

SISTEM PENUNJANG KEPUTUSAN KESEHATAN UNTUK HIPERTENSI MENGUNAKAN ALGORITMA C4.5

Hilda Amalia¹, Evicienna²

¹⁾ Manajemen Informatika, AMIK Bina Sarana Informatika Jakarta
Jl. Ciledug Raya No. 168 Ulujami, Jakrta Selatan
hilda.ham@bsi.ac.id

²⁾ Komputerisasi Akuntansi, AMIK Bina Sarana Informatika Jakarta
Jl. Ciledug Raya No. 168 Ulujami, Jakrta Selatan
evicienna.eca@bsi.ac.id

ABSTRACT

Hypertension is a cardiovascular disease that causes 4.5% of the global disease burden. Hypertension is a major risk factor for heart problems and can be said as the "silent killer" because there are no specific signs and can cause serious illness if left untreated for a long time. Decision support system for hypertension can be used to obtain the results of the decision of the cases, one of them using the decision tree method. Hypertension data will be processed using the method of decision tree algorithm C4.5 through software RapidMiner and will result in a decision support rules, the value of accuracy, and AUC than the rule. After testing the accuracy of the values obtained on the C4.5 algorithm by 76.6%, AUC values for 0862 with a good level of diagnostic classification.

Keywords: Hypertension, decision support systems, Algorithm C4.5

I. PENDAHULUAN

Hipertensi dikenal secara luas sebagai penyakit kardiovaskular. Diperkirakan telah menyebabkan 4,5% dari beban penyakit secara global, dan prevelensinya hampir sama di Negara berkembang maupun Negara maju. Hipertensi merupakan salah satu faktor risiko utama gangguan jantung[2]. Hipertensi disebut juga dengan “*silent killer*” dikarenakan tidak adanya tanda-tanda khusus dan dapat menyebabkan penyakit yang serius jika tidak ditangani dalam waktu yang lama[1]. Oleh karena itu penting untuk membuat suatu sistem penunjang pengambilan keputusan untuk memprediksi hipertensi. Data hipertensi akan diolah menggunakan metode *decision tree* melalui *software RapidMiner*, sehingga diperoleh aturan-aturan dalam sistem penunjang keputusan.

Data mining adalah tentang memecahkan masalah dengan menganalisis data yang sudah ada dalam database[12]. Perkembangan gaya hidup masyarakat telah menghasilkan sebuah masyarakat yang sangat membutuhkan informasi. Informasi yang dibutuhkan masyarakat berasal dari kumpulan data-data yang bernilai. Data-data yang menghasilkan informasi tersebut hanya tersimpan dan terkunci dalam sebuah databases. Padahal dari data-data tersebut terdapat data potensial yang dapat kita jelajahi lebih lanjut dan menghasilkan informasi yang sangat berharga. Setiap saat data-data tersebut

bertambah dan menumpuk membentuk sebuah bukit data dan hanya menjadi sebuah “kuburan data”. Untuk itu diperlukan teknik data mining untuk menemukan informasi yang berharga dari tumpukan data dan diperlukan sebuah sistem pakar dalam pendagnosisan penyakit.

Data mining didefinisikan sebagai proses untuk menemukan pola pada data. Proses harus otomatis atau (lebih biasanya) semi-otomatis. Pola ditemukan harus lebih bermakna karena mereka mengakibatkan beberapa keuntungan[12]. Salah satu teknik data mining yang sering digunakan adalah algoritma C4.5.

Dalam penelitian ini akan dilakukan analisa data hipertensi menggunakan klasifikasi data mining yakni algoritma C4.5 dengan menggunakan tiga parameter yaitu usia, berat badan dan jenis kelamin.

