

SISTEM PENENTUAN TINGKAT KESEJAHTERAAN ANAK MENGUNAKAN ALGORITMA C 4.5

Yuli Murdianingsih

Jurusan Manajemen Informatika STMIK Subang, Jawa Barat
Jl. Marsinu No 5 Subang, 40112 telp (0260)417853
Email:yuli_murdianingsih@yahoo.com

Abstrak

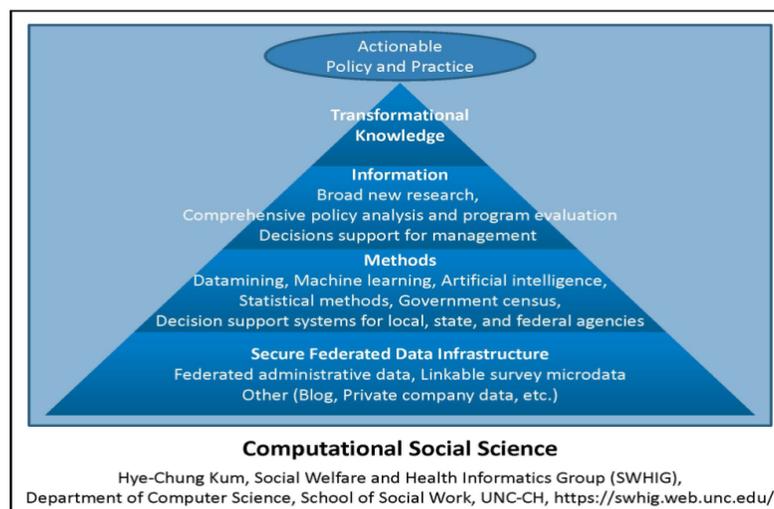
Implementasi algoritma C 4.5 dalam menentukan tingkat kesejahteraan anak merupakan sebuah upaya preventif terhadap masa depan bangsa yang tergantung pada masa depan anak-anak. Penelitian ini bertujuan untuk memperoleh model sistem yang memiliki nilai akurasi tinggi lebih dari 80%. Sehingga diharapkan pengambilan keputusan penentuan tingkat kesejahteraan anak memiliki kepastian kebesaran yang memadai. Sistem yang diharapkan berupa basis data relasional yang menggambarkan aktivitas penentuan tingkat kesejahteraan anak. Data yang dikumpulkan dalam penelitian ini berupa data primer yang diperoleh dari hasil wawancara langsung dengan responden menggunakan kuesioner. Data responden dalam penelitian ini adalah sampel data sebanyak 212 data, dimana pengambilan data dilakukan secara acak (random) dari populasi 14291 KK. Dari 212 data tersebut 176 data digunakan sebagai data training dan 36 data lainnya. Dilakukan penghitungan entropy dan gain untuk memperoleh pohon keputusan. Diperoleh model sistem yang memiliki tingkat akurasi tinggi yaitu sebesar 95,65%. Diperoleh basis data relasional untuk merealisasikan sistem penentuan keputusan penentuankesejahteraan anak.

Key word: C 4.5, tingkat kesejahteraan anak, akurasi

1. PENDAHULUAN

Upaya mengentaskan kemiskinan seolah-olah merupakan usaha yang tidak ada ujungnya, sehingga usaha untuk mewujudkan kesejahteraan anakpun mengalami kendala baik teknis maupun struktural. Kendala ini terutama dialami oleh sebagian wilayah yang letaknya jauh dari pusat. Kota Bandung khususnya kecamatan cicadas merupakan daerah dengan kondisi terkumuh di kota Bandung, dimana kekumuhan merupakan salah satu indikator kemiskinan. Dengan menggunakan parameter tingkat pemenuhan kebutuhan anak berupa fisik, emosional, intelektual dan sosial spiritual diharapkan dapat mengungkap deskripsi tingkat pencapaian kesejahteraan anak.

Masalah sosial sudah menjadi bagian tidak terpisahkan dari informatika bahkan sudah di definisikan menjadi cabang ilmu yaitu ilmu komputasi social [11]. Di Universitas Carolina Utara telah didefinisikan ilmu komputasi social dengan hirarki seperti diperlihatkan pada Gambar 1.2, dimana menyertakan data mining sebagai salah satu metode dalam ilmu komputasi social sebagai bagian dalam pengambilan keputusan. Adapun arsitektur KDD (Knowledge Discovery and Data Mining) untuk system kesejahteraan social diperlihatkan dalam Gambar 1.



