

## Perbandingan Metode *k*-NN dan *Naïve Bayes* dalam Klasifikasi Penentuan Calon Pendoror Darah

Nahya Nur\*<sup>1</sup>, Nur Syahra<sup>2</sup>, Asmawati<sup>3</sup>

<sup>1,2,3</sup>Program Studi Teknik Informatika Universitas Sulawesi Barat

<sup>1</sup>[nahya.nur@unsulbar.ac.id](mailto:nahya.nur@unsulbar.ac.id), <sup>2</sup>[nursyahra98@gmail.com](mailto:nursyahra98@gmail.com), <sup>3</sup>[asmawati.s@unsulbar.ac.id](mailto:asmawati.s@unsulbar.ac.id)

### Abstrak

*Donor darah adalah proses pengambilan darah dari seorang yang sehat dengan dinyatakan memenuhi syarat dengan meninjau beberapa faktor seperti pengukuran tekanan darah, golongan darah, kadar hemoglobin (Hb), status pendonor, konsultasi medis dan sebagainya. Klasifikasi adalah proses menemukan fungsi atau model yang dapat membedakan suatu konsep atau kelas data untuk memprediksi kelas suatu objek yang tidak diketahui. Penelitian ini dilakukan untuk mencari metode terbaik berdasarkan perbandingan nilai akurasi menggunakan 100 data latih donor darah dari RSUD Majene yang diproses dengan mengimplementasikan metode K-Nearest Neighbor dan Naive Bayes. Hasil analisis menunjukkan bahwa algoritma K-Nearest Neighbor menunjukkan performa yang lebih baik dengan tingkat akurasi sebesar 86% sedangkan algoritma Naive Bayes memperoleh nilai akurasi sebesar 76%.*

**Kata kunci**— Donor darah, Klasifikasi Naive Bayes, K-Nearest Neighbor

### Abstract

*Blood donation is the process of taking blood from a healthy person who is declared eligible by reviewing several factors such as measurement of blood pressure, blood group, hemoglobin (Hb) level, donor status, medical consultation and so on. Classification is the process of finding a function or model that can distinguish a concept or data class to predict the unknown class of an object. This research was conducted to find the best method based on the comparison of accuracy values using 100 blood donor training data from the Majene Regional Hospital which was processed by implementing the K-Nearest Neighbor and Naive Bayes methods. The analysis results show that the K-Nearest Neighbor algorithm shows better performance with an accuracy rate of 86% while the Naive Bayes algorithm gets an accuracy value of 76%.*

**Keywords**— Blood donation, Classification, Naive Bayes, K-Nearest Neighbor

## 1. PENDAHULUAN

Donor darah adalah proses transfuse darah dari seseorang yang sehat ke seseorang yang membutuhkan dengan beberapa komponen darah. Manfaat mendonorkan darah cukup banyak untuk kesehatan, diantaranya adanya regenerasi sel-sel darah dalam tubuh menjadi lebih cepat. Meski demikian, manfaat tersebut masih belum diketahui oleh kebanyakan orang. Persediaan darah kadang menjadi kendala sehingga harus dipastikan bahwa persediaan darah harus mencukupi. Akan tetapi, tidak semua orang dapat mendonorkan darahnya. Serangkaian pemeriksaan harus dijalani

oleh pendonor, seperti pengukuran tekanan darah, golongan darah, kadar hemoglobin (Hb), status pendonor maupun konsultasi medis [1].

Permasalahan yang sering dihadapi petugas bagian pemeriksa tahap awal yaitu kendala dalam menentukan status calon pendonor apakah layak atau tidak layak untuk mendonorkan darahnya disebabkan oleh keterbatasan jumlah petugas yang terlibat. Untuk itu diperlukan sebuah sistem pengklasifikasian untuk menentukan status donor calon pendonor darah. Metode penyelesaian masalah yang digunakan dalam sistem ini memanfaatkan metode *K-NN* dan metode *naïve bayes*. Kedua metode tersebut merupakan metode pengklasifikasian yang memiliki tingkat keberhasilan yang tinggi serta kelebihan-kelebihan lain yang dimiliki oleh masing-masing metode. Pembuatan sistem ini juga diharapkan mampu memudahkan dalam proses pengambilan keputusan untuk menentukan status calon pendonor yang pada awalnya masih manual menjadi suatu sistem yang terkomputerisasi.

