

## IMPLEMENTASI DATA MINING UNTUK MENGETAHUI FAKTOR KELAYAKAN DONOR DARAH UTD KOTA TASIKMALAYA MENGUNAKAN ALGORITMA C4.5

Abdul Latif<sup>1</sup>; Dini Silvi Purnia<sup>2</sup>

Program Studi Sistem Informasi<sup>1,2</sup>  
Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri<sup>1,2</sup>  
<http://nusamandiri.ac.id><sup>1,2</sup>  
abdullatif56@gmail.com<sup>1</sup> dini.dlv@nusamandiri.ac.id<sup>2</sup>



**Abstract**— Blood donor is a humanity activity which a person voluntarily donates blood to be stored in a blood bank which is then used for blood transfusions. UTD PMI Tasikmalaya is a health service facility that organizes blood donors, blood providers, and blood distribution. When there are many voluntary donors at UTD PMI Tasikmalaya, long lines often occur during blood donor feasibility checks. Sometimes long lines make donors voluntarily undo their intention to donate, so we need a model for the efficiency of blood donor feasibility checks. Data Mining is a combination of a number of computer disciplines that is defined as the process of discovering new patterns from very large data sets. So in this research, writer was used data mining with Rapidminer software and using the Decision Tree C4.5 Algorithm method to predict of blood donor feasibility based on Age, Weight, Blood Pressure, and Hemoglobin. From the results of this research, hemoglobin is the variable that most determines the eligibility of blood donors with an accuracy rate of 97.69% which means the accuracy of this model is very good

**Keywords:** Blood Donor, C4.5 Algorithm, Data Mining, Decision Tree, Prediction.

**Abstrak**— Donor darah merupakan kegiatan kemanusiaan dimana seseorang dengan sukarela menyumbangkan darahnya untuk disimpan di bank darah yang kemudian digunakan untuk transfusi darah. UTD PMI Tasikmalaya merupakan fasilitas layanan kesehatan yang menyelenggarakan donor darah, penyedia darah, dan pendistribusian darah. Ketika banyaknya pendonor sukarela di UTD PMI Tasikmalaya, maka antrean panjang sering terjadi pada saat pemeriksaan kelayakan donor darah. Terkadang antrean panjang membuat pendonor sukarela mengurungkan niatnya untuk mendonor, sehingga diperlukan sebuah model untuk efisiensi pemeriksaan kelayakan donor darah. Data Mining merupakan gabungan sejumlah disiplin ilmu komputer yang didefinisikan sebagai proses penemuan pola-pola baru dari kumpulan-kumpulan data sangat besar. Sehingga dalam penelitian ini, peneliti memanfaatkan data mining menggunakan software Rapidminer dan menggunakan metode Decision Tree Algoritma C4.5 untuk menentukan prediksi kelayakan donor darah berdasarkan Usia, Berat Badan, Tekanan Darah, dan Hemoglobin. Dari hasil penelitian ini, hemoglobin adalah variabel paling menentukan kelayakan donor darah dengan tingkat akurasi 97,69% yang berarti akurasi model ini sangat baik

**Kata kunci:** Algoritma C4.5, Data Mining, Decision Tree, Donor Darah, Prediksi

### PENDAHULUAN

Donor darah merupakan kegiatan kemanusiaan dimana seseorang dengan sukarela menyumbangkan darahnya untuk disimpan di bank darah untuk kemudian digunakan untuk transfusi darah. Menurut Departemen Kesehatan Republik Indonesia dalam (Bayususetyo et al., 2017):

Donor darah adalah proses menyalurkan darah atau unsur-unsur darah dari satu orang ke

sistem peredaran orang lainnya. Banyak orang yang tidak tahu tentang manfaat donor darah bagi kesehatan (Wardati et al., 2019). Padahal dengan melakukan donor darah, maka sel-sel darah di dalam tubuh menjadi lebih cepat terganti (Bayususetyo et al., 2017) dengan yang baru. Dengan meningkatnya permintaan suplai darah di masyarakat, persediaan darah yang mencukupi sangat dibutuhkan (Atmaja et al., 2017)(Devi, 2018). Meskipun demikian, pendonor harus terlebih dahulu menjalani pemeriksaan kesehatan,

baik pengukuran tekanan darah, golongan darah, kadar hemoglobin (Hb) maupun konsultasi medis.

