

# Komparasi Algoritma C4.5 dan *Naïve Bayes* dalam Penentuan Status Kelayakan Donor Darah

<sup>1</sup>Kartika Handayani\*, <sup>2</sup>Lisnawanty, <sup>3</sup>Abdul Latif, <sup>4</sup>Muhammad Rifqi Firdaus, <sup>5</sup>Fuad Nur Hasan

<sup>1,3</sup>Teknologi Informasi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

<sup>2</sup>Sistem Informasi Akuntansi, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

<sup>4</sup>Teknologi Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

<sup>5</sup>Ilmu Komputer, Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

Jl. Kramat Raya No.98, Jakarta Pusat

Email: [kartika.kth@bsi.ac.id](mailto:kartika.kth@bsi.ac.id)

(received: 8 Juni 2021, revised: 25 Juli 2021, accepted: 5 September 2021)

## Abstrak

Donor darah merupakan kegiatan kemanusiaan dimana seseorang dengan sukarela menyumbangkan darahnya untuk disimpan di bank darah yang kemudian digunakan untuk transfusi darah. UDD (Unit Donor Darah) PMI Kota Pontianak merupakan tempat pelayanan donor darah dari masyarakat Kota Pontianak. Dalam prakteknya, tidak semua masyarakat yang ingin mendonorkan darah dapat berhasil mendonorkan darahnya. Dalam memprediksi layak atau tidaknya masyarakat untuk mendonorkan darahnya dapat dilakukan dengan klasifikasi *data mining* untuk mengetahui faktor yang paling mempengaruhi prediksi donor darah. Penelitian ini menggunakan metode klasifikasi algoritma C4.5 dan *Naïve Bayes* kemudian dilakukan perbandingan dua metode tersebut menggunakan *confusion matrix*, AUC dan uji beda t-test dengan analisa *software rapidminer* berdasarkan umur, jenis kelamin, berat badan, tekanan darah, dan hemoglobin. Dari hasil penelitian ini, hemoglobin adalah variabel paling menentukan kelayakan donor darah kemudian tekanan darah. Algoritma terbaik dalam kasus ini adalah *Naïve Bayes* dengan akurasi 93,26%, sedangkan tingkat akurasi C4.5 93,22%. *Naïve Bayes* termasuk dalam predikat *good classification* dengan AUC sebesar 0.833, sedangkan C4.5 termasuk dalam predikat *fair classification* dengan AUC sebesar 0.758. Dari hasil uji beda t-test diperoleh hasil 0.841 yang menyatakan bahwa tidak ada perbedaan signifikan dalam penentuan klasifikasi status kelayakan donor darah untuk kedua algoritma.

**Kata kunci:** prediksi, donor darah, c4.5, *naïve bayes*

## Abstract

*Blood donation is a humanitarian activity in which someone voluntarily donates blood to be stored in a blood bank which is then used for blood transfusions. UDD (Blood Donation Unit) PMI Pontianak City is a blood donor service area of the Pontianak City community. In practice, not all people who want to donate blood can successfully donate blood. In predicting the feasibility of whether or not the community to donate blood can be done with the classification of data mining to determine the factors that most influence the prediction of blood donors. This study uses the C4.5 algorithm and Naïve Bayes classification method, then compares the two methods using a confusion matrix, AUC and t-test different test with rapidminer software analysis based on age, sex, weight, blood pressure, and hemoglobin. From the results of this study, hemoglobin is the most determining variable of eligibility for blood donation then blood pressure. The best algorithm in this case is Naïve Bayes with an accuracy of 93.26%, while the accuracy rate of C4.5 is 93.22%. Naïve Bayes is included in the category of good class certification with AUC of 0.833, while C4.5 is included in the category of fair class certification with AUC of 0.758. From the results of the t-test different test results obtained 0.841 which states that there is no significant difference in determining the classification of blood donor eligibility status for the two algorithms.*

**Keywords:** prediction, blood donor, c4.5, *naïve bayes*

## 1 Pendahuluan

Standar World Health Organization (WHO) untuk ketersediaan jumlah kantong darah pada suatu negara adalah 2% dari populasi nasional. Artinya, Indonesia membutuhkan kantong darah sekitar 4,5 juta dalam satu tahun [1]. Donor darah merupakan aktivitas menyalurkan atau unsur-unsur darah untuk disimpan di bank darah yang bisa digunakan bagi pasien yang membutuhkan. Manfaat donor darah bagi kesehatan yang kurang diketahui adalah sel-sel darah di dalam tubuh menjadi lebih cepat terganti dengan yang baru. Dengan meningkatnya permintaan suplai darah di masyarakat, persediaan darah yang mencukupi sangat dibutuhkan. Meskipun demikian, pendonor harus terlebih dahulu menjalani pemeriksaan kesehatan, baik pengukuran tekanan darah, golongan darah, kadar hemoglobin (Hb) maupun konsultasi medis [2]. Palang Merah Indonesia (PMI) adalah organisasi kepalangmeraan yang ikut bertanggung jawab menyediakan kebutuhan stok darah di setiap daerah. Melalui Unit Donor Darah (UDD), PMI melakukan upaya memenuhi ketersediaan darah untuk kebutuhan pelayanan kesehatan [3].

UDD (Unit Donor Darah) PMI Kota Pontianak merupakan tempat pelayanan donor darah dari masyarakat Kota Pontianak yang ingin mendonorkan darahnya secara sukarela maupun untuk transfusi keluarga atau kerabat yang sedang membutuhkan. Dalam prakteknya, tidak semua masyarakat yang ingin mendonorkan darah dapat berhasil mendonorkan darahnya. Terdapat klasifikasi layak atau tidak layak dalam menentukan keberhasilan proses donor darah. Kriteria pendonor darah yang ditentukan PMI diantaranya sehat Jasmani dan Rohani, rentang usia dari 17 – 60 tahun dan sampai 65 tahun untuk pendonor yang rutin mendonorkan darahnya, berat badan minimal 45 Kg, tekanan darah (tensi darah) normal dengan kadar Hb (Hemoglobin) minimal 12,5 g. Orang – orang yang mempunyai penyakit jantung dan paru, kanker, tekanan darah tinggi (hipertensi), kencing manis (diabetes melitus), kelainan darah, epilepsi atau sering kejang, pernah menderita hepatitis B atau C, mengidap sifilis, ketergantungan narkoba, kecanduan minuman beralkohol, mengidap HIV/AIDS tidak disarankan dokter melakukan donor darah [4][5].