Gambar 1 Hirarki ilmu komputasi sosial [11]

Dengan demikian sangat memungkinkan data pemenuhan kesejahteraan di kota Bandung khususnya di kecamatan Cicadas, dikelola menggunakan teknik data mining, sebagai rintisan untuk Indonesia. Data sosial dengan kecenderungan yang terus membesar memerlukan pemanfaatan teknik data mining. Data sosial yang belum dikelola secara optimal berpengaruh pada pelayanan atau program baik untuk penanganan kemiskinan ataupun pelayanan bagi anak yang dilakukan secara sporadis, diskontinu dan responsif atau bahkan terhambat.

Belum ada penelitian terkait dengan domain masalah kesejahteraan anak, untuk klasifikasi terpenuhi dan tidak terpenuhi, menggunakan parameter pemenuhan kebutuhan intelektual, emosional, social spiritual dan fisik yang melakukan pengukuran akurasi. Penelitian[4] hanya menghitung akurasi aturan asosiasi, menghitung bobot parameter dan kemiripan model system kesejahteraan, penelitian[5] menentukan nilai factor care dari masing-masing parameter kesejahteraan anak.

Dari penelitian yang pernah dilakukan [13] algoritma C 4.5 memiliki kemampuan dalam melakukan klasifikasi dengan nilai akurasi hampir 80 % dan nilai ini dianggap memiliki kehandalan. Sementara penelitian dibidang social terutama terhadap pemenuhan kebutuhan dasar anak yang menggunakan algoritma C4.5 belum ditemukan.

Tujuan dari penelitian ini adalah membuat model system untuk menentukan tingkat pemenuhan kebutuhan dasar anak dengan menggunakan algoritma C4.5 yang memiliki nilai akurasi >80%. Dengan harapan pemanfaatan data mining untuk pengambilan keputusan pemenuhan tingkat kebutuhan anak dapat mendukung instansi terkait atau pengambil kebijakan, memberikan pelayanan yang diperlukan secara sistematis, berkesinambungan dan terintegrasi

Penerapan algoritma C4.5 pada data pemenuhan kebutuhan anak dapat memperkaya khazanah penggunaan metode bidang informatika dalam bidang ilmu sosial. Dengan menggunakan algoritma C 4.5 dalam mengelola data pemenuhan kebutuhan anak diharapkan dapat memberikan nilai tambah pada pengelolaan data karena dapat diekstraksi menjadi pengetahuan dalam pengambilan keputusan pemenuhan kebutuhan anak. Algoritma C 4.5 diharapkan dapat mempercepat proses pengambilan keputusan, tingkat pemenuhan kebutuhan anak berdasarkan parameter pemenuhan kebutuhan fisik, intelektual, emosional dan sosial spiritual.

2. TINJAUAN PUSTAKA

Craw [4] menyampaikan pohon keputusan menggunakan algoritma C 4.5 dalam menentukan keputusan bermain tenis berdasarkan pada outlook, temperature, humadity dan windy. Untuk memperoleh root pada pohon keputusan dan seluruh node yang lain yang tertera pada pohon keputusan, Craw melakukan penghitungan Entropy, seperti pada Rumus 1, melakukan penghitungan Gain informasi seperti pada Rumus 2.

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

Dimana:
 S : himpunan kasus
 n : jumlah partisi S
 pi : proporsi dari Si terhadap S

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

Dimana:
 S : himpunan kasus
 A : atribut
 n : jumlah partisi atribut A
 |Si| : jumlah kasus pada partisi ke-i
 |S| : jumlah kasus dalam S

Seleksi atribut melalui teori entropi dilakukan dengan menentukan gain informasi (*Information gain*), gain informasi yang tertinggi menjadi acuan dalam menentukan sebuah atribut n sebagai root atau node-node lainnya dalam sebuah pohon keputusan. Dilakukan proses klasifikasi yang meliputi dua tahap, yaitu tahap pertama adalah proses *learning* yaitu data training dianalisa oleh algoritma klasifikasi dan menghasilkan sejumlah aturan, yaitu proses menganalisa data training oleh algoritma klasifikasi, kemudian tahap kedua yaitu *training* adalah menguji akurasi dari algoritma klasifikasi dengan memasukkan sejumlah data [8].