Beberapa penelitian telah dilakukan sebelumnya dalam membandingkan metode klasifikasi penentuan calon pendonor darah. Penelitian-penelitian dengan konsep klasifikasi yang menggunakan metode *naïve bayes* dan *k-NN* diantaranya: Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012 [2], penelitian dalam menentukan calon pendonor darah menggunakan algoritma Naïve Bayes dengan Studi Kasus PMI Semarang [3], penelitian tentang analisis performa algoritma naïve bayes untuk penentuan kelayakan pendonor darah [4], serta penelitian lain untuk melakukan prediksi calon pendonor darah potensial dengan menggunakan beberapa Algoritma, diantaranya *Naïve Bayes*, *K-nearest Neighbors* dan *Decision Tree C4.5* [5]. Selain itu penelitian lain yang menerapkan Algoritma *Classifier* dalam mengklasifikasi donor darah berbasis web juga sudah pernah dilakukan sebelumnya pada Studi Kasus Palang Merah Indonesia Kota Kebumen [6].

Dari beberapa penelitian yang telah dilakukan, tingkat keberhasilan metode diukur dengan menghitung nilai akurasi serta laju error. Pada penelitian ini akan dilakukan Perbandingan Metode Klasifikasi Naïve Bayes dan K-Nearest Neighbor untuk memperoleh nilai akurasi terbaik dari kedua metode yang dibandingkan dan diharapkan mampu membantu mempermudah proses penyeleksian calon pendonor darah yang benar-benar layak berdasarkan kriteria yang sudah ditentukan.

### 1.1 K-Nearest Neighbor (K-NN)

Algoritma *k-NN* merupakan salah satu metode klasifikasi yang sederhana tetapi dapat memberikan hasil yang baik [2]. Proses klasifikasi dilakukan dengan memperhitungkan kelas atau label *k* tetangga terdekatnya. Dekat jauhnya tetangga dihitung berdasarkan jarak, salah satunya adalah *jarak Euclidean*. Jarak Euclidean merupakan suatu interpretasi kedekatan jarak diantara dua objek atau kemiripan antara data *training* dan data *testing*. Semakin kecil jarak yang dihasilkan, maka semakin mirip data yang dibandingkan [7]. Jarak Euclidean ditunjukkan pada persamaan (1) sebagai berikut

$$Dist(a, b) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}, \quad (1)$$

dimana :

$D(a,b)$  : jarak Euclidean antara data a dan data b  
 $x$  dan  $y$  : data (fitur) yang dibandingkan

$n$  : total data yang dibandingkan

### 1.2 Naïve Bayes

*Naïve bayes* adalah suatu metode klasifikasi berbasis probabilitas yang sederhana, akan tetapi memiliki performa yang baik, yang ditandai dengan tingkat akurasi yang cukup tinggi meskipun diaplikasikan dalam jumlah data yang cukup besar [8]. Formula *Naïve Bayes* ditunjukkan pada persamaan (2) sebagai berikut

$$P(Y|X) = \frac{P(Y) \prod_{i=1}^q P(X_i|Y)}{P(X)}, \quad (2)$$

dimana

$P(Y|X)$  adalah probabilitas data dengan titik  $X$  pada kelas  $Y$

$P(X)$  adalah probabilitas pada  $X$  nilainya selalu tetap.

$P(Y)$  adalah probabilitas awal pada kelas  $Y$

$\prod_{i=1}^q P(X_i|Y)$  adalah probabilitas independent kelas  $Y$  dari semua fitur dalam titik  $X$ .

### 1.3 Confusion Matriks

*Confusion Matriks* adalah teknik untuk perhitungan kinerja suatu algoritma klasifikasi dengan menggunakan matriks, dimana jumlah kolom pada matriks konfusi harus sesuai dengan jumlah kolom pada data set. *Confusion Matriks* pada dasarnya menghasilkan dua nilai yaitu: nilai akurasi dan laju error.