Salah satu lembaga atau instansi yang melayani serta menyediakan fasilitas untuk donor darah adalah Unit Transfusi Darah (UTD) Palang Merah Indonesia (PMI). UTD PMI Kota Tasikmalaya merupakan tempat pelayanan donor darah dari masyarakat Kota Tasikmalaya yang donor secara sukarela maupun untuk transfusi keluarga yang sedang membutuhkan.

Ketika kebutuhan darah semakin banyak, ataupun ketika terdapat donor masal, tentunya menimbulkan antrean yang sangat panjang dan menyita waktu yang cukup lama untuk melakukan pemeriksaan kelayakan donor darah. Dalam kegiatan donor darah sukarela secara masal di tempat ramai, misalkan di acara "*car free day*" (Paraswati, 2014) antrean yang panjang membuat para pendonor sukarela terkadang bosan dan mengurungkan niatnya untuk mendonor karena alasan waktu dan sebagainya.

Berdasarkan uraian diatas, untuk memaksimalkan pelayanan serta terpenuhinya kebutuhan akan darah semakin cepat, maka diperlukan sebuah prediksi cepat untuk mengetahui kelayakan pendonor PMI Kota Tasikmalaya. Untuk menentukan kelayakan pendonor bisa di klasifikasikan secara cepat menggunakan prediksi berdasarkan kriteria paling berpengaruh dalam menentukan kelayakan pendonor. Untuk mendapatkan prediksi kriteria paling berpengaruh maka arsip formulir beberapa bulan lalu digunakan sebagai bahan penelitian menggunakan *data mining* untuk memprediksi kelayakan calon pendonor darah berdasarkan berat badan, usia, tekanan darah (*sistol* dan *diastol*), serta kadar hemoglobin calon pendonor. Sehingga dapat dilakukan proses *mining* untuk menentukan pola prediksi menentukan kelayakan pendonor dengan cepat. Serta pendaftaran calon pendonor menjadi lebih efisien dan penanganan bisa dilakukan dengan lebih cepat.

Untuk memprediksi layak atau tidaknya pendonor, maka digunakan Algoritma C4.5 (Atmaja et al., 2017) berdasarkan : berat badan, usia, tekanan darah (*sistol* dan *diastol*), serta kadar hemoglobin calon pendonor sebagai variabel pendukung.

Algoritma C4.5 akan menghasilkan pohon keputusan. Pohon keputusan atau *decision tree* adalah pohon yang digunakan sebagai prosedur penalaran untuk mendapatkan jawaban dari masalah yang dimasukkan. Sehingga dengan pohon keputusan, manusia bisa dengan mudah diidentifikasi dan dilihat dari faktor-faktor yang mempengaruhi layak atau tidaknya seorang calon pendonor.

## BAHAN DAN METODE

### Metode Penelitian

Metode penelitian yang dilakukan penulis diawali dengan mengumpulkan data dengan berbagai langkah, diantaranya:

#### 1) Studi Pustaka

Langkah pertama yang dilakukan penulis adalah mencari literatur penelitian terdahulu untuk mencari kriteria apa saja yang bisa menentukan kelayakan pendonor, dan merangkumnya ke dalam skripsi ini. Studi pustaka ini juga dilakukan untuk melengkapi teori-teori yang dibutuhkan, penulis melakukan metode studi kepustakaan dengan cara menggunakan referensi dari buku-buku, jurnal, internet, dan berbagai literatur lainnya sesuai dengan isi dari penyusunan skripsi tanpa melewati penulisan sumbernya.

#### 2) Wawancara

Penulis melakukan tanya jawab dengan pengelola dan pekerja UTD Kota Tasikmalaya serta kepada beberapa ahli terkait kriteria kelayakan donor darah.

#### 3) Dokumentasi

Membuat sebuah simpulan kriteria kelayakan calon pendonor berdasarkan data yang diperoleh dari lembar formulir yang diisi calon pendonor.