Berikutnya pengujian dilakukan menggunakan *Confusion matrix* [14], merupakan metode untuk mengevaluasi model klasifikasi pada data mining dengan menghasilkan nilai prediksi benar dan prediksi salah jika dibandingkan ke nilai tujuan (*target value*) dalam data, seperti diperlihatkan dalam Tabel 1.

Tabel 1. Confusion Matric

Confusion Matrix		Target			
		Positive	Negative		
Model	Positive	a	b	Positive Predictive Value	$a/(a+b)$
	Negative	c	d	Negative Predictive Value	$d/(c+d)$
		Sensitivity	Specificity	Accuracy = $(a+d)/(a+b+c+d)$	
		$a/(a+c)$	$d/(b+d)$		

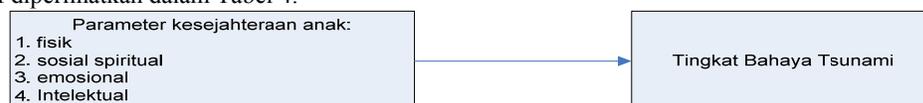
Berdasarkan pada Berry [1] dan Lee [12] tingkat pemenuhan kebutuhan anak meliputi parameter kebutuhan fisik, intelektual, emosional dan sosial spiritual sedangkan indikator pemenuhan kesejahteraan anak mengacu pada indikator yang telah ditentukan oleh BKKBN, seperti diperlihatkan dalam Tabel 2.

Tabel 2. Indikator pemenuhan kebutuhan anak

Kebutuhan Anak		Parameter
FISIK	Sandang	Anggota keluarga memiliki pakaian yang berbeda untuk di rumah, sekolah/bekerja dan untuk bepergian
		Anak mendapat 1 stel baju baru setiap tahun
	Pangan	Pada umumnya makan 2x sehari atau lebih
		Alokasi untuk makan tidak kurang dari
	Papan	Tipe rumah yang ditempati
Kesehatan	Status rumah yang ditempati	
	Bila ada yang sakit dibawa ke sarana kesehatan 3 bulan terakhir tidak sakit/sehat	
INTELEKTUAL		Anak usia 7-15 tahun bersekolah
EMOSIONAL	Merasa aman	Membandingkan anak dengan saudara atau orang lain
		Mengkritik atau mencari kesalahan
	Kasih sayang dan penerimaan	Meluangkan waktu untuk bersama anak
		Menerima/memberikan peluang pada anak untuk mengemukakan pendapat
SOSIAL SPIRITUAL		Menyediakan/membelikan bahan mainan untuk anak
		Memberi peluang pada anak untuk bermain bersama teman sebaya
		Mengarahkan anak untuk mengikuti kegiatan keagamaan di sekitar rumah

3. METODE PENELITIAN

Jenis data yang dikumpulkan dalam penelitian ini berupa data primer yang diperoleh dari hasil wawancara langsung dengan responden menggunakan kuesioner. Data responden dalam penelitian ini adalah sampel data sebanyak 212 data, dimana pengambilan data dilakukan secara acak (*random*) dari populasi 14291 KK. Dari 212 data tersebut 176 data digunakan sebagai data training dan 36 data lainnya digunakan sebagai data testing. Potongan data diperlihatkan pada Tabel 3, hasil konversi data kebesaran kualitatif diperlihatkan dalam Tabel 4.



Gambar 2. Empat parameter penentuan tingkat kesejahteraan anak

Tabel 2 Kategori atribut

Atribut	Nilai Angka	Kategorikal
Fisik	17 – 21	Memadai
	12 – 16	Cukup
	≤ 11	Tidak memadai
Intelektual	7 – 9	Memadai
	5 - 6	Cukup
	≤ 4	Tidak memadai
Emosional	9 – 10	Memadai
	6 – 8	Cukup
	≤ 5	Tidak memadai
Sosial Spiritual	7 – 9	Memadai
	5 - 6	Cukup
	≤ 4	Tidak memadai