Rumus untuk menghitung akurasi dan laju error masing-masing ditunjukkan pada persamaan (3) dan (4)

$$\text{Akurasi} = \frac{\text{jumlah data yang diprediksi secara benar}}{\text{jumlah prediksi yang dilakukan}} \quad (3)$$

$$\text{Laju Error} = \frac{\text{Jumlah data yang diprediksi Salah}}{\text{Jumlah prediksi yang dilakukan}} \quad (4)$$

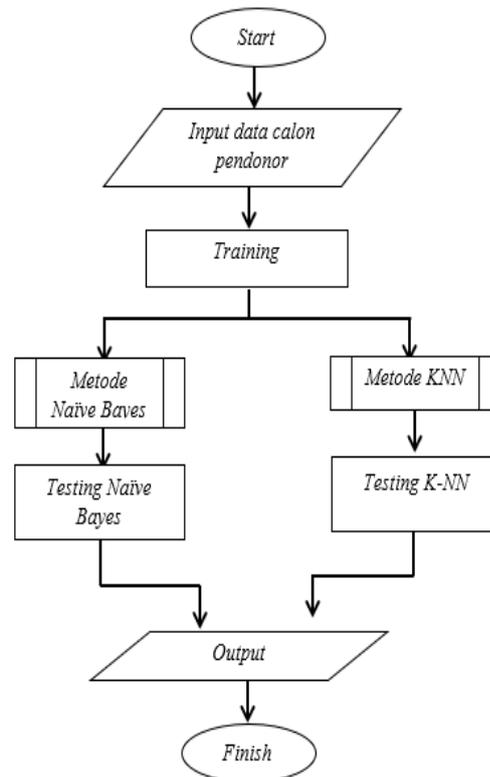
## 2. METODE

Secara umum, sistem yang akan dibuat ditunjukkan pada gambar 1. Sistem yang dibuat berbasis *website* menggunakan bahasa pemrograman *PHP* dengan menerapkan dua metode yaitu *naïve bayes* dan *KNN*.

Adapun data yang akan digunakan dalam penelitian ini adalah bersumber dari data yang didapatkan dari Unit Transfusi Darah RSUD Kab. Majene dan data pendonor darah yang akan diolah dalam klasifikasi penentuan calon pendonor darah. Selain itu dalam penelitian ini juga membutuhkan data pendukung yang bersumber dari buku, jurnal, dan literatur lainnya yang relevan dengan penelitian ini berkaitan dengan fitur apa yang mendukung dalam dalam klasifikasi penentuan calon pendonor darah.

Data yang diperoleh kemudian di-*training* dan menghasilkan dua model *classifier* yaitu *naïve bayes* dan *k-NN*. Proses klasifikasi kemudian dilakukan dengan

menggunakan data testing yang mengacu pada model klasifikasi yang sudah dibangun sebelumnya. Akan diperoleh dua hasil klasifikasi berdasarkan metode *naïve bayes* dan metode *k-NN*. Hasil klasifikasi tersebut dibandingkan untuk mengukur performa metode yang mana yang lebih baik berdasarkan nilai akurasi dan laju error.



Gambar 1. Diagram Alir Penyelesaian penentuan calon pendonor darah

### 3. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data pendonor darah Unit Transfusi Darah Rumah Sakit Umum Daerah Kab. Majene Provinsi Sulawesi Barat Tahun 2019-2020 berdasarkan data Unit Transfusi Darah oleh Rumah Sakit Umum Daerah Kab. Majene pada tahun 2020. Terdapat 7 parameter atau atribut yang dalam penentuan calon pendonor darah yang diperlihatkan pada tabel 1. Keseluruhan data yang digunakan sebanyak 150 data yang dibagi menjadi 100 data untuk data training dan 50 data untuk data testing untuk kedua metode yang digunakan. Terdapat dua kelas dalam proses klasifikasi ini, yaitu kelas layak dan tidak layak. Kelas layak dimaksudkan sebagai kelas kategori calon pendonor yang diperbolehkan untuk melakukan donor darah sedangkan kelas tidak layak sebagai kelas untuk calon pendonor yang tidak diperbolehkan untuk mendonorkan darah.