### Metode Analisis Data

Data pendonor yang didapatkan dari proses pengumpulan data sebanyak 2.548 record dan memiliki 19 atribut. Selanjutnya untuk mendapatkan data yang berkualitas, maka diperlukan tahap analisa di antaranya sebagai berikut:

#### Data Cleaning

Data yang berkualitas sangat mempengaruhi kualitas keputusan yang akan diperoleh. Data yang tidak berkualitas adalah data yang nilai atributnya hilang, error dan data yang tidak konsisten dalam pengisian atributnya. Untuk mendapatkan data yang berkualitas dilakukan tahapan dalam data cleaning sebagai berikut:

- Menghilangkan data yang tidak lengkap terdapat 75 data dari 2.548 yang tergolong tidak lengkap, dan 17 data dari 2.473 tidak diisi dengan tensi sebagai kurang lengkap yang tersisa 2.456 record.
- Mengisi nilai-nilai yang hilang, data yang tidak lengkap (missing value).

### Data Integration and Transformation

Teknik yang digunakan untuk menganalisis data korelasi, atribut yang redudan disebut integration, sedangkan transformation berguna untuk meningkatkan akurasi dan efisiensi.

Algoritma C4.5 memiliki kelebihan memproses data yang bernilai nominal, ordinal maupun kontinyu. Sehingga nilai-nilai setiap atribut yang terdapat pada dataset tidak perlu ditransformasikan

**Data Reduction**

Dataset dapat direduksi dengan mengurangi jumlah atribut dan record supaya menjadi lebih sedikit tetapi tetap bersifat informatif. Memperoleh representasi dalam bentuk volume data yang telah berkurang jumlahnya namun tetap mendapatkan hasil analisis yang sama. Deskretisasi data merupakan bagian dari reduksi data bagian penting untuk data numerik. Berikut adalah keseluruhan atribut sebelum dilakukan reduction data. Berdasarkan standar minimal untuk donor darah yang digunakan PMI dan Departemen Kesehatan Republik Indonesia, data yang berkaitan dengan standar kelayakan donor darah tersebut parameternya adalah usia, berat badan, tekanan darah atau tensi, dan hemoglobin atau Hb, dan status sebagai label. Kemudian menghapus atribut selain dari atribut yang digunakan. Dapat dilihat pada tabel 1

Tabel 1. Atribut Data Penelitian

No	Atribut	Proses	Penjelasan
1	No.	Data Cleaning	No. Transfusi
2	Tanggal	Data Cleaning	Tanggal Transfusi
3	ID	Data Cleaning	ID Pendoron
4	Nama	Data Cleaning	Nama Pendoron
5	Alamat	Data Cleaning	Alamat Pendoron
6	No. HP	Data Cleaning	No. Hp Pendoron
7	Umur	Digunakan	Umur Pendoron
8	Gol (Rh)	Atribut	Golongan Darah (Rhesus)
9	JK	Data Cleaning	Jenis Kelamin (Pria/Wanita)
10	DP/DS	Data Cleaning	Donor Pengganti/ Sukarela
11	Donor Ke-	Data Cleaning	Donor ke berapa kalinya
12	No. Kantong	Data Cleaning	No. Kantong darah
13	HB	Data Cleaning	Kadar Hemoglobin Pendoron
14	Tensi	Digunakan	Tekanan Darah Pendoron
15	Berat Badan	Atribut	Berat Badan Pendoron
16	CC	Digunakan	Jumlah darah yang diambil
17	Petugas	Atribut	Nama Petugas Transfusi
18	Keterangan	Digunakan	Keterangan
19	Status	Digunakan Atribut	Sebagai Variabel Dependen (Layak dan Tidak Layak) Status

Sumber: (Latif & Purnia, 2019)

**HASIL DAN PEMBAHASAN**

**a. Pengolahan Data**

Dalam penelitian ini data yang digunakan adalah data pendonor darah dari UTD PMI Kota Tasikmalaya pada tahun 2019 sebanyak 2.456 record sampel. Data sampel pendonor darah dapat dilihat pada Tabel 1. Tabel 2 adalah pengelompokan dari pembagian dari masing-masing variabel.