II. LANDASAN TEORI

a. Sistem Penunjang Keputusan (*Decision Support System*)

Sistem Penunjang Keputusan (SPK) atau dikenal juga dengan *Decision Support System* (DSS) adalah salah satu subsistem dari Sistem Informasi Berbasis Komputer atau *Computer Based System Information* (CBIS) yang dapat menyediakan informasi yang berguna bagi proses pengambilan keputusan ketika menghadapi sebuah masalah semi terstruktur yang spesifik. Informasi sebagai output dari SPK,

dapat disajikan dalam bentuk laporan yang dihasilkan melalui perhitungan model matematika[6].

b. Data Mining

Data mining telah menarik banyak perhatian dalam dunia sistem informasi dan dalam masyarakat secara keseluruhan dalam beberapa tahun terakhir, karena ketersediaan luas dalam jumlah besar data dan kebutuhan segera untuk mengubah data tersebut menjadi informasi yang berguna dan pengetahuan. Informasi dan pengetahuan yang diperoleh dapat digunakan untuk aplikasi mulai dari pasar analisis, deteksi penipuan, dan retensi pelanggan, untuk pengendalian produksi dan ilmu pengetahuan eksplorasi[4]. Adanya ketersediaan data yang melimpah, kebutuhan akan informasi atau pengetahuan sebagai sarana pendukung dalam pengambilan keputusan baik bagi individu, organisasi, perusahaan dan pemerintahan.

Dalam data mining data disimpan secara elektronik dan diolah secara otomatis, atau setidaknya disimpan dalam komputer. Data mining adalah tentang menyelesaikan masalah dengan menganalisa data yang telah ada dalam database[12].

c. Sistem Pakar

Sistem pakar merupakan suatu program aplikasi komputerisasi yang berusaha menirukan proses penalaran dari seorang ahlinya dalam memecahkan masalah spesifikasi atau bisa dikatakan merupakan duplikat dari seorang pakar karena pengetahuannya disimpan didalam basis pengetahuan untuk diproses pemecahan masalah (Hamdani,2010)[3].

d. Algoritma C4.5

Algoritma C4.5 dan pohon keputusan (*decision tree*) merupakan dua mode yang tidak terpisahkan, karena untuk membangun sebuah pohon keputusan, dibutuhkan algoritma C4.5. Dari akhir tahun 1970 sampai awal 1980-an J.Ross Quinlan, melakukan pengembangan terhadap algoritma *decision tree* yakni ID3 (*Iterative Dichotomiser*). Kemudian Quinlan juga menghadirkan algoritma C4.5, yang menjadi awal dari algoritma *supervised learning* yang terbaru. Di tahun 1984 sebuah kelompok statistik (L.Breiman, J. Fridman, R. Olshen dan C.Stone) mempublikasikan *Classification and Regresssion Tree* (CART), yang menggambarkan generasi *binary decision tree*[6].

Desicion Tree menyerupai struktur *flowchart*, yang masing-masing internal node-nya dinyatakan sebagai atribut pengujian, setiap cabang mewakili output dari pengujian, dan setiap node daun (*terminal node*) menentukan label class. Node paling atas dari sebuah pohon

adalah node akar[4]. Salah satu metode klasifikasi yang menarik melibatkan konstruksi pohon keputusan, koleksi node keputusan, terhubung oleh cabang-cabang, memperpanjang bawah dari simpul akar sampai berakhir di node daun. Dimulai di node root, yang oleh konvensi ditempatkan di bagian atas dari diagram pohon keputusan, atribut diuji pada node keputusan, dengan setiap hasil yang mungkin dihasilkan dalam suatu cabang. Setiap cabang kemudian mengarah baik ke node lain keputusan atau ke node daun untuk mengakhiri[8].

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5, yaitu[7]:

1. Menyiapkan data training. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai gain dari masing-masing atribut, nilai gain yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung gain dari atribut, hitung dahulu nilai entropy yaitu:

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

Keterangan:

S : himpunan kasus

A : atribut

N : jumlah partisi S

Pi : proporsi dari Si terhadap S

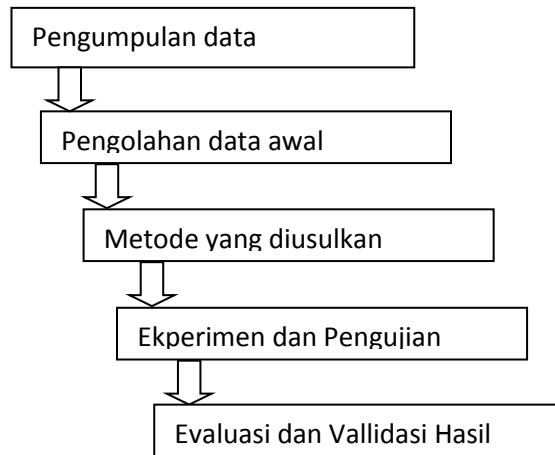
3. Kemudian hitung nilai gain dengan metode informasi gain:

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

4. Ulangi langkah ke-2 hingga semua tupelo terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:
 - a. Semua tupel dalam node N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam tupel yang dipartisi lagi.
 - c. Tidak ada tupel di dalam cabang yang kosong.