Tabel 3 Data responden bulan November 2012

NO	NAMA	FISIK	INTELEKTUAL	EMOSIONAL	SOSIAL SPIRITUAL	PENDAPATAN	CLASS
1	TINI	16	7	8	7	1.5 -2JT	TERPENUHI
2	ENGKOS	18	6	9	8	1.5 - 2	TERPENUHI
3	IDA	16		8	5	2JT	
4	NANDANG	13	7	6	5	1JT-1.450	TIDAK TERPENUHI
5	AGUS	21	sdh kerja	9		>2JT	
6	ETI	14	5	7	6	1JT-1.450	TIDAK TERPENUHI
7	RAMDANI	21	sdh kerja	8	9	>2 JT	
8	RUDIYANTO	10	5		7	1.5 - 2	
9	SARNO	19	7	5	6	1.5-2	TERPENUHI
10	AYI SULAEMAN	12	5	6	6	1-4.50	TIDAK TERPENUHI
11	ENDANG	20	6	9	9	>2	TERPENUHI
12	IWAN	13	7	6	5	1-1.450	TIDAK TERPENUHI
13	IRMA	18	Balita	7	9	1-1.450	
14	GUSTINI	21	8	9	8	>2	TERPENUHI

Tabel 4 Data training bulan November 2012

NO	NAMA	FISIK	INTELEKTUAL	EMOSIONAL	SOSIAL SPIRITUAL	CLASS
1	TINI	CUKUP	MEMADAI	CUKUP	MEMADAI	TERPENUHI
2	ENGKOS	MEMADAI	CUKUP	MEMADAI	MEMADAI	TERPENUHI
3	NANDANG	CUKUP	MEMADAI	CUKUP	CUKUP	TIDAK TERPENUHI
4	ETI	CUKUP	CUKUP	CUKUP	CUKUP	TIDAK TERPENUHI
5	SARNO	MEMADAI	MEMADAI	TIDAK MEMADAI	CUKUP	TERPENUHI
6	AYI SULAEMAN	CUKUP	CUKUP	CUKUP	CUKUP	TIDAK TERPENUHI
7	ENDANG	MEMADAI	CUKUP	MEMADAI	MEMADAI	TERPENUHI
8	IWAN	CUKUP	MEMADAI	CUKUP	CUKUP	TIDAK TERPENUHI
9	GUSTINI	MEMADAI	MEMADAI	MEMADAI	MEMADAI	TERPENUHI

10	MILA	MEMADAI	MEMADAI	CUKUP	CUKUP	TERPENUHI
11	RATNA	MEMADAI	TIDAK MEMADAI	CUKUP	CUKUP	TIDAK TERPENUHI
12	USEP	MEMADAI	CUKUP	MEMADAI	MEMADAI	TERPENUHI
13	SUSI	MEMADAI	MEMADAI	MEMADAI	MEMADAI	TERPENUHI
14	MUFIDAH	MEMADAI	CUKUP	MEMADAI	CUKUP	TERPENUHI

Data hasil konversi dalam Tabel 4 (potongan data) digunakan untuk menghitung entropy total menggunakan Rumus 1, diperlihatkan dalam paparan sebagai berikut:

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 &= -p_1 * \log_2 p_1 - p_2 * \log_2 p_2 \\
 &= -\left(\frac{S_1}{S}\right) * \log_2 \left(\frac{S_1}{S}\right) - \left(\frac{S_2}{S}\right) * \log_2 \left(\frac{S_2}{S}\right) \\
 &= -\frac{106}{150} * \log_2 \left(\frac{106}{150}\right) - \frac{44}{150} * \log_2 \left(\frac{44}{150}\right) \\
 &= 0.353968 + 0.51902 \\
 &= 0.872988
 \end{aligned}$$

Menghitung **Entropy untuk atribut fisik** ($S_{memadai}$, S_{cukup} , $S_{tidak\ memadai}$):

$$\begin{aligned}
 Entropy(S_{memadai}) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 &= -\frac{96}{103} * \log_2 \left(\frac{96}{103}\right) - \frac{7}{103} * \log_2 \left(\frac{7}{103}\right) \\
 &= 0.094637 + 0.263631 \\
 &= 0.358269
 \end{aligned}$$

$$\begin{aligned}
 Entropy(S_{cukup}) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 &= -\frac{10}{15} * \log_2 \left(\frac{10}{15}\right) - \frac{5}{15} * \log_2 \left(\frac{5}{15}\right) \\
 &= 0.389975 + 0.528321 \\
 &= 0.918296
 \end{aligned}$$

$$\begin{aligned}
 Entropy(S_{tidakmemadai}) &= \sum_{i=1}^n -p_i * \log_2 p_i \\
 &= -\frac{0}{32} * \log_2 \left(\frac{0}{32}\right) - \frac{32}{32} * \log_2 \left(\frac{32}{32}\right) \\
 &= 0 + 0 \\
 &= 0
 \end{aligned}$$