Teknik *Machine Learning* dapat meningkatkan pendeteksian otomatis. Dengan sistem ini, kemungkinan kesalahan dalam diagnosis dihindari [6]. Klasifikasi adalah sebuah teknik untuk menentukan keanggotaan kelompok berdasarkan data-data yang sudah ada. Konsep dasar dari klasifikasi adalah beberapa data yang memiliki struktur data yang mirip akan memiliki klasifikasi yang mirip pula [7]. Algoritma C4.5 dan *Naïve Bayes* merupakan dua algoritma populer yang digunakan dalam memprediksi data berdasarkan klasifikasi [8]. 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 [9]. Sedangkan, *Naïve Bayes* merupakan salah satu metoda pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistik. *Naïve Bayes* dapat memprediksi probabilitas layak atau tidaknya seorang calon pendonor berdasarkan data pendonor sebelumnya [10][11][11].

Tujuan penelitian ini adalah untuk menentukan calon pendonor pada UDD PMI Kota Pontianak termasuk layak atau tidak layak melakukan donor darah menggunakan algoritma *Naïve Bayes* dan C.45 dan membandingkan kedua algoritma untuk mendapatkan tingkat akurasi yang paling tinggi. Parameter yang digunakan dalam penelitian ini diantaranya, usia, tensi darah, jumlah hb (hemoglobin) dan berat badan. Pengujian yang dilakukan dengan menggunakan *confusion matrix* yang menguji validasi atau kebenaran antara data dengan hasil prediksi.

## 2 Tinjauan Literatur

Penelitian terkait penentuan kelayakan donor darah pernah dilakukan oleh Dhimas Bayususetyo, Rukun Santoso, dan Tarno dengan menggunakan algoritma *Naïve Bayes* [12] Membahas mengenai prediksi dan klasifikasi seseorang apakah bisa mendonorkan darahnya atau tidak, berdasarkan kadar hemoglobin, tensi atas, tensi bawah, berat badan, dan usia yang dimilikinya sebagai variabel pendukung. Data yang digunakan dalam penelitian ini merupakan data primer, yaitu data calon pendonor darah di Kota Semarang terhitung dari tanggal 16 – 27 September 2016. Data tersebut diperoleh dari mobile unit UDD PMI Kota Semarang yang mengadakan kegiatan donor darah di Bank Danamon Kota Semarang, SMA 2 Semarang, R.S Telogorejo, dan Kampus Polines. Variabel yang

<http://sistemasi.ftik.unisi.ac.id>

digunakan dalam penelitian ini terdiri dari status calon pendonor sebagai variabel dependen dan berat badan, usia, tensi atas/sistole, tensi bawah/diastole, dan kadar hemoglobin sebagai variabel independen. Setelah dilakukan proses klasifikasi terhadap data uji menggunakan dua pendekatan yaitu pendekatan fungsi densitas normal dan pendekatan selisih peluang kumulatif dengan proporsi 80% data latih dan 20% data uji diperoleh nilai *Matthews Correlation Coefficient* (MCC). MCC untuk pendekatan fungsi densitas yaitu sebesar 0,8985841 atau mendekati +1, yang berarti klasifikasi berjalan baik.

Selanjutnya Naïve Bayes Berbasis Genetic Algorithm juga digunakan dalam penentuan klasifikasi donor darah yang ditulis oleh Hilda Amalia [13]. Dataset yang digunakan dalam penelitian ini diperoleh dari situs penyedia dataset diperuntukan untuk keperluan penelitian yaitu pada <https://archive.ics.uci.edu/ml/datasets/Blood+Transfusion+Service+Center> yang merupakan data donor darah yang terdiri dari enam atribut yaitu frekuensi(bulan) sejak terkahir kali mendonorkan, frekuensi(total jumlah mendonorkan darah), total volume darah yang didonorkan, waktu(bulan) sejak pertama kali mendonorkan darahnya dan satu class label yaitu apakah akan mendonorkan darahnya pada maret 2007. Total data yang terdapat pada dataset tranfusi darah adalah 749 record. Dari hasil pengolahan dataset donor darah menggunakan naïve bayes diperoleh nilai akurasi yaitu sebesar 74,08% dan nilai AUC sebesar 0,709. Sedangkan dari hasil pengolahan nilai akurasi dan nilai AUC untuk peningkatan optimasi genetic algorithm pada metode naïve bayes, diperoleh nilai akurasi sebesar 76,48% dan nilai AUC 0,623.

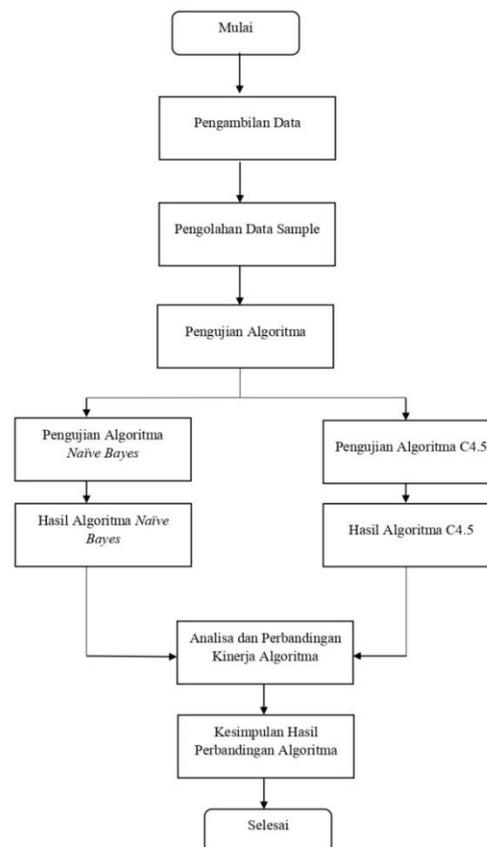
Prediksi kelayakan donor darah juga pernah diteliti menggunakan algoritma C4.5 oleh Anita Febriani, Tiara Trimadya Rahmawati dan Eka Sabna [14]. Penelitian berdasarkan: usia, berat badan, kadar hemoglobin (HB), tekanan darah (sistolik dan diastolik) yang dimilikinya sebagai variabel pendukung. Data usia, berat badan, kadar Hemoglobin, dan tekanan darah, hasilnya diambil dari beberapa literatur tentang usia, berat badan, kadar Hemoglobin, dan tekanan darah. Data yang digunakan adalah data calon pendonor darah dari UTD PMI Kota Pekanbaru sebanyak 1839 data. Berdasarkan hasil *confision matrix* didapat akurasi dari pohon keputusan yang dihasilkan dari C4.5 untuk penentuan kelayakan calon pendonor darah adalah 94,02%.