III. METODOLOGI PENELITIAN.

Pada penelitian ini akan digunakan data kesehatan yaitu usia, berat badan dan jenis kelamin. Berikut adalah tahapan penelitian yang dilakukan dalam penelitian ini:



Gambar 1. Tahapan Penelitian

a. Pengumpulan data

Teknik pengumpulan data ialah teknik atau cara-cara yang dapat digunakan untuk menggunakan data[9]. Data yang dikumpulkan diperoleh dari buku catatan kesehatan pasien dan jumlah sampel yang dikumpulkan sebanyak 85 sampel. Setiap keluarga mempunyai satu buku dan di dalam satu buku tersebut terdapat lebih dari satu orang yang dicatat. Di buku kesehatan tersebut terdapat data-data tentang pasien yang menderita hipertensi yaitu berupa jenis kelamin, berat badan, usia dan tekanan darah pasien. Dalam pengumpulan data terdapat sumber data, sumber data yang dihimpun langsung oleh peneliti disebut dengan sumber primer, sedangkan apabila melalui tangan kedua disebut sumber sekunder[9].

Berikut disajikan data-data hipertensi yang dikumpulkan:

Tabel 1. Data Hipertensi

Jenis Kelamin	Umur	berat badan	Tekanan Darah
P	35	55	110/70
L	46	80	130/90
P	26	52	120/70
P	31	60	120/70
L	44	55	170/100
L	45	70	150/100
L	44	72	140/90
L	42	65	160/110
L	70	70	150/100
P	49	65	170/100
P	30	55	110/80

P	22	52	120/80
P	35	54	120/70
P	35	60	150/100
L	70	65	180/120
L	44	40	120/80
P	50	70	140/100
L	42	65	150/100
P	45	57	120/70
P	40	70	130/80
P	46	70	150/100
P	43	52	140/100
P	40	70	120/80
L	40	80	150/100
L	52	60	110/70
P	25	50	110/80
P	35	52	120/80
P	28	45	110/70
P	70	40	160/100
L	70	72	190/110
P	26	60	120/80
P	45	65	110/70
P	41	45	110/80
L	23	60	110/70
P	28	56	120/70
L	52	65	120/80
L	40	73	130/80
P	32	57	110/70
P	44	70	150/100
P	34	40	110/70
L	29	56	120/80
P	37	50	150/100
L	32	57	130/70
L	42	60	120/70
L	35	61	110/70
P	47	56	110/70
P	26	55	120/80
L	45	72	150/100
L	56	63	150/100
P	62	40	160/100
L	45	48	120/80
L	44	52	120/70
L	61	64	140/100
L	41	67	120/70

P	46	50	110/70
P	40	50	120/70
P	32	51	110/70
L	45	60	120/80
P	25	65	130/80
P	41	45	120/70

Keterangan:

Usia : disajikan dalam tahun
 Berat badan : disajikan dalam Kilogram (Kg)
 Jenis Kelamin : L(Laki-Laki) dan P(Perempuan)

b. Pengolahan Awal Data

Data yang diperoleh kemudian diolah, tidak semua data digunakan, jumlah data yang berhasil dikumpulkan adalah sebanyak 85 data. Kemudian data tersebut diolah sehingga menghasilkan 60 data yang akan dianalisa lebih lanjut. Hal ini dikarenakan pada data awal yang berjumlah 85 data dilakukan teknik pengolahan awal data yaitu bertujuan untuk memperoleh data yang berkualitas. Berikut teknik yang dilakukan dalam pengolahan awal data[11]:

1. Data Validation: yaitu teknik menghilangkan menghapus data yang tidak lengkap(outer/noise). Dalam penelitian ini terdapat 25 data yang tidak lengkap sehingga dihapus, sehingga data yang digunakan untuk analisa adalah 60 data.
2. Data Integration and Transformation: yaitu teknik mengubah dan mengabungkan data sehingga diperoleh data yang valid. Dalam penelitian ini dilakukan pengubahan data dari numeric menjadi kategorikal.