#### 3.1 Tahap Pengujian Akurasi

Hasil pengujian terhadap metode *Naïve Bayes* dilakukan dengan menggunakan Bahasa pemrograman *PHP*. Pada penelitian ini pengujian dilakukan sebanyak 1 (satu) kali dengan jumlah data *training* 100 data dan data testing sebanyak 50 data calon pendonor darah. Proses uji coba dilakukan untuk mengetahui hasil akurasi terhadap metode klasifikasi yang terbentuk dari 50 data testing yang digunakan untuk memprediksi sebuah data baru yang belum diketahui label kelasnya berdasarkan pengetahuan dan pembelajaran dari data sebelumnya (*data training*).

Tabel 1 Kriteria (fitur)

No.	Nama Kriteria	Sub Kriteria
1.	Berat Badan	> 45
2.	Kadar Hemoglobin	12gr untuk perempuan 12,5gr untuk laki-laki
3.	Tekanan Darah Sistolik	110-160 mmHg
4.	Tekanan Darah Diastolik	70-100 mmHg
5.	Usia	17-60 tahun
6.	JenisKelamin	Perempuan Laki-laki
7.	Jenis Penyakit	Ada Tidak ada

Bedasarkan hasil perhitungan nilai preferensi sebelumnya yang telah di *testing* menggunakan algoritma *naïve bayes*, sebanyak 38 data diklasifikasikan layak dan 12 data untuk kelas tidak layak. Hasil akurasi diketahui dari berapa jumlah data prediksi yang benar dibagi dengan jumlah prediksi yang dilakukan. Untuk menghitung akurasi dan laju error dapat menghitung menggunakan persamaan (3) dan (4)

$$Akurasi = \frac{38}{50} = 0.76 \text{ atau } 76\%$$

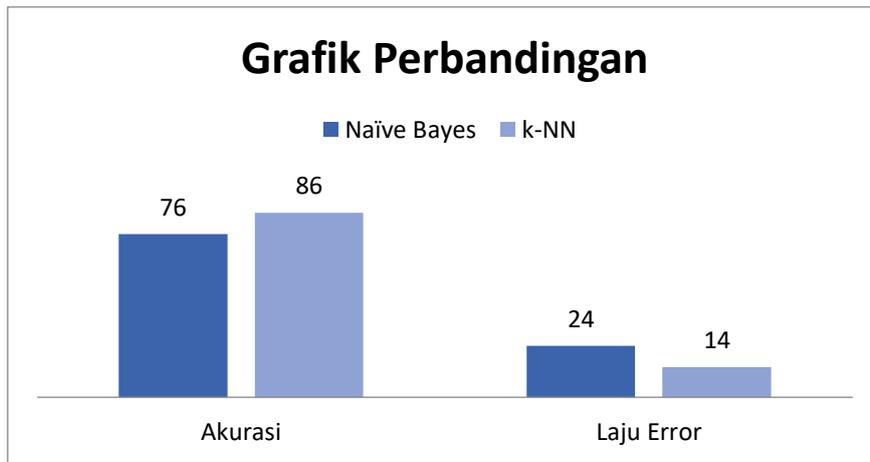
$$Laju Error = \frac{12}{50} = 0.24 \text{ atau } 24\%$$

Hasil pengujian menggunakan metode *K-Nearest Neighbor* dengan menggunakan 100 data *training* dan 50 data *testing* dari calon pendonor darah, terdapat 43 data untuk kelas layak dan 7 data yang diklasifikasikan tidak layak untuk

mendonorkan darahnya berdasarkan jumlah kategori terbanyak dari tetangga terdekat. Jumlah tetangga terdekat  $k$  yang digunakan dalam penelitian ini adalah  $k = 3$ . Hasil akurasi diketahui dari jumlah prediksi data secara benar dan dibagi dengan jumlah data yang di uji. Untuk menghitung nilai akurasi dapat menggunakan persamaan 3 dan perhitungan laju error menggunakan persamaan 4

$$Akurasi = \frac{43}{50} = 0.86 \text{ atau } 86\%$$

$$Laju\ Error = \frac{7}{50} = 0.14 \text{ atau } 14\%$$



Gambar 2. Grafik perbandingan nilai akurasi dan laju error metode naïve bayes dan knn

### 3.2 Implementasi Sistem

Implementasi system adalah penerapan model system klasifikasi untuk rekomendasi penentuan calon pendonor darah menggunakan perhitungan metode *Naïve Bayes* dan *K-Nearest Neighbor*.