Penelitian terkait donor darah menggunakan C4.5 juga dilakukan yang dilakukan oleh Ketut Jaya Atmaja, Ida Bagus Gede Anandita, dan Ni Kadek Ceryna [15]. Membahas tentang menentukan pola prediksi pendonor yang berpotensi menjadi pendonor tetap. Sehingga dapat difokuskan dimana penyebaran informasi harus dilakukan agar lebih efisien dalam menambah jumlah pendonor darah. Metode yang dipakai adalah metode klasifikasi menggunakan algoritma C4.5. Adapun atribut data yang dipakai adalah jenis kelamin, pekerjaan, golongan darah, jumlah donor, dan umur. Banyak data yang dipakai adalah 600 data, dimana *data training* berjumlah 500 data, dan data set berjumlah 100 data. Klasifikasi dilakukan berdasarkan pekerjaan. Hasil yang didapatkan dari penelitian ini bahwa pegawai swasta dengan umur diatas 26 tahun paling banyak menjadi pendonor.

Berdasarkan peneltian terkait, dapat disimpulkan bahwa penggunaan algoritma C4.5 dan Naïve Bayes pada kasus klasifikasi status kelayakan donor darah menghasilkan kinerja yang cukup baik. Sehingga pada penelitian yang diusulkan dilakukan perbandingan antara algoritma C4.5 dan Naïve Bayes dengan data Donor Darah UDD Kota Pontianak untuk mendapatkan tingkat akurasi prediksi yang paling tinggi dengan parameter yang digunakan yaitu usia, tensi darah, jumlah hb (hemoglobin) dan berat badan. Penelian ini menggunakan *software rapid miner* untuk pengolahan data dan perhitungan manual menggunakan *microsoft excel* untuk menguji kesesuaian hasil.

### 3 Metode Penelitian

Metode penelitian yang dilakukan untuk menganalisis permasalahan diatas menggunakan teknik *Machine Learning* dengan algoritma *Naïve Bayes* dan C4.5 seperti gambar 1:



**Gambar 1. Alur Diagram**

Berdasarkan gambar 1, tahapan penelitian yang dilakukan oleh penulis dalam melakukan penelitian ini seperti kegiatan di bawah ini :

### **3.1. Pengumpulan Data**

Pengumpulan data dilakukan secara langsung dari *database* Sistem Informasi Donor Darah UDD PMI Kota Pontianak. Data berupa histori pendonor Januari 2019 – Oktober 2019 dengan jumlah data sebanyak 26.714 *record* dengan parameter Nama, Umur, Jenis Kelamin, HB, Tensi, Berat Badan dan Status.

### **3.2. Pengolahan Data Sample**

Pengolahan data sample dilakukan dengan proses *Data Cleaning*, *Data Integration and Transformation* dan *Data Reduction*

### **3.3. Pengujian Algoritma**

Pengujian Algoritma dilakukan dengan perhitungan manual menggunakan *Microsoft Excel* dan *Rapid Miner* pada kedua algoritma menggunakan data yang sudah diolah pada tahap sebelumnya.

### **3.4. Analisa dan Kesimpulan Perbandingan Algoritma**

Evaluasi digunakan untuk melakukan pengamatan dan menganalisa hasil kerja kedua algoritma pada *Rapid Miner*. Validasi dilakukan untuk melakukan pengukuran hasil prediksi.

*Cross Validation*, merupakan pengujian standar yang dilakukan untuk memprediksi *error rate* [6]. Pengukuran hasil validasi dengan menggunakan ROC dan *Confusion matrix* sampai mendapatkan tingkat akurasi yang tertinggi. Sebuah metode umum untuk menghitung daerah dibawah kurva ROC adalah *Area Under Curve (AUC)* dimana bidang yang berada dibawah kurve mempunyai nilai yang selalu berada pada nilai 0,0 dan 1,0. Semakin tinggi luasnya maka akan semakin baik nilai klasifikasinya. *Performance* keakurasian AUC (Gorunescu, 2010) dapat diklasifikasikan menjadi lima kelompok yaitu:

1. 0,90 – 1,00 = *Exellent Classification*
2. 0,80 – 0,90 = *Good Classification*
3. 0,70 – 0,80 = *Fair Classification*
4. 0,60 – 0,70 = *Poor Classification*
5. 0,50 – 0,60 = *Failure Classification*

*Confusion matrix* adalah sebuah tabel yang menyatakan jumlah data uji yang benar diklasifikasikan dan jumlah data uji yang salah diklasifikasikan untuk melihat nilai akurasi yang dihasilkan.

## 4 Hasil dan Pembahasan

Hasil penelitian ini membahas mengenai data yang telah dikumpulkan selama kegiatan penelitian, dan mengimplementasikan data hasil penelitian ke dalam metode prediksi yaitu metode *data mining* menggunakan perbandingan Algoritma *Naive Bayes* dan *C4.5*. Semua yang berkaitan dengan pengujian akan dibahas pada bab ini mulai dari pengolahan dataset hingga pengujian data pada *software*.