Tabel 2. Data sampel dari data pendonor

No	Usia	Jenis Kelamin	HB	Tensi	Berat Badan	Status
1	Dewasa	Pria	Normal	Normal	Gemuk	Layak
2	Remaja	Pria	Normal	Normal	Sedang	Layak
3	Remaja	Pria	Normal	Normal	Gemuk	Layak
4	Remaja	Pria	Normal	Normal	Gemuk	Layak
5	Remaja	Pria	Normal	Normal	Obesitas	Layak
6	Remaja	Pria	Normal	Normal	Obesitas	Layak
7	Remaja	Wanita	Normal	Normal	Sedang	Layak
8	Dewasa 2	Pria	Normal	Normal	Gemuk	Layak
9	Remaja	Pria	Normal	Normal	Gemuk	Layak
...						
2.456	Remaja	Pria	Normal	Normal	Gemuk	Layak

Sumber: (UTD PMI Kab Tasikmalaya, 2019)

Tabel 3. Pembagian Kelas dan variabel

No	Atribut	Nilai
1	Umur	Remaja (17-30), Dewasa(31-50), Dewasa2(50-60), Lansia (>60)
2	Berat Badan	Kurus (berat badan < 45 kg), Sedang (berat badan 45-60 kg), Gemuk (berat badan 61-80 kg), Obesitas (berat badan > 80 kg)
3	Hemoglobin	Rendah (Hb < 12,5), Normal (Hb 12,5-17), Tinggi (Hb >18)
4	Tekanan Darah	Rendah (Tensi < 110/70), Normal (Tensi 110/70-150/90), Tinggi (Tensi > 160/100)

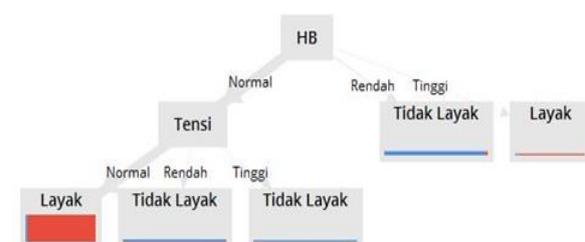
Sumber: (UTD PMI Kab Tasikmalaya, 2019)

Sumber : wawancara PMI Tasikmalaya (2019)

Atribut usia, berat badan, kadar hemoglobin, dan tekanan darah berdasarkan referensi dari WHO, JNC 7, dan DEPKES RI. Kemudian dirangkum dan disesuaikan dengan data yang dimiliki.

**b. Pemodelan**

Pemodelan dilakukan menggunakan *RapidMiner*. Hasilnya membentuk pohon keputusan dan terlihat hemoglobin yang menjadi node akar.



Sumber: (Latif & Purnia, 2019)

Gambar 1. **Decision Tree** Hasil Pengujian

Status kelayakan pendonor darah di UTD PMI Kota Tasikmalaya sebanyak 2035 orang layak dan 421 orang tidak layak. Dilihat dari Tensi darah dengan tensi tinggi sebanyak 46, normal 2274 dan rendah 136. Sedangkan untuk Hb atau hemoglobin, Hb tinggi sebanyak 22, Hb rendah 225 dan Hb Normal 2209. Hasil klasifikasi decision tree dari ExampleSet dimana yang menjadi root node adalah Hb atau Hemoglobin yang terbagi atas 3 cabang nilai atribut yaitu Tinggi, Normal, dan Rendah. Sedangkan Atribut Tensi atau tekanan darah berada dibawahnya. Bisa disimpulkan bahwa layak atau tidaknya calon pendonor mendonorkan darahnya bisa di prediksi dari tingkat Hb kemudian Tensi atau tekanan darahnya.

**Evaluasi dan Validasi Hasil**

Untuk mencari keseluruhan hasil klasifikasi pada seluruh data testing, digunakan software RapidMiner yang menghasilkan confusion matrix. Confusion matrix adalah suatu metode yang biasanya digunakan untuk melakukan perhitungan akurasi pada konsep data mining atau Sistem Pendukung Keputusan. Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Berdasarkan confision matrix pada Tabel 3, didapat akurasi dari pohon keputusan yang dihasilkan dari C4.5 untuk penentuan kelayakan calon pendonor darah adalah

97,69%. Hal ini berarti akurasi dari model ini dinyatakan sangat baik.