Menghitung **Information Gain untuk nilai fisik**:

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

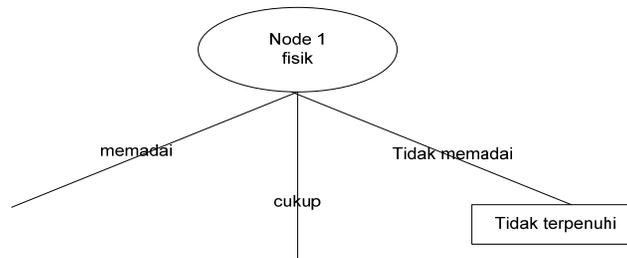
$$\begin{aligned}
 &= Entropy(total) - \sum_{i=1}^3 \frac{S_{fisik}}{S_{total}} * Entropy(fisik) \\
 &= 0.872988 - \left(\frac{103}{150}\right) * 0.358269 - \left(\frac{15}{150}\right) * 0.918296 - \left(\frac{32}{150}\right) * 0 \\
 &= 0.535148
 \end{aligned}$$

Demikian juga untuk atribut yang lainnya dilakukan proses perhitungan serupa, sehingga pada fase pencarian akar diperoleh rangkuman hasil perhitungan seperti diperlihatkan dalam Tabel 5.

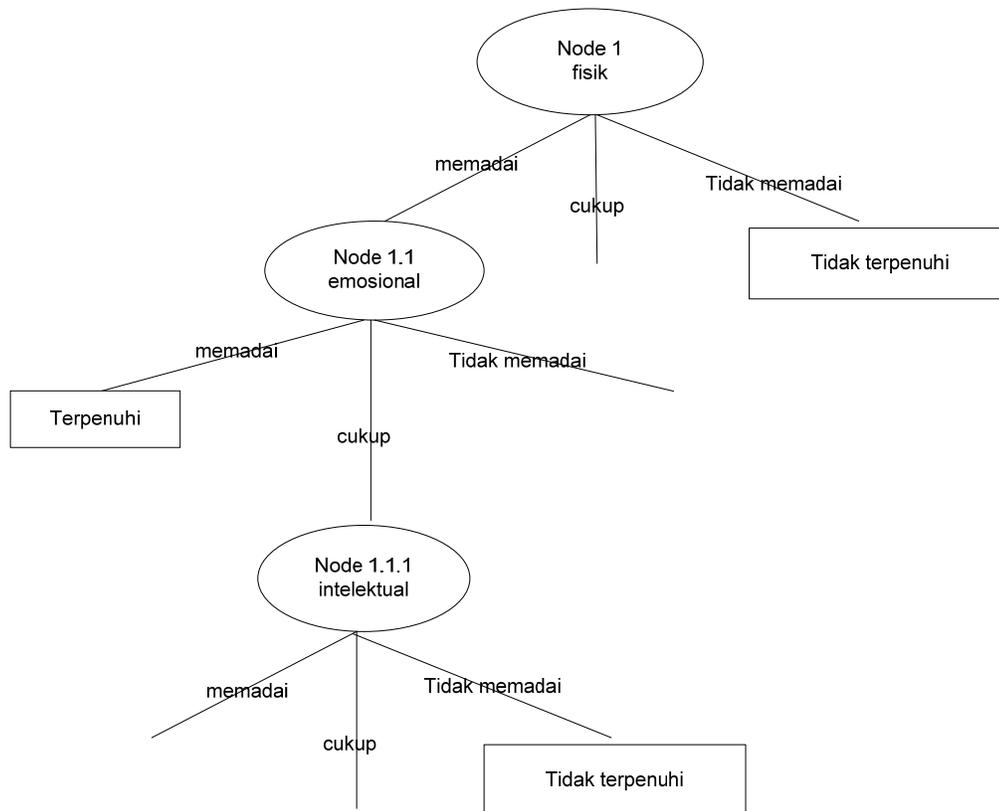
Tabel 5 Rangkuman hasil perhitungan entropy dan gain untuk root pohon keputusan