### 4.1. Pengumpulan Data

Pengambilan data dilakukan dengan Aplikasi Sistem Informasi Donor Darah UDD PMI Pontianak dengan cara *export* data dari *database* kedalam bentuk *excel*. Pengambilan data dilakukan oleh petugas kemudian data dalam bentuk *excel* merupakan data mentah dan masih harus diolah kembali untuk menghasilkan informasi. Dalam pengambilan data, tidak semua atribut data di *export* oleh petugas. Hal ini guna menjaga kerahasiaan data pribadi pendonor. Data yang diambil merupakan data pendonor sejak Januari 2019 – Oktober 2019 dengan jumlah data sebanyak 26.714 *record* sample. Berikut contoh data *primer* yang diperoleh pada tabel 1.

Tabel 1. Contoh Data Donor Darah

| No | Nama | Umur | Gol  |        | DS Donor |     |          | HB   | Tensi  | Berat Badan | DG MU |
|----|------|------|------|--------|----------|-----|----------|------|--------|-------------|-------|
|    |      |      | (Rh) | JK     | DP       | Ke- | Status   |      |        |             |       |
| 1  |      | 44   | O+   | Pria   | DS       | 13  | Berhasil | 12,5 | 110/80 | 50          | DG    |
| 2  |      | 29   | AB+  | Pria   | DS       | 11  | Berhasil | 12,5 | 110/90 | 50          | DG    |
| 3  |      | 48   | AB+  | Pria   | DP       | 2   | Berhasil | 12,5 | 110/80 | 50          | DG    |
| 4  |      | 33   | O+   | Pria   | DS       | 10  | Berhasil | 12,5 | 110/80 | 50          | DG    |
| 5  |      | 46   | B+   | Pria   | DS       | 17  | Berhasil | 12,5 | 110/80 | 50          | DG    |
| 6  |      | 24   | A+   | Pria   | DS       | 3   | Berhasil | 12,5 | 110/80 | 50          | DG    |
| 7  |      | 54   | O+   | Pria   | DS       | 94  | Berhasil | 12,5 | 120/80 | 50          | DG    |
| 8  |      | 37   | O+   | Pria   | DS       | 34  | Berhasil | 12,5 | 110/90 | 50          | DG    |
| 9  |      | 48   | B+   | Pria   | DP       | 10  | Berhasil | 12,5 | 110/80 | 50          | DG    |
| 10 |      | 35   | B+   | Pria   | DP       | 1   | Berhasil | 12,5 | 120/80 | 50          | DG    |
| 11 |      | 57   | A+   | Pria   | DS       | 75  | Berhasil | 12,5 | 130/90 | 50          | DG    |
| 12 |      | 37   | A+   | Pria   | DS       | 24  | Berhasil | 12,5 | 120/80 | 50          | DG    |
| 13 |      | 25   | B+   | Pria   | DS       | 3   | Berhasil | 12,5 | 120/80 | 50          | DG    |
| 14 |      | 38   | A+   | Pria   | DS       | 34  | Berhasil | 12,5 | 120/80 | 50          | DG    |
| 15 |      | 31   | A+   | Wanita | DS       | 10  | Berhasil | 12,5 | 120/80 | 50          | DG    |

### 4.2. Pengolahan Data Sample

Data sampel ini adalah data yang akan digunakan dalam pengujian yang didapatkan dari sistem informasi donor darah, kemudian dibuat dalam format *excel* yang baru untuk dirapikan.

Beberapa tahapan yang dilakukan penulis untuk mengolah data agar *dataset* menjadi data informasi yang nantinya bisa digunakan untuk peujian.

#### 1. Data Cleaning

Seperti yang sudah dibahas pada bab sebelumnya *data cleaning* digunakan untuk menghasilkan data berkualitas. Untuk mendapatkan data yang berkualitas penulis melakukan tahapan dalam *data cleaning* sebagai berikut :

- a. Menghilangkan data yang tidak lengkap. Terdapat 83 data dari 26.714 yang tergolong tidak lengkap, data yang tidak lengkap ini seperti tensi, hb dan berat badan yang tidak diisi/kosong.
  - b. Menghilangkan data yang *noise*. Terdapat 25 data dari 26.714 yang tergolong *noise*. Sehingga tersisa 26.606 *record*.
2. *Data Reduction*  
*Dataset* direduksi dengan mengurangi jumlah atribut dan *record* sehingga data yang awalnya terdapat 10 atribut direduksi menjadi 5 atribut dengan atribut status sebagai *label/class*. Setelah mengalami reduksi dan *data cleaning*, data yang dihasilkan lebih informatif dengan menghilangkan beberapa atribut yang tidak digunakan serta menghapus data yang tidak lengkap.
3. *Data Transformation*  
*Data transformation* atau proses perubahan data dalam kategori tertentu yang berfungsi mengubah data yang kompleks dengan tidak menghilangkan isi, sehingga lebih mudah diolah. Berikut contoh *data transformation* yang dilakukan pada tabel 2.

**Tabel 2. Hasil Data Transformation**

| No | Jenis Kelamin | Umur      | HB     | Tensi  | Berat  | Status      |
|----|---------------|-----------|--------|--------|--------|-------------|
| 1  | Pria          | Remaja    | Normal | Normal | Sedang | Layak       |
| 2  | Pria          | Dewasa    | Normal | Normal | Sedang | Layak       |
| 3  | Pria          | Remaja    | Normal | Normal | Sedang | Layak       |
| 4  | Pria          | Remaja    | Normal | Normal | Sedang | Layak       |
| 5  | Pria          | Dewasa    | Normal | Normal | Sedang | Layak       |
| 6  | Pria          | Anak-Anak | Normal | Normal | Sedang | Layak       |
| 7  | Pria          | Remaja    | Normal | Normal | Sedang | Layak       |
| 8  | Pria          | Remaja    | Normal | Normal | Sedang | Layak       |
| 9  | Pria          | Dewasa    | Normal | Normal | Sedang | Layak       |
| 10 | Wanita        | Dewasa    | Normal | Normal | Sedang | Layak       |
| 11 | Pria          | Remaja    | Normal | Normal | Sedang | Layak       |
| 12 | Pria          | Dewasa    | Normal | Normal | Sedang | Layak       |
| 13 | Wanita        | Dewasa    | Normal | Normal | Sedang | Layak       |
| 14 | Pria          | Dewasa    | Rendah | Normal | Sedang | Tidak Layak |
| 15 | Pria          | Dewasa 2  | Rendah | Normal | Sedang | Tidak Layak |
| 16 | Pria          | Dewasa    | Rendah | Normal | Sedang | Tidak Layak |
| 17 | Wanita        | Dewasa    | Rendah | Normal | Sedang | Tidak Layak |
| 18 | Pria          | Dewasa    | Rendah | Normal | Sedang | Tidak Layak |
| 19 | Pria          | Dewasa 2  | Tinggi | Normal | Sedang | Tidak Layak |

