970950594	0,074783318
		Kawin	42	18	24	0,985228136	
		Cerai Mati	18	14	4	0,764204507	
		Cerai Hidup	3	1	2	0,918295834	
	Pendidikan	Belum/Tidak Tamat SD	26	15	11	0,982858609	0,007933203
		SD/MI/Sederajat	39	19	20	0,999525689	
		SLTP/MTs/Sederajat	3	2	1	0,918295834	
	Pekerjaan	Buruh pertanian tidak tetap	11	3	8	0,845350937	0,505520695
		Buruh tidak tetap non pertanian	6	0	6	0	
		Tidak bekerja	39	33	6	0,619382195	
		Usaha sendiri	11	0	11	0	
		Usaha dengan buruh tetap/tidak tetap	1	0	1	0	
	Keterampilan	Bidang pertanian/perikanan	32	12	20	0,954434003	0,114569884

				0,8497511
Lainnya	29	21	8	37
Pertukangan	1	0	1	0
Menjahit	1	1	0	0
				0,9709505
Kerajinan tangan	5	2	3	94

3.2 Evaluasi dan validasi hasil

Setelah dilakukan beberapa percobaan, maka akan dilakukan pengujian tingkat akurasi untuk melihat kinerja dari metode Algoritma ID3. Kemudian dipilih salah satu model dengan akurasi tertinggi untuk diimplementasikan ke dalam aplikasi. Evaluasi tingkat akurasi ditampilkan dengan *confusion matrix* dan kurva ROC (*Receiver Operating Characteristics*)/AUC (*Area Under Cover*) menggunakan RapidMiner. Teknik validasi yang digunakan yaitu *cross validation* pada *RapidMiner*. Berikut adalah pengujian tingkat akurasi algoritma ID3 dengan *RapidMiner*.

1. Akurasi Algoritma ID3 dengan 55% data

Tabel 3. Hasil *Accuracy Confusion Matrix Algoritma ID3*
Accuracy:94.29% +/- 9.48% (mikro: 94.12%)

	True ditolak	True diterima	<i>Class precesion</i>
Pred. Tolak	30	1	96.77%
Pred. diterima	3	34	91.89%
<i>Class recall</i>	90.91%	97.14%	

Dari tabel 1 dari 68 data *training* hasil akurasi *confusion matrix Algoritma ID3* menggunakan RapidMiner dengan 55% data *training* sebesar 94,29 %. Tingkat akurasi Algoritma ID3 juga dapat dilihat pada Gambar 1. kurva ROC/AUC.



Gambar 1. Kurva ROC/AUC dengan 55 % Data Training

Hasil yang didapat dari pengolahan AUC dengan menggunakan 55 % data *training* sebesar 0,567 dapat dilihat pada gambar 2.

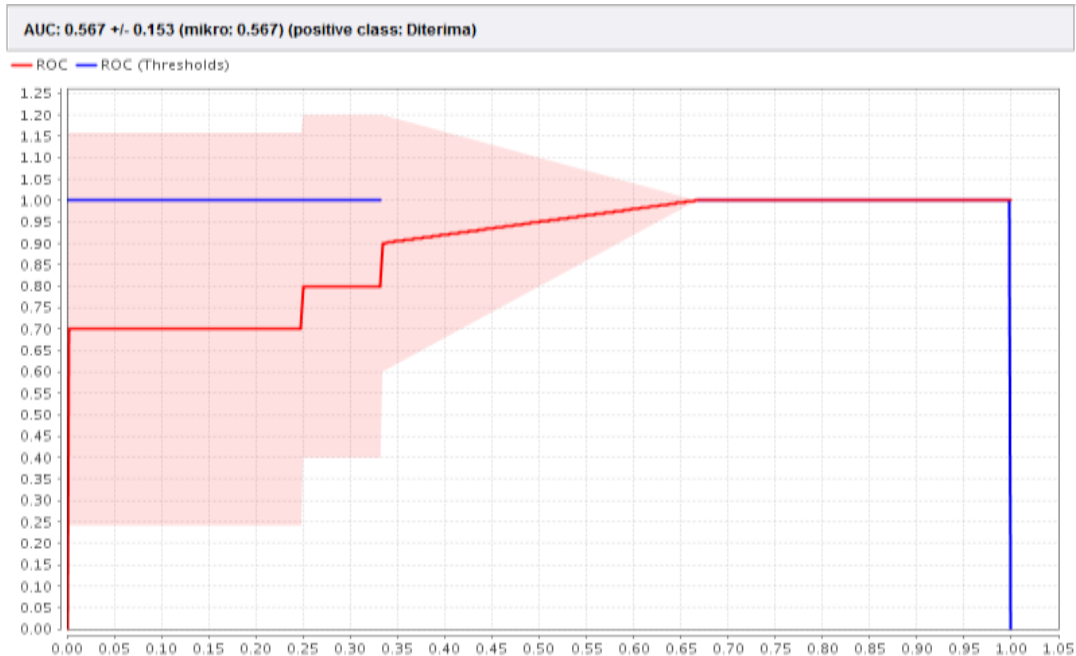
2. Akurasi Algoritma ID3 dengan 65% data