Tabel 4. Hasil evaluasi

Table View Plot View

accuracy: 97.69%

	true Tidak Layak	true Layak	class precision
pred. Tidak Layak	110	1	99.10%
pred. Layak	16	609	97.44%
class recall	87.30%	99.84%	

Sumber: (Latif & Purnia, 2019)

**Perhitungan Entropy dan Gain**

Dalam pembentukan sebuah pohon keputusan atau decision tree diperlukan perhitungan entropy dan gain dimana gain yang paling tertinggi akan menjadi root node. Karena peneliti menggunakan software sebagai analisa jadi peneliti perlu membuktikan mengapa Hb atau hemoglobin menjadi root node dari decision tree. Entropy (S) = 0,6609359 . Hitung entropy semua atribut, kemudian setelah semua entropy didapatkan, hitung gain dan mendapatkan hasil Gain (S,A) = 0,25027259 .

Berikut ini adalah penjelasan lebih terperinci dalam pembentukan decision tree dengan menggunakan algoritma C.45 yang dituangkan dalam tabel 5.

Tabel 5. Perhitungan Node 1

Node	Atribut	Nilai	Sum	Layak	Tidak Layak	Entrophy	Gain
1	Total		2456	2035	421	0,6609359	
	Usia						0,00535305
		Remaja	973	772	201	0,7348898	
		Dewasa	1240	1067	173	0,5829702	
		Dewasa 2	243	196	47	0,7085623	
	HB						0,25027259
		Rendah	225	5	220	0,1537422	
		Normal	2209	2009	200	0,4382654	
		Tinggi	22	21	1	0,2667650	
	Tensi						0,20310138
		Rendah	136	1	135	0,0626827	
		Normal	2274	2033	241	0,4876721	
		Tinggi	46	1	45	0,1510970	
	Berat Badan						0,01162637
		Kurus	2	0	2	0,0000000	
		Sedang	712	544	168	0,7882487	
		Gemuk	1326	1123	203	0,6175243	
		Obesitas	416	368	48	0,5159469	

Sumber: (Latif & Purnia, 2019)

Berdasarkan Tabel 5, entropy totalnya adalah 0,6609359 langkah selanjutnya yaitu menghitung entropy dari tiap-tiap atribut untuk menentukan node dari sebuah pohon keputusan. Dari hasil

perhitungan diatas diketahui gain yang paling besar adalah Hb (Hemoglobin) dimana atribut Hb mendapat hasil 0,25027259, hasil gain dari Hb didapatkan dari hasil normalisasi dengan

menggunakan gain ratio. Setelah node akar atau node 1 ditentukan, kemudian cari node selanjutnya. Ketika Hb bernilai "Rendah" sudah mengklasifikasikan kasus menjadi 1 keputusan yaitu "Tidak Layak" donor darah, sedangkan untuk Hb "Tinggi" hanya satu data yang menunjukkan Tidak Layak terpengaruhi oleh berat badan yang kurus dan 21 Layak sehingga dengan mudah

diklasifikasi bahwa Hb Tinggi masih dikategorikan layak.

Sedangkan untuk Hb dengan nilai "Normal" masih memiliki 2 keputusan yaitu "Layak" dan "Tidak Layak" sehingga perlu dilakukan perhitungan lagi. Ulangi lagi proses perhitungan Entropy untuk seluruh atribut dengan Hb "Normal".

Tabel 6. Perhitungan Node 1.1

Node	Atribut	Nilai	Sum	Layak	Tidak Layak	Entropy	Gain
1.1	HB		2209	2009	200	0,4382654	
		Usia					
		Remaja	848	762	86	0,4734654	
		Dewasa	1148	1055	93	0,4057296	
		Dewasa 2	213	192	21	0,4645149	
		Tensi					0,27132134
		Rendah	104	1	103	0,0782325	
		Normal	2062	2008	54	0,1748997	
		Tinggi	43	43	0	0,0000000	
		Berat Badan					0,00141924
		Kurus	0	0	0	0,0000000	
		Sedang	602	539	63	0,4835684	
	Gemuk	1217	1106	111	0,4404888		
	Obesitas	2209	364	26	0,3533593		