### 4.3. Pengujian Algoritma

Dalam pengujian ini, akan dilakukan pengujian perhitungan berdasarkan rumus dan ketentuan algoritma *Naïve bayes* dan C4.5. Berikut langkah pengujian yang dilakukan:

#### 4.3.1. Pengujian Algoritma Naïve Bayes

*Naïve bayes* adalah algoritma pertama yang akan dihitung. Langkah-langkah yang akan dilakukan adalah menghitung nilai probabilitas prior, yaitu probabilitas nilai data pendonor darah yang “Layak” dan “Tidak Layak”, berikut perhitungannya pada tabel 3.

**Tabel 3. Perhitungan Algoritma Naïve Bayes**

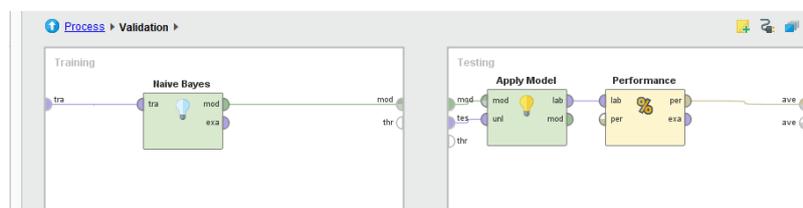
| Atribut       | Kasus | Layak | Tidak Layak | P(X Ci) |             |
|---------------|-------|-------|-------------|---------|-------------|
|               |       |       |             | Layak   | Tidak Layak |
| Total Kasus   | 26606 | 22910 | 3696        | 0.861   | 0.139       |
| Jenis Kelamin |       |       |             |         |             |
| Pria          | 20102 | 18096 | 2006        | 0.790   | 0.543       |
| Wanita        | 6504  | 4814  | 1690        | 0.210   | 0.457       |
| Umur          |       |       |             |         |             |
| Remaja        | 8538  | 7044  | 1494        | 0.307   | 0.404       |
| Dewasa        | 7387  | 6495  | 892         | 0.284   | 0.241       |
| Dewasa 2      | 6596  | 5793  | 803         | 0.253   | 0.217       |
| Lansia        | 4085  | 3578  | 507         | 0.156   | 0.137       |
| HB            |       |       |             |         |             |
| Normal        | 25476 | 22904 | 2572        | 1.000   | 0.696       |
| Tinggi        | 287   | 0     | 287         | 0.000   | 0.078       |
| Rendah        | 843   | 6     | 837         | 0.000   | 0.226       |
| Tensi         |       |       |             |         |             |
| Normal        | 25750 | 22901 | 2849        | 1.000   | 0.771       |
| Tinggi        | 272   | 2     | 270         | 0.000   | 0.073       |
| Rendah        | 584   | 7     | 577         | 0.000   | 0.156       |
| Berat         |       |       |             |         |             |
| Kurus         | 12    | 0     | 12          | 0.000   | 0.003       |
| Sedang        | 22822 | 19353 | 3469        | 0.845   | 0.939       |
| Gemuk         | 2620  | 2459  | 161         | 0.107   | 0.044       |
| Gemuk 2       | 1152  | 1098  | 54          | 0.048   | 0.015       |

Pada tabel 4 diperoleh hasil perhitungan nilai probabilitas “Layak” atau “Tidak Layak” dari masing masing atribut.

**Tabel 4. Perhitungan Nilai Probabilitas Prior Data Testing**

| Data X Untuk Kasus Baru |        | P(X Ci) |             |
|-------------------------|--------|---------|-------------|
| Atribut                 | Nilai  | Layak   | Tidak Layak |
| Jenis Kelamin           | Pria   | 0.790   | 0.543       |
| Usia                    | Dewasa | 0.284   | 0.241       |
| Hemoglobin              | Normal | 1.000   | 0.696       |
| Tensi                   | Rendah | 0.000   | 0.156       |
| Berat                   | Sedang | 0.845   | 0.939       |

Setelah dilakukan pengujian perhitungan menggunakan *Microsoft Excel*, dilakukan pula pengujian menggunakan Aplikasi RapidMiner, berikut pengujian menggunakan RapidMiner pada gambar 2.



**Gambar 2. Subproses Training dan Testing Algoritma Naïve bayes**

accuracy: 93.26% +/- 0.49% (mikroc: 93.26%)

|                   | true Layak | true Tidak Layak | class precision |
|-------------------|------------|------------------|-----------------|
| pred. Layak       | 22895      | 1779             | 92.79%          |
| pred. Tidak Layak | 15         | 1917             | 99.22%          |
| class recall      | 99.93%     | 51.87%           |                 |

**Gambar 3. Confusion Matrix Algoritma Naïve bayes**

Berdasarkan gambar 3, dapat dijelaskan bahwa dari 26606 data sebanyak 22895 data diprediksikan sesuai yaitu 22895 data “Layak” dan 1779 data yang diprediksikan “Layak” tetapi ternyata “Tidak Layak”. Dan sebanyak 15 data diprediksi “Tidak Layak” ternyata termasuk klasifikasi “Layak” dan sebanyak 1917 data diprediksi sesuai yaitu “Tidak Layak”.