Tabel 2. Kategori Tekanan Darah

Kategori	Sistolik (mmHg)	Diastolik (mmHg)
Normal	< 120	< 80
Pre-hipertensi	120 – 139	80 – 89
Hipertensi Stage 1	140 – 159	90 – 99
Hipertensi Stage 2	>=160	>=100

Tabel 2. diatas adalah kategori tekanan darah menurut *Joint Comitte on Detection Evaluation and Treatment of High Blood Presure*[10]. Dari tabel diatas maka dapat

dikategorikan bahwa seseorang yang menderita hipertensi adalah dengan tekanan sistolik diatas 139 mmHg atau tekanan diastolik yang diatas 89 mmHg. Dibawah tekanan diatas maka dapat dikategorikan tidak menderita hipertensi.

Setelah didapat kriteria untuk data usia dan berat badan data mentah yang diperoleh dikelompokkan ke dalam kriteria – kriteria di atas, sehingga diperoleh data sebagai berikut:

Tabel 3. Hasil Pengelompokan Kriteria

Jenis Kelamin	Umur	berat badan	Hasil
P	35	55	TIDAK
L	46	80	YA
P	26	52	TIDAK
P	31	60	TIDAK
L	44	55	YA
L	45	70	YA
L	44	72	YA
L	42	65	YA
L	70	70	YA
P	49	65	YA
P	30	55	TIDAK
P	22	52	TIDAK
P	35	54	TIDAK
P	35	60	YA
L	70	65	YA
L	44	40	TIDAK
P	50	70	YA
L	42	65	YA
P	45	57	TIDAK
P	40	70	TIDAK
P	46	70	YA
P	43	52	YA
P	40	70	TIDAK
L	40	80	YA
L	52	60	TIDAK
P	25	50	TIDAK
P	35	52	TIDAK
P	28	45	TIDAK
P	70	40	YA
L	70	72	YA
P	26	60	TIDAK
P	45	65	TIDAK

P	41	45	TIDAK
L	23	60	TIDAK
P	28	56	TIDAK
L	52	65	TIDAK
L	40	73	TIDAK
P	32	57	TIDAK
P	44	70	YA
P	34	40	TIDAK
L	29	56	TIDAK
P	37	50	YA
L	32	57	TIDAK
L	42	60	TIDAK
L	35	61	TIDAK
P	47	56	TIDAK
P	26	55	TIDAK
L	45	72	YA
L	56	63	YA
P	62	40	YA
L	45	48	TIDAK
L	44	52	TIDAK
L	61	64	YA
L	41	67	TIDAK
P	46	50	TIDAK
P	40	50	TIDAK
P	32	51	TIDAK
L	45	60	TIDAK
P	25	65	TIDAK
P	41	45	TIDAK

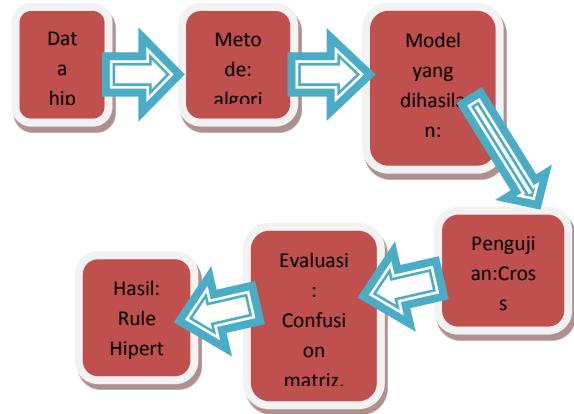
Tabel atribut yang digunakan, terlihat pada tabel dibawah ini:

Tabel 4. Pengelompokan Atribut

Atribut	Nilai
Jenis kelamin	Perempuan, Laki-Laki
Usia	<=41 dan >41 <=54 dan >54
Berat badan	<=62 dan > 62

c. Metode Yang Diusulkan

Dalam penelitian ini metode yang diusulkan untuk menganalisa data dan menemukan pola yang tersembunyi dari data hipertensi adalah algoritma C4.5. Berikut bagan ilustrasi untuk metode yang diusulkan:



Gambar 2. Metode Usulan

IV. HASIL PENELITIAN DAN PEMBAHASAN

a. Eksperimen Dan Pengujian Metode

Eksperimen dan Pengujian Metode Algoritma C4.5

1. Menyiapkan data training, yaitu menggunakan data pada tabel data ya
2. Menghitung Entropy keseluruhan total kasus yang “Ya” Hipertensi dan “Tidak” Hipertensi. Dari data training diketahui jumlah data yang “Ya” Hipertensi adalah 30 dan jumlah data yang “Tidak” Hipertensi berjumlah 30 data. Data keseluruhan adalah 60. Berikut adalah perhitungan entropinya:

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

$$= -22/60 \log(2) 22/60 - (38/60 \log(2) 38/60)$$

$$= 0,948078244$$

3. Menghitung nilai entropi dan nilai *gain* masing-masing atribut. Nilai *gain* tertinggi adalah atribut yang menjadi *root* dari pohon keputusan yang akan dibuat. Berikut perhitungan entropi untuk atribut usia.

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

$$E_{Perempuan} [9,25] = -9/60 \log(2) 9/60 - (25/60 \log(2) 25/60) = 0,94$$

$$E_{laki-laki} [13,13] = 13/60 \log(2) 13/60 - (13/60 \log(2) 13/60) = 0,96$$

$$Entropi Jenis Kelamin [35,26] = 35/60 * 0,94 + 26/60 * 0,96 = 0,96$$

$$Gain Jenis Kelamin = 0,95 - 0,96 = -0,013$$

Tabel 5 Perhitungan entropi atribut dan gain

Simpul	Kasus	Ya	Tidak	Entropi	Gain
Total data	60	22	38	0,95	
Jenis Kelamin					
Perempuan	35	9	25	0,94	-
Laki-Laki	26	13	13	0,96	0,013
Umur					
<=54	53	15	38	0,5	0,46
>54	7	7	0	0,36	
<=41,5	30	3	27	0,155	0,38
>41,5	30	19	11	0,97	
Berat Badan					
<=62,5	37	6	31	0,82	0,11
>62,5	23	16	7	0,87	

Dari tabel perhitungan entropi diatas diketahui gain tertinggi yaitu 0,46 oleh karena itu atribut berat badan <= 54 dan > 54 merupakan root dari pohon keputusan yang dihasilkan.

- Ulangi langkah kedua sampai tidak bisa terpartisi, yaitu menghitung entropi dan gain untuk atribut selanjutnya untuk menentukan simpul 1.1 dan simpul 1.2, nilai dihitung berdasarkan umur > 54. Berikut tabel data yang termasuk kategori umur > 54:

Tabel 6. Kategori Umur > 54

Jenis Kelamin	Umur	berat badan	hasil
L	70	70	YA
L	70	65	YA
P	70	40	YA
L	70	72	YA
L	56	63	YA
P	62	40	YA
L	61	64	YA

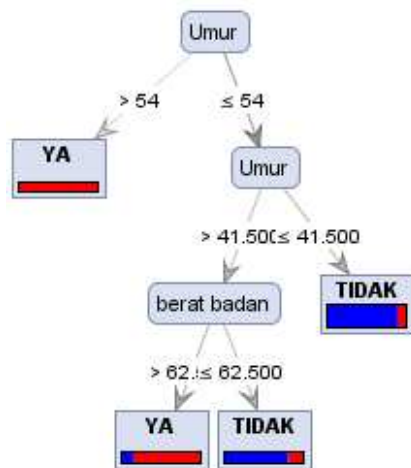
Dilihat dari tabel 4.2 umur > 54 hasil yang diperoleh adalah Ya untuk semua data sehingga atribut umur > 54 tidak dapat dipartisi lagi. Berikut tabel data yang termasuk pada kategori umur <= 54:

Tabel 7 Kategori Umur <= 54

Jenis Kelamin	Umur	berat badan	hasil
P	35	55	TIDAK
L	46	80	YA
P	26	52	TIDAK
P	31	60	TIDAK
L	44	55	YA
L	45	70	YA
L	44	72	YA
L	42	65	YA
P	49	65	YA
P	30	55	TIDAK
P	22	52	TIDAK
P	35	54	TIDAK
P	35	60	YA
L	44	40	TIDAK
P	50	70	YA
L	42	65	YA
P	45	57	TIDAK
P	40	70	TIDAK
P	46	70	YA
P	43	52	YA
P	40	70	TIDAK
L	40	80	YA
L	52	60	TIDAK
P	25	50	TIDAK
P	35	52	TIDAK
P	28	45	TIDAK
P	26	60	TIDAK
P	45	65	TIDAK
P	41	45	TIDAK
L	23	60	TIDAK
P	28	56	TIDAK
L	52	65	TIDAK
L	40	73	TIDAK
P	32	57	TIDAK
P	44	70	YA
P	34	40	TIDAK
L	29	56	TIDAK
P	37	50	YA
L	32	57	TIDAK

L	42	60	TIDAK
L	35	61	TIDAK
P	47	56	TIDAK
P	26	55	TIDAK
L	45	72	YA
L	45	48	TIDAK
L	44	52	TIDAK
L	41	67	TIDAK
P	46	50	TIDAK
P	40	50	TIDAK
P	32	51	TIDAK
L	45	60	TIDAK
P	25	65	TIDAK
P	41	45	TIDAK

Berikut adalah pohon keputusan yang dihasilkan dari olahan RapidMiner:



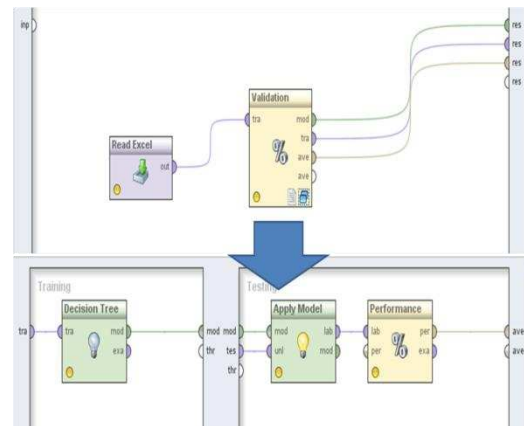
Gambar 3 Pohon Keputusan Hipertensi

Dari pohon keputusan tersebut berikut adalah rule yang diperoleh:

1. R1: jika umur > 54 maka hasil YA Hipertensi
2. R2: Jika umur ≤ 54 dan umur > 41,5 dan berat badan > 62,5 maka hasil Ya Hipertensi
3. R3: jika umur ≤ 54 dan umur > 41,5 dan berat badan ≤ 62,5 maka hasil Tidak Hipertensi
4. R4: jika umur ≤ 54 dan umur ≤ 41,5 maka hasil Tidak Hipertensi.

b. Evaluasi dan Validasi

Berikut pengujian yang dilakukan menggunakan cross validation dengan tools RapidMiner.



Gambar 4. Desain Pengujian RapidMiner

Metode klasifikasi bisa dievaluasi berdasarkan kriteria seperti tingkat akurasi, kecepatan, kehandalan, skabilitas dan interpretabilitas[10]. Berikut tabel confusion Matrix algoritma C4.5, dari tabel diketahui tingkat akurasi 76,6.

Tabel 8. Cunfusion Matrix Algoritma C4.5

accuracy 76.67% +/- 18.56% (mean 76.67%)			
	true TIDAK	true YA	class precision
pred TIDAK	33	9	78.57%
pred YA	5	13	72.22%
class recall	88.84%	59.09%	

Dalam penelitian ini dilakukan prediksi data hipertensi menggunakan algoritma C4.5, hasil yang didapat dari 60 kasus adalah: diketahui sebanyak 33 kasus diprediksi Tidak Hipertensi sesuai dengan data yang ada, sebanyak 9 kasus diprediksi Tidak Hipertensi tetapi pada data yang ada kenyataannya Ya Hipertensi. Sebanyak 5 kasus diprediksi Ya Hipertensi pada kenyataannya Tidak Hipertensi, dan sebanyak 13 kasus diprediksi Tidak Hipertensi pada kenyataannya yaitu sesuai dengan data yang ada.