Sumber: (Latif & Purnia, 2019)

**Implikasi Penelitian**

Berdasarkan hasil penelitian ini UTD PMI Kota Tasikmalaya memprediksi kelayakan calon pendonor berdasarkan data pendonor tahun 2019, menggunakan data mining Algoritma C4.5 dengan hasil yang menunjukkan bahwa akurasi untuk penentuan kelayakan calon pendonor adalah 97,69%. Diketahui bahwa algoritma C4.5 dengan 5 atribut dapat dijadikan alternatif untuk pemodelan klasifikasi data mining pada dataset prediksi kelayakan calon pendonor darah. Hal ini ditunjukkan dengan hasil akurasi yang cukup tinggi dan masuk ke dalam kategori klasifikasi yang sangat baik. Hasil penelitian ini dapat diterapkan dan dikembangkan dengan membuat program atau aplikasi untuk mengimplementasikan model yang telah dibuat sehingga di masa yang akan datang seleksi calon pendonor darah bisa dilakukan dengan efektif dan efisien.

**KESIMPULAN**

Berdasarkan hasil penelitian tentang prediksi faktor kelayakan donor darah calon pendonor di UTD PMI Kota Tasikmalaya, didapatkan 5 atribut yang digunakan yaitu umur, berat badan, *hemoglobin*, tekanan darah, dan status sebagai label. Data didapatkan sejumlah 2.456 *record* dari data pendonor pada tahun 2019. Algoritma C4.5 bekerja dengan baik dengan tingkat akurasi sebesar 97,69%. *Entropy* total yang dihasilkan sebesar 0,6609359 dan *Hemoglobin*

merupakan atribut yang paling berpengaruh dalam penentuan kelayakan calon pendonor darah karena Hb atau *hemoglobin* merupakan *gain* dengan nilai tertinggi, yaitu 0,25027259.

**REFERENSI**

Atmaja, K. J., Anandita, I. B. G., & Dewi, N. K. C. (2017). Penerapan Data Mining Untuk Memprediksi Potensi Pendonor Darah Menjadi Pendonor Tetap Menggunakan Metode Decision Tree C. 45. *S@ CIES*, 7(2), 101-108. <https://jurnal.stiki-indonesia.ac.id/index.php/sacies/article/view/284>

Bayususetyo, D., Santoso, R., & Tarno, T. (2017). Klasifikasi Calon Pendonor Darah Menggunakan Metode Naïve Bayes Classifier (Studi Kasus: Calon Pendonor Darah Di Kota Semarang). *Jurnal Gaussian*, 6(2), 193-200. <https://ejournal3.undip.ac.id/index.php/gaussian/article/view/16948>

Devi, A. K. (2018). Peramalan Kebutuhan Darah Jenis Packet Red Cells (PRC) di PMI Kota Surabaya dengan Metode Jaringan Syaraf Tiruan Propagasi Balik. *Zeta-Math Journal*, 4(1), 7-11. <http://journal.uim.ac.id/index.php/Zeta/article/view/21>

- Latif, A., & Purnia, D. S. (2019). *Laporan Akhir Penelitian Mandiri: Implementasi Data Mining Untuk Mengetahui Faktor Kelayakan Donor Darah Utd Kota Tasikmalaya Menggunakan Algoritma C4.5*.
- Paraswati, A. (2014). *MASYARAKAT SERBU BUS DONOR DARAH DI CAR FREE DAY*. [Http://www.Pmi.or.Id/](http://www.pmi.or.id/).  
<http://www.pmi.or.id/masyarakat-serbu-bus-donor-darah-di-car-free-day/>
- UTD PMI Kab Tasikmalaya. (2019). *Sistem Informasi Donor Darah (SIDORDAR)*.
- Wardati, W., Nur'aini, N., & Hadi, A. J. (2019). Faktor Yang Memengaruhi Perilaku Donor Darah di Unit Transfusi Darah RS Dr. Fauziah Bireuen. *MPPKI (Media Publikasi Promosi Kesehatan Indonesia): The Indonesian Journal of Health Promotion*, 2(3), 181-185. <https://jurnal.unismuhpalu.ac.id/index.php/MPPKI/article/view/804>