#### 4.3.2. Pengujian Algoritma C4.5

C4.5 adalah algoritma kedua yang akan dihitung, dimulai dengan perhitungan nilai *entropy* dan *gain* untuk penentuan node 1 kemudian dilanjutkan dengan perhitungan *entropy* dan *gain* untuk penentuan node 1.1 berikut hasil perhitungan pada tabel 5 dan tabel 6.

**Tabel 5. Perhitungan Nilai Entropy dan Gain untuk Penentuan Node 1**

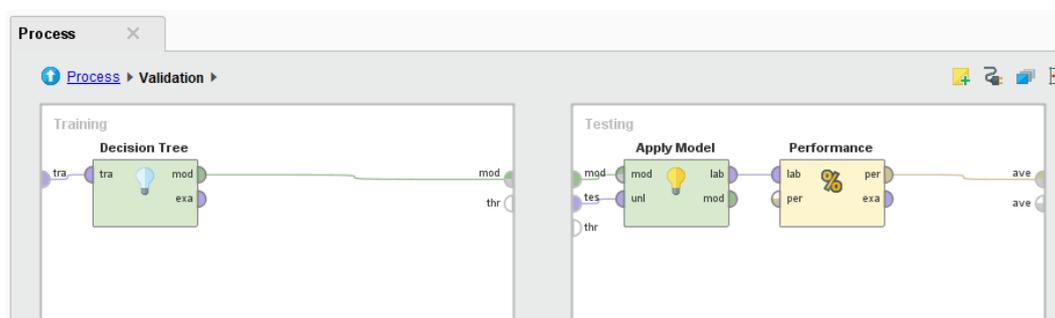
| Node | Atribut       | Kasus | Layak | Tidak Layak | Entropy  | Gain             |
|------|---------------|-------|-------|-------------|----------|------------------|
| 1    | Total Kasus   | 26606 | 22910 | 3696        | 0.581393 |                  |
|      | Jenis Kelamin |       |       |             |          | 0.0255027        |
|      | Pria          | 20102 | 18096 | 2006        | 0.468333 |                  |
|      | Wanita        | 6504  | 4814  | 1690        | 0.826505 |                  |
|      | Umur          |       |       |             |          | 0.0035964        |
|      | Remaja        | 8538  | 7044  | 1494        | 0.668977 |                  |
|      | Dewasa        | 7387  | 6495  | 892         | 0.531521 |                  |
|      | Dewasa 2      | 6596  | 5793  | 803         | 0.534343 |                  |
|      | Lansia        | 4085  | 3578  | 507         | 0.541068 |                  |
|      | HB            |       |       |             |          | <b>0.1274832</b> |
|      | Normal        | 25476 | 22904 | 2572        | 0.472024 |                  |
|      | Tinggi        | 287   | 0     | 287         | 0        |                  |
|      | Rendah        | 843   | 6     | 837         | 0.06101  |                  |
|      | Tensi         |       |       |             |          | 0.0927475        |
|      | Normal        | 25749 | 22900 | 2849        | 0.501859 |                  |
|      | Tinggi        | 273   | 3     | 270         | 0.087281 |                  |
|      | Rendah        | 584   | 7     | 577         | 0.093691 |                  |
|      | Berat         |       |       |             |          | 0.0093789        |
|      | Kurus         | 12    | 0     | 12          | 0        |                  |
|      | Sedang        | 22822 | 19353 | 3469        | 0.614829 |                  |
|      | Gemuk         | 2620  | 2459  | 161         | 0.333176 |                  |
|      | Gemuk 2       | 1152  | 1098  | 54          | 0.272971 |                  |

Dari tabel perhitungan *entropy* dan *gain* diatas, diketahui nilai *gain* tertinggi pada atribut hemoglobin yakni 0.1274832. Maka *node* 1 untuk pohon keputusan adalah *hemoglobin*. Kemudian dilakukan lagi perhitungan nilai *entropy* dan *gain* untuk menentukan *node* 1.1 seperti tabel 6 berikut:

**Tabel 6. Perhitungan Nilai Entropy dan Gain untuk Penentuan Node 1.1**

| ode | Atribut       | Kasus | Layak | Tidak Layak | Entropy  | Gain             |
|-----|---------------|-------|-------|-------------|----------|------------------|
| 1.1 | HB            |       |       |             |          |                  |
|     | Normal        | 25476 | 22904 | 2572        | 0.472024 |                  |
|     | Jenis Kelamin |       |       |             |          | 0.0130896        |
|     | Pria          | 19609 | 18091 | 1518        | 0.392999 |                  |
|     | Wanita        | 5867  | 4813  | 1054        | 0.679308 |                  |
|     | Umur          |       |       |             |          | 0.0020205        |
|     | Remaja        | 8047  | 7043  | 1004        | 0.542910 |                  |
|     | Dewasa        | 7132  | 6493  | 639         | 0.435121 |                  |
|     | Dewasa 2      | 6359  | 5792  | 567         | 0.433676 |                  |
|     | Lansia        | 3938  | 3576  | 362         | 0.442862 |                  |
|     | Tensi         |       |       |             |          | <b>0.1055548</b> |
|     | Normal        | 24683 | 22894 | 1789        | 0.375108 |                  |
|     | Tinggi        | 262   | 3     | 259         | 0.090262 |                  |
|     | Rendah        | 531   | 7     | 524         | 0.101221 |                  |
|     | Berat         |       |       |             |          | 0.0158569        |
|     | Kurus         | 11    | 0     | 11          | 0        |                  |
|     | Sedang        | 21802 | 19350 | 2452        | 0.507312 |                  |
|     | Gemuk         | 2538  | 2456  | 82          | 0.205842 |                  |
|     | Gemuk 2       | 1125  | 1098  | 0           | 0.034206 |                  |

Dari tabel perhitungan *entropy* dan *gain* diatas, diketahui nilai *gain* tertinggi untuk perhitungan *node* 1.1 pada atribut Tensi yakni 0.1055548. Maka *node* 1.1 untuk pohon keputusan adalah Tensi. Setelah dilakukan pengujian perhitungan menggunakan *Microsoft Excel*, dilakukan pula pengujian menggunakan Aplikasi RapidMiner, berikut pengujian menggunakan RapidMiner pada gambar 4.