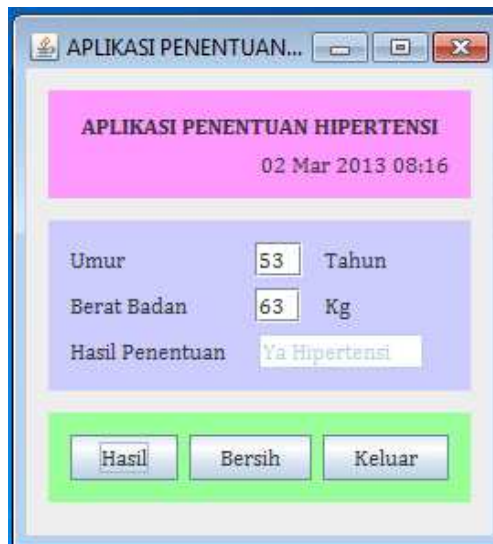
Berikut grafik Area Under Curve (AUC) yang dihasilkan dari tools RapidMiner:



Gambar 4. Nilai AUC dalam Grafik ROC

Dari gambar 4.3 diatas diketahui nilai AUC yang dihasilkan adalah 0,862 dengan tingkat diagnosa *good classification*.

Berikut aplikasi yang dihasilkan berdasarkan rule yang ada menggunakan aplikasi java NeatBeans:



Gambar 5. Aplikasi Penentuan Hipertensi

Rule yang digunakan pada gambar 5 adalah sesuai dengan rule yang didapat dari hasil olahan RapidMiner. Input umur dan berat badan sesuai dengan ketentuan yang telah ditetapkan sebelumnya. Dan hasil penentuan yang didapatpun sesuai dengan rule yang telah ditetapkan juga.

V. KESIMPULAN

Dalam penelitian ini dilakukan prediksi pengolahan data hipertensi dengan menggunakan tiga parameter yaitu jenis, kelamin, berat badan dan umur. Data diolah menggunakan metode algoritma C4.5, dari hasil running data menggunakan tools Rapidminer diketahui tingkat akurasi yaitu 0,79% dan AUC 0,862 dengan tingkat diagnosa *good classification*. Dan

dibuatkan aplikasi prediksi hipertensi berdasarkan rule yang ada.

DAFTAR PUSTAKA

- Abdullah, A. A., Zakaria, Z., & Mohammad, F. N. (2011). Design and Development of Fuzzy Expert System for Diagnosis of Hypertension. Second International Conference on Intelligent Systems, Modelling and Simulation , 978-0-7695-4336-9.
- Depkes. (2006). Pharmaceutical Care Untuk Penyakit Hipertensi. Jakarta: Direktorat Bina Farmasi Komunitas Dan Klinik Ditjen Bina Kefarmasian Dan Alat Kesehatan Departemen Kesehatan.
- Hamdani. (2010). Sistem Pakar Untuk Diagnosa Penyakit Mata Pada Manusia. Jurnal Informatika Mulawarman. Vol 5 No. 2 Juli 2010 - 13
- Han, J., & Kamber, M. (2007). Data Mining Concepts and Techniques. San Fransisco: Mofgan Kaufan Publisher.
- Jr, McLeod Raymond. Sistem Informasi Manajemen. Edisi ketujuh. Jilid satu. PT. Penhelindo.Jakarta. 2001.
- Kusrini, & Luthfi, E. T. (2009). Algoritma Data Mining. Yogyakarta: Andi Publishing.
- Larose, D. T. (2005). Discovering Knowledge in Databases. New Jersey: John Willey & Sons Inc.
- Riduwan. (2008). Metode dan Teknik Menyusun Tesis. Bandung: Alfabeta.
- The Joint National Committee on detection, Evaluation and Treatment of High Blood Pressure: The Seventh Report of the Joint National Committee on Detection, Evaluation and Treatment of High Blood Pressure; Hypertension 2003;42:1206-52.
- Vercellis, C. (2009). Business Intelligent: Data Mining and Optimization for Decision Making. Southern Gate: John Willey & Sons Inc.
- Witten, H. I., Eibe, F., & Hall, A. M. (2011). Data Mining Machine Learning Tools and Techiques. Burlington: Morgan Kaufmann Publisher.