#### 1. Form Halaman Nilai Atribut

Gambar 3 menunjukkan halaman pemberian nilai setiap atribut dimana pada halaman ini admin akan diarahkan untuk melakukan penginputan nilai berdasarkan setiap atribut. Pada system ini halaman nilai atribut dikategorikan sebagai kategorikal maka system akan menginput nilai atribut seperti jenis kelamin untuk membedakan jenis kelamin perempuan dan laki-laki.

Naive Bayes   Atribut   Nilai Atribut   Dataset   Perhitungan   Password   Logout				
Nilai Atribut				
No	Kode	Nama Atribut	Nama Nilai Atribut	Aksi
1	A06	Jenis Kelamin	Laki-Laki	 
2	A06	Jenis Kelamin	Perempuan	 
3	A07	Penyakit	Ada	 
4	A07	Penyakit	Tidak Ada	 
5	A08	Status	Layak	 
6	A08	Status	Tidak Layak	 

Gambar 3. Tampilan Halaman Form Atribut

2. Tampilan Form Dataset

Gambar 4 merupakan halaman dataset adalah dimana admin akan menginput dataset dengan kata lain adalah calon pendonor darah yang dijadikan sebagai data *training* dalam pengklasifikasian calon pendonor darah menggunakan metode *Naïve bayes*

Nomor	Nama	Kadar Hemoglobin	Berat Badan	Usia	Tekanan Darah Sistolik	Tekanan Darah Diastolik	Jenis Kelamin	Penyakit	Status	Aksi
1	Iham	14	60	33	110	70	Laki-Laki	Tidak Ada	Layak	
2	Tamsil	17	30	59	110	70	Laki-Laki	Tidak Ada	Layak	
3	Jufry	15	70	37	120	80	Laki-Laki	Tidak Ada	Layak	
4	Iham Akbar	16	85	25	150	70	Laki-Laki	Tidak Ada	Layak	
5	Aswar	17	60	22	120	80	Laki-Laki	Tidak Ada	Layak	
6	Muhammad Ridwan	11	50	41	100	70	Laki-Laki	Tidak Ada	Tidak Layak	
7	Firman	15	50	19	110	70	Laki-Laki	Tidak Ada	Layak	
8	Musthopa Kamel	16	65	26	120	80	Laki-Laki	Ada	Layak	
9	Marwah Asruddin	10	80	25	100	60	Perempuan	Ada	Tidak Layak	
10	Herewati	13	47	22	110	80	Perempuan	Tidak Ada	Layak	

Gambar 4. Halaman Form Dataset

3. Tampilan Form Perhitungan

Halaman ini berfungsi untuk mengolah data perhitungan naïve bayes pengklasifikasian calon pendonor darah. Halaman Perhitungan adalah halaman untuk mengolah data *testing*. Setelah melakukan penginputan data *testing* maka perhitungan otomatis akan dilakukan.

Gambar 5. Halaman Perhitungan menggunakan metode naïve bayes

#### 4. KESIMPULAN

Berdasarkan hasil analisis dan pembahasan, maka dapat diambil kesimpulan bahwa *algoritma K-Nearest Neighbor* lebih baik dalam melakukan klasifikasi data calon pendonor dara berdasarkan tingkat akurasi yang lebih tinggi serta laju error yang lebih kecil dibandingkan dengan metode naïve bayes. Tingkat akurasi *algoritma K-Nearest Neighbor* sebesar 86% serta laju error yaitu 14% sedangkan untuk algoritma *Naïve Bayes* memiliki akurasi 76% dan laju error 24%.

#### REFERENSI

- [1] Depkes RI.(2009). Donor Darah, Hidup Sehat Sambil Beramal. Jakarta. [www.health.detik.com](http://www.health.detik.com) (diakses pada Jam 19.44 7 July 2010)
- [2] Putri, R.L., Suparti., & Rahmawati, R.(2014). Perbandingan Metode Klasifikasi Naïve Bayes