node	Atribut	jml kasus S	terpenuhi	tdk terpenuhi	Entrophy	Gain
1	Total	150	106	44	0.872988	
	Fisik					0.535148
	Memadai	103	96	7	0.358269	
	Cukup	15	10	5	0.918296	
	tdk memadai	32	0	32	0	
	intelektual					0.159055
	Memadai	76	65	11	0.596511	
	Cukup	65	41	24	0.95008	
	tdk memadai	9	0	9	0	
	emosional					0.413633
	Memadai	72	69	3	0.058842	
	Cukup	64	36	28	0.988699	
	tdk memadai	14	1	13	0.099278	
	sosial spiritual					0.18958
	Memadai	34	32	2	0.322757	
	Cukup	102	70	32	0.897427	
	tdk memadai	14	4	10	0	

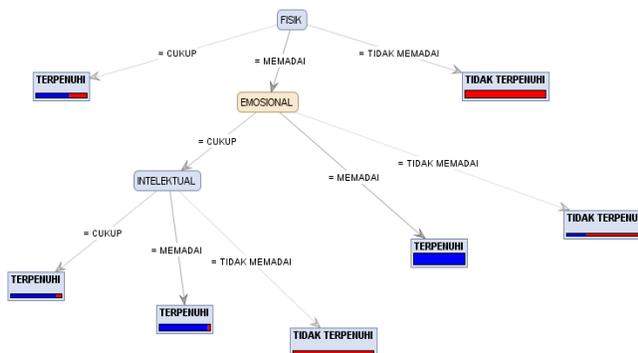
Dari rangkuman penghitungan entropy dan gain seperti Tabel 5, diperoleh root, seperti terlihat pada Gambar 3.



Gambar 3 Menentukan root



Gambar 4 menentukan node 1.1.1

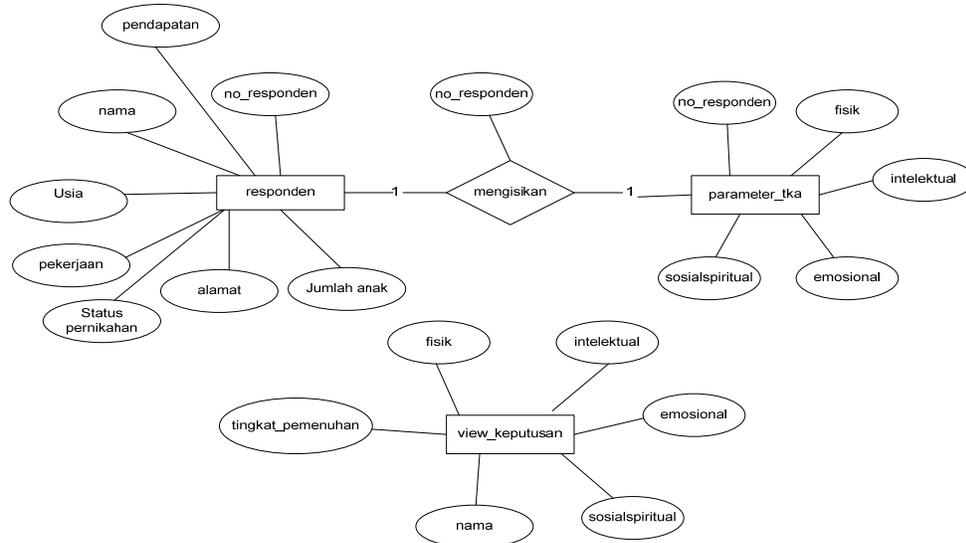


Gambar 5 pohon keputusan hasil rapidminer

Dari pohon keputusan pada gambar 5, ditentukan aturan-aturan sebagai berikut :