**Gambar 4. Subproses Training dan Testing Algoritma C4.5**

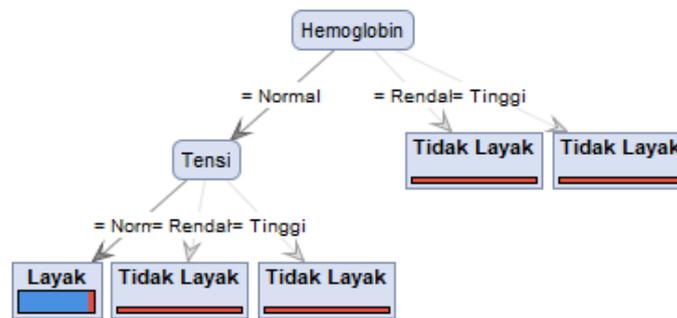
accuracy: 93.22% +/- 0.48% (mikro: 93.22%)

|                   | true Layak | true Tidak Layak | class precision |
|-------------------|------------|------------------|-----------------|
| pred. Layak       | 22895      | 1789             | 92.75%          |
| pred. Tidak Layak | 15         | 1907             | 99.22%          |
| class recall      | 99.93%     | 51.60%           |                 |

**Gambar 5. Confusion Matrix Algoritma C4.5**

Berdasarkan gambar 5, dapat dijelaskan bahwa dari 26606 data sebanyak 22895 data diprediksikan sesuai yaitu 22895 data “Layak” dan 1789 data yang diprediksikan “Layak” tetapi ternyata “Tidak Layak”. Dan sebanyak 15 data diprediksi “Tidak Layak” ternyata termasuk klasifikasi “Layak” dan sebanyak 1907 data diprediksi sesuai yaitu “Tidak Layak”.

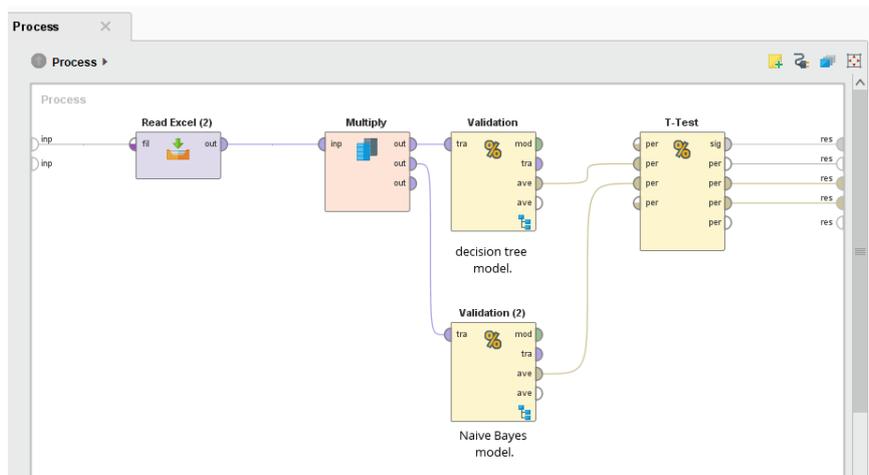
Dari proses analisa C4.5 menggunakan RapidMiner juga diketahui pohon keputusan yang terbentuk dari data donor darah yang telah masukkan. Pohon keputusan (*decision tree*) yang terbentuk sesuai dengan perhitungan manual yang telah dilakukan sebelumnya yakni dengan *node* 1 adalah *Hemoglobin* dan *node* 1.1 adalah *Tensi*. Berikut pohon keputusan yang sudah terbentuk dalam gambar 6.



**Gambar 6. Pohon Keputusan (Decision Tree)**

#### 4.4. Analisa Perbandingan Algoritma

RapidMiner juga digunakan untuk melakukan Uji Beda T-Test. T-Test dilakukan untuk menentukan dua sampel yang tidak berhubungan yang memiliki nilai rata-rata berbeda. Caranya dengan membandingkan perbedaan dari ke dua sampel tersebut. Berikut adalah model uji beda T-Test menggunakan *rapidminer* pada gambar 7 dan gambar 8.



**Gambar 7. Model Uji Beda T-Test Algoritma C4.5 dan Naïve Bayes**

| A           | B    | C           |
|-------------|------|-------------|
|             | C4.5 | Naïve Bayes |
| C4.5        | -    | 0.841       |
| Naïve Bayes | -    | -           |

**Gambar 8. Hasil Uji Beda T-Test Algoritma C4.5 dan Naïve Bayes**

Nilai alpha yang digunakan adalah 0.05. jika probabilitasnya  $>0.05$  maka  $H_0$  diterima (tidak ada perbedaan yang signifikan) sedangkan jika probabilitasnya  $<0.05$  maka  $H_0$  ditolak (ada perbedaan yang signifikan). Berdasarkan dari analisa pengujian masing-masing algoritma di atas maka dapat dibandingkan hasil kedua algoritma pada tabel berikut:

**Tabel 7. Perbandingan Kinerja Algoritma**

|                         | C4.5   | Naïve Bayes |
|-------------------------|--------|-------------|
| <i>Accuracy</i>         | 93,22% | 92,26%      |
| <i>Area Under Curve</i> | 0,758  | 0,833       |

Berdasarkan tabel 7, perbandingan berdasarkan tingkat *accuracy*, dalam penentuan klasifikasi donor darah pada UDD PMI Kota Pontianak dengan 26.606 dataset yang terdiri dari 6 atribut yaitu Umur, Jenis Kelamin, Berat Badan, Hemoglobin, Tensi dan Status, Algoritma *Naïve Bayes* lebih unggul dengan hasil *accuracy* 93,26%. Sedangkan Algoritma C4.5 menghasilkan *accuracy* 93,22%.