Tabel 2. Hasil *Accuracy Confusion Matrix Algoritma ID3*
Accuracy:96.29% +/- 5.73% (mikro: 96.25%)

	True ditolak	True diterima	<i>Class precesion</i>
Pred. Tolak	36	1	97.30%

Pred. diterima	2	41	95.35%
<i>Class recall</i>	94.74%	97.62%	

Dari tabel 2 dari 80 data *training* hasil akurasi *confusion matrix* Algoritma ID3 menggunakan RapidMiner dengan 65% data *training* sebesar 96,25 %. Tingkat akurasi Algoritma ID3 juga dapat dilihat pada kurva ROC/AUC Gambar 2.



Gambar 2. Kurva ROC/AUC dengan 65 % Data Training

Hasil yang didapat dari pengolahan AUC dengan menggunakan 65 % data *training* sebesar 0,567 dapat dilihat pada gambar 3.

3. Akurasi Algoritma ID3 dengan 75% data

Tabel 3. Hasil Accuracy Confusion Matrix Algoritma ID3
Accuracy:97.89% +/- 4.23% (mikro: 97.83%)

	True ditolak	True diterima	<i>Class precision</i>
Pred. Tolak	44	1	97.78%
Pred. diterima	1	46	97.87%
<i>Class recall</i>	97.78%	97.87%	

Dari tabel 3 dari 92 data *training* hasil akurasi *confusion matrix* Algoritma ID3 menggunakan RapidMiner dengan 75% data *training* sebesar 97,89 %. Tingkat akurasi Algoritma ID3 juga dapat dilihat pada kurva ROC/AUC berikut ini:



Gambar 3. Kurva ROC/AUC dengan 75 % Data Training

Hasil yang didapat dari pengolahan AUC dengan menggunakan 75 % data *training* sebesar 0,575 dapat dilihat pada gambar 4.

4. Akurasi Algoritma ID3 dengan 85% data

Tabel 4. Hasil *Accuracy Confusion Matrix Algoritma ID3*

Accuracy:97.27% +/- 4.17% (mikro: 97.14%)

	True ditolak	True diterima	<i>Class precesion</i>
Pred. Tolak	50	1	98.04%
Pred. diterima	2	52	96.30%
<i>Class recall</i>	96.15%	98.11%	

Dari tabel 4 dari 105 data *training* hasil akurasi *confusion matrix Algoritma ID3* menggunakan RapidMiner dengan 85% data *training* sebesar 97,27 %. Tingkat akurasi Algoritma ID3 juga dapat dilihat pada kurva ROC/AUC berikut ini:



Gambar 4. Kurva ROC/AUC dengan 85 % Data Training

Hasil yang didapat dari pengolahan AUC dengan menggunakan 85 % data *training* sebesar 0,550 dapat dilihat pada gambar 4.5

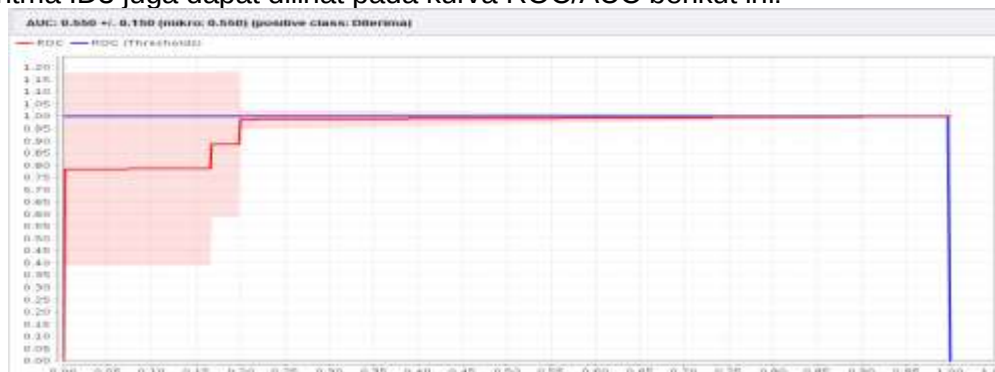
5. Akurasi Algoritma ID3 dengan 95% data

Tabel 5. Hasil *Accuracy Confusion Matrix Algoritma ID3*

Accuracy:96.52% +/- 4.27% (mikro: 96.58%)

	True ditolak	True diterima	<i>Class precesion</i>
Pred. Tolak	56	2	96.55%
Pred. diterima	2	52	96.61%
<i>Class recall</i>	96.55%	96.61%	

Dari tabel 5 dari 117 data *training* hasil akurasi *confusion matrix Algoritma ID3* menggunakan RapidMiner dengan 95% data *training* sebesar 97,27 %. Tingkat akurasi Algoritma ID3 juga dapat dilihat pada kurva ROC/AUC berikut ini:



Gambar 5. Kurva ROC/AUC dengan 95 % Data Training

Hasil yang didapat dari pengolahan AUC dengan menggunakan 95 % data *training* sebesar 0,550 dapat dilihat pada gambar 6. Dari beberapa percobaan yang telah dilakukan, diperoleh tingkat akurasi dari tiap-tiap model seperti yang terlihat pada tabel dibawah berikut ini:

Tabel 6. Evaluasi Algoritma

Evaluasi Algoritma	Jumlah Data Training	Tingkat Akurasi
Rasio Data Training	55 %	94,89 %
	65 %	96,25 %
	75 %	97,89 %
	85 %	97,27 %
	95 %	96,52 %

Tabel diatas menunjukkan akurasi dari algoritma ID3 untuk klasifikasi penerima program Bantuan lanjut usia berdasarkan rasio data *training* dengan interval 55%, 65%, 75%, 85% dan 95%. Dari tabel tersebut diketahui bahwa algoritma id3 memiliki akurasi terbaik pada rasio data *training* sebesar 75% menggunakan 92 data dengan nilai akurasi sebesar 97,89%. Aplikasi yang dibangun berdasarkan pada rule yang telah diperoleh tools *RapidMiner* dengan akurasi tertinggi yaitu pada data *training* 75 % dengan data sebesar 92 dan nilai akurasi sebesar 97,89 %.

4. KESIMPULAN

Berdasarkan pembahasan terhadap hasil penentuan penerima program bantuan daerah menggunakan metode Algoritma ID3, maka dapat disimpulkan sebagai berikut:

- 1 Penentuan penerima program bantuan lanjut usia pada Dinas Sosial Kabupaten Kukar Kartanegara menggunakan varabel umur jenis kelamin, status perkainan, pendidikan terakhir, pekerjaan dan keterampilan.
- 2 Dalam melakukan Klasifikasi metode ID3 menggunakan rapidminer menghasilkan akurasi sebesar 97,89 % dengan 92 data sehingga algoritma ini sangat baik digunakan dalam klasifikasi penentuan penerima program bantuan lanjut usia.
- 3 Aplikasi ini diharapkan dapat membantu dalam memaksimalkan kinerja pihak Dinas Sosial dalam proses seleksi khususnya dalam hal menentukan calon penerima program bantuan yang akan diterima secara lebih obyektif dan sesuai dengan yang diharapkan.

5. SARAN

Penelitian yang dilakukan ini menggunakan metode Algoritma ID3, untuk penelitian selanjutnya bisa menggunakan metode algoritma yang lain. Penelitian selanjutnya dapat mengembangkan aplikasi ini untuk lebih baik lagi sesuai dengan kebutuhan seperti menggunakan lebih dari satu prograam bantuan.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik. 2016. *Kemiskinan dan Ketimpangan*. BPS-Statistics Indonesia.
- [2] Badan Pusat Statistik. 2016. *Kutai Kartanegara Dalam Angka 2016*. Tenggara: Badan Pusat Statistik Kabupaten Tenggara Kutai Kartanegara.
- [3] Prasetyo, E. (2012). *Data Mining dan Aplikasi Menggunakan MATLAB*. Yogyakarta: Andi.

- [4] Yusuf Elmande1, Prabowo Pudjo Widodo. 2012. "Pemilihan Criteria Splitting dalam Algoritma Iterative Dichotomiser 3 (ID3) untuk Penentuan Kualitas Beras : Studi Kasus pada Perum Bulog Divre Lampung." *TELEMATIKA MKOM VOL.4 No.1*, ISSN: 2085-725X.
 - [5] Ture, M., Fusun, T., Imran, K., 2009. Using Kaplan–Meier analysis together with decision tree methods (C&RT, CHAID, QUEST, C4.5 and ID3) in determining recurrence-free survival of breast cancer patients, *Expert Systems with Applications* 36, 2017–2026.
 - [6] Defiyanti, S., Pardede, C.D.L., 2009. *Perbandingan Kinerja Algoritma ID3 dan C4.5 Dalam Klasifikasi Spam-Mail, Skripsi*, Universitas Gunadarma, Jakarta.
 - [7] Gorunescu, F. (2011). *Data mining (Intelligent System Refrence Library Volume 12)*. Craiova:Springe.
 - [8] Wati, Masna, and Abdul Hadi. "Implementasi Algoritma Naive Bayesian Dalam Penentuan Penerima Program Bantuan Pemerintah." *STMIK KHARISMA Makassar 3.1* (2017): 22-26.
 - [9] Sukerti, Ni Kadek . 2015. "Penerapan Fuzzy Topsis Untuk Seleksi Penerima Bantuan Kemiskinan." *Jurnal Informatika*, Vol.15, No.2.
 - [10] Wahyuni, Sri. 2011. "Sistem Pendukung Keputusan Penerima Raskin Menggunakan Metode Topsis (Technique For Order Preference Similarity To Ideal Solution) (Studi Kasus : Kelurahan Simpang Baru)." Tugas Akhir.
-