- R1 : IF (fisik=memadai) AND (emosional=memadai) THEN terpenuhi
- R2 : IF (fisik=memadai) AND (emosional=tidak memadai) THEN tidak terpenuhi
- R3 : IF (fisik=memadai) AND (emosional=cukup) AND (intelektual=memadai) THEN terpenuhi
- R4 : IF (fisik=memadai) AND (emosional=cukup) AND (intelektual=cukup) THEN terpenuhi
- R5 : IF (fisik=memadai) AND (emosional=cukup) AND (intelektual=tidak memadai) THEN tidak terpenuhi
- R6 : IF(fisik=cukup) THEN terpenuhi
- R7 : IF (fisik=tidak memadai) THEN tidak terpenuhi

Model data pada Gambar 6, merupakan rancangan basis data relasional yang akan diimplementasikan pada DBMS MySQL, terdiri dari dua entitas yaitu entitas responden dan entitas parameter_tka, sedangkan view keputusan diperoleh melalui eksekusi code view dari Gambar 7.



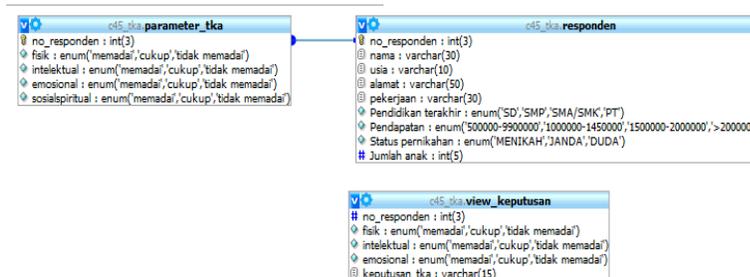
Gambar 6 Model data

create view view_keputusan as select no_responden, fisik, intelektual, emosional, CASE WHEN fisik='memadai' AND emosional='memadai' THEN 'terpenuhi' WHEN fisik='memadai' AND emosional='tidak memadai' THEN 'tidak terpenuhi' WHEN fisik='memadai' AND emosional='cukup' AND intelektual='memadai' THEN 'terpenuhi' WHEN fisik='memadai' AND emosional='cukup' AND intelektual='cukup' THEN 'terpenuhi' WHEN fisik='memadai' AND emosional='cukup' AND intelektual='tidak memadai' THEN 'tidak terpenuhi' WHEN fisik='cukup' THEN 'terpenuhi' WHEN fisik='tidak memadai' THEN 'tidak terpenuhi' END as keputusan_tka from parameter_tka

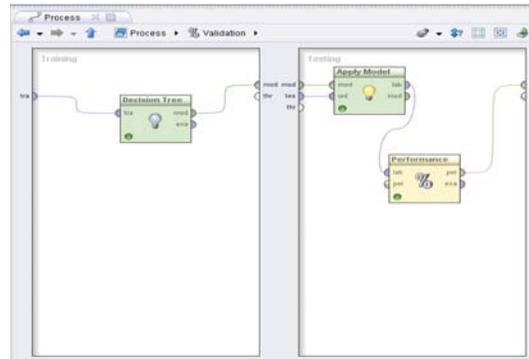
Gambar 7 Rancangan query view dalam DBMS MySQL

4. IMPLEMENTASI DAN PEMBAHASAN

Dilakukan implementasi basis data sistem penentuan tingkat kesejahteraan anak dengan dua table data masing-masing tabel responden, parameter Tingkat Kesejahteraan Anak (TKA) dan satu view, view keputusan, seperti diperlihatkan pada Gambar 7. Pada gambar 7 di tentukan model sistem pengujian menggunakan rapidminer.



Gambar 7 Implementasi basis data relasional dalam DBMS MySQL

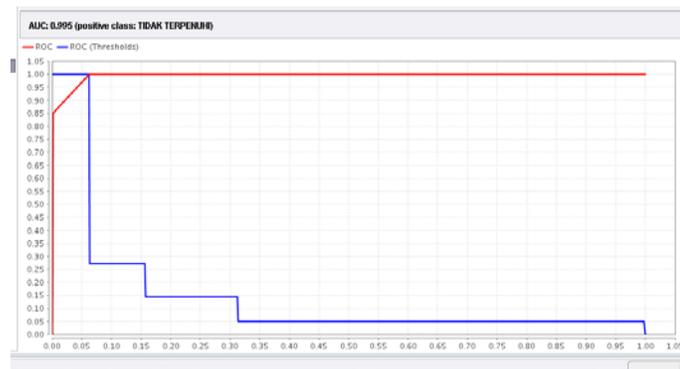


Gambar 8 model pada rapidminer

Akurasi dari model system pada Gambar 8 diperoleh sebesar 95, 65 % seperti diperlihatkan dalam Gambar 9 dan kurva ROC diperoleh pada Gambar 10.

accuracy: 95.56%			
	true TERPENUHI	true TIDAK TERPENUHI	class precision
pred. TERPENUHI	32	2	94.12%
pred. TIDAK TERPENUHI	0	11	100.00%
class recall	100.00%	84.62%	

Gambar 9 Nilai akurasi yang diperoleh dari model pada gambar



Gambar 10 Kurva ROC dari model pada gambar 8

5. KESIMPULAN

Algoritma C 4.5 dapat digunakan untuk menentukan tingkat kesejahteraan anak. Dengan algoritma C 4.5 diperoleh pohon keputusan yang dapat menyederhanakan aturan penentuan tingkat kesejahteraan anak. Diperoleh model sistem yang memiliki akurasi tinggi yaitu sebesar 95, 65 %. Penggunaan tools rapidminer dapat melengkapi proses penentuan pohon keputusan jika mendapat kendala menentukan keputusan, dengan cara diambil yang paling banyak jumlah kasusnya. Dapat diimplementasikan basis data relasional untuk tingkat kesejahteraan anak.

6. DAFTAR PUSTAKA

- [1] Berry, J., *Social Work with Children*, LSW-London, 1972.
- [2] Badan Koordinasi Keluarga Berencana Nasional (BKKBN) Indonesia.
- [3] Commission To Promote Sustainable Child Welfare, *Child Welfare Statistical Neighbour Model: Rationale & Prototype*, Working Paper Commission To Promote Sustainable Child Welfare. [online] 2012.[Cited: ,2013]
- [4] Chang, C., A study of applying data mining to early intervention for developmentally-delayed children, *ELSEVIER international journal of Expert Systems with Applications* 33 (2007) 407–412. [online] 2007.[Cited: ,2013]

- [5] Chor, B., McClelland, G., Jordan, N., Using CANS as a Placement Decision Support Algorithm to Predict Outcome of Youth in Child Welfare Placed in Residential Treatment, Northwestern University. [online] 2009.[Cited: ,2013]
- [6] Craw, S., Case Base Reasoning : Lecture 3: CBR Case-Base Indexing. <http://www.comp.rgu.ac.uk/staff/smc/teaching/kbp3/>. [online] 2005.[Cited: ,2013]
- [7] Eddyono,S.W., Pengantar Konvensi Hak Anak. <http://www.kontras.org/baru/Kovens%20Hak%20Anak.pdf>. [online] 2005.[Cited: ,2013]
- [8] Han, J., & Kamber, M., *Data Mining Concept and Tehniques*. San Fransisco: Morgan Kauffman, 2006.
- [9] Hurlock, E. B., *Psikologi Anak*, Jakarta : Erlangga, 1997.
- [10] Kadushin, A., *Child Welfare Services*, New York, 1974.
- [11] Kum, H., Moore, T.D., Dineen, M., Smith, S., Stewart, C.J., Decision Support Systems and Informatics for Child Welfare Administrative Data to Inform Practice and Policy, https://www.pal-tech.com/intranet/OCAN/3452_Kum,_H.-Decision_SUpport_Systems_and_Informatics.pdf. [online] 2012.[Cited: ,2013]
- [12] Lee, C., *Pertumbuhan dan Perkembangan Anak*, Jakarta : Penerbit Arcan, 1989.
- [13] Ramadan, R., Penerapan Pohon Untuk Klasifikasi Dokumen Teks Berbahasa Inggris, publikasi ilmiah. [online] 2007.[Cited: ,2013]
- [14] Sayad, S., Model Evalution-Clasification, dapat diakses pada: http://www.saedsayad.com/model_evaluation_c.htm. [online] 2013.[Cited: ,2013]
- [15] Suharto, E., *Pembangunan, Kebijakan Sosial dan Pekerjaan Sosial spektrum pemikiran*, Bandung : LSP-STKS, 1997.
- [16] UU no 4 tahun 1979 tentang Kesejahteraan Anak
- [17] Welling, L., Thomson, L., (2003), PHP And MYSQL Web Development, depeleoper's Library.