## 5 Kesimpulan

Berdasarkan penelitian yang telah dilakukan menggunakan 26.606 dataset donor darah UDD PMI Kota Pontianak Januari – Oktober 2019 dengan 6 atribut yaitu Umur, Jenis Kelamin, Berat Badan, *Hemoglobin*, Tensi dan Status dapat disimpulkan bahwa berdasarkan nilai *gain* tertinggi sebesar 0.1274832 pada perhitungan *node* 1 yang merupakan nilai *gain Hemoglobin*. Maka *Hemoglobin* merupakan faktor yang paling berpengaruh dalam penentuan klasifikasi status donor darah pada UDD PMI Kota Pontianak. Setelah *Hemoglobin* faktor yang paling mempengaruhi selanjutnya adalah Tensi (tekanan darah) yang memperoleh nilai *gain* sebesar 0.1055548 pada perhitungan *node* 1.1. Algoritma *Naïve Bayes* menghasilkan *accuracy* sebesar 93,26% lebih unggul dari Algoritma C4.5 menghasilkan *accuracy* sebesar 93,22%. Berdasarkan *Performance* keakurasian AUC penentuan klasifikasi status donor darah pada UDD PMI Kota Pontianak, Algoritma *Naïve Bayes* termasuk algoritma yang akurat dengan AUC sebesar 0.833 yang termasuk dalam predikat *Good Classification*, sedangkan Algoritma C4.5 dengan AUC sebesar 0,758 termasuk dalam predikat *Fair Classification* (0,70 – 0,80). Berdasarkan Uji Beda T-Test, diperoleh hasil 0.841. Hasil ini menunjukkan bahwa tidak terdapat perbedaan signifikan penerapan Algoritma C4.5 dan *Naïve Bayes* dalam penentuan status kelayakan donor darah pada UDD PMI Kota Pontianak. Sehingga  $H_0$  diterima dan  $H_1$  ditolak.

## Referensi

- [1] M. R. Firdaus, A. Latif, and W. Gata, “Klasifikasi Kelayakan Calon Pendonor Darah Menggunakan Neura L Network,” *Sistemasi*, vol. 9, no. 2, p. 362, 2020, doi: 10.32520/stmsi.v9i2.840.
- [2] D. Bayususetyo, R. Santoso, and T. Tarno, “Klasifikasi calon pendonor darah menggunakan metode naïve bayes classifier (Studi Kasus : Calon Pendonor Darah di Kota Semarang),” *J. Gaussian*, vol. 6, no. 2, pp. 193–200, 2017, [Online]. Available: <https://ejournal3.undip.ac.id/index.php/gaussian/article/view/16948>.
- [3] R. R. K. Dewi, M. I. Kartasurya, and A. Mawarni, “Analisis Kebijakan Donor Darah Dan Implementasi Program Rekrutmen Donor Di Unit Donor Darah ( UDD PMI ) Kota

- Pontianak,” *Manaj. Kesehat. Indones.*, vol. 4, no. 02, pp. 109–117, 2016, [Online]. Available: <https://ejournal.undip.ac.id/index.php/jmki/article/view/13601/10255>.
- [4] K. B. Utomo, “Perancangan Sistem Informasi Bank Darah Hidup untuk Mempercepat Penyediaan Calon Penyumbang Darah Dengan Ketepatan yang Tinggi (Studi di PMI Kota Samarinda),” *J. Inform. Mulawarman*, vol. 5, p. 22, 2010.
- [5] PMI, “Tentang Donor Darah.” .
- [6] F. S. Nugraha, M. J. Shidiq, and S. Rahayu, “Analisi Algoritma Klasifikasi Neural Network untuk Diagnosa Penyakit Kanker Payudara,” *J. PILAR Nusa Mandiri*, 2019.
- [7] Y. I. K. Kurniawan, “Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. x, no. 30, pp. 1–8, 2018, doi: 10.25126/jtiik.
- [8] S. Bahri, D. Marisa Midyanti, R. Hidayati, J. Sistem Komputer, and F. Mipa, “Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak,” *Semin. Nas. Apl. Teknol. Inf.*, pp. 24–31, 2018.
- [9] Asmira, “Penerapan Data Mining untuk Mengklasifikasi Pola Nasabah Menggunakan Algoritma C4,5 pada Bank BRI Unit Andounohu Kendari,” *J. Sist. Komput. dan Sist. Inf.*, vol. 1, no. 1, pp. 22–28, 2019, [Online]. Available: <file:///C:/Users/LUKINO~1/AppData/Local/Temp/316-828-1-SM-2.pdf>.
- [10] S. Lestari and M. Badrul, “Implementasi Klasifikasi Naive Bayes Untuk Prediksi,” vol. 7, no. 1, pp. 8–16, 2020.
- [11] P. A. Rahayuningsih, “Komparasi Algoritma Klasifikasi Data Mining Untuk Memprediksi Tingkat Kematian Dini Kanker Dengan Dataset Early Death Cancer,” *J. Tek. Inform. Kaputama*, vol. 3, no. 2, 2019, [Online]. Available: <http://jurnal.kaputama.ac.id/index.php/JTIK/article/view/169>.
- [12] D. Bayususetyo, R. Santoso, and Tarno, “Klasifikasi Calon Pendorong Darah Menggunakan Metode Naive Bayes Classifier,” *J. Gaussian*, vol. 6, 2AD.
- [13] H. Amalia, “Penerapan Naive Bayes Berbasis Genetic Algorithm Untuk Penentuan Klasifikasi Donor Darah,” *J. Tek. Komput. AMIK BSI*, vol. III, no. 1, pp. 70–76, 2017, [Online]. Available: <https://pdfs.semanticscholar.org/2157/29026e7e58b60b322267534aaa28bb448f41.pdf>.
- [14] A. Febriani, T. T. Rahmawati, and E. Sabna, “Implementation of Data Mining to Predict The Feasibility of Blood Donors Using C4.5 Algorithm,” *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, p. 41, 2018, doi: 10.24014/ijaidm.v1i1.4562.
- [15] K. J. Atmaja, I. B. G. Anandita, and N. K. C. Dewi, “Penerapan Data Mining Untuk Memprediksi Potensi Pendorong Darah Menjadi Pendorong Tetap Menggunakan Metode Decision Tree C.45,” *S@Cies*, vol. 7, no. 2, pp. 101–108, 2018, doi: 10.31598/sacies.v7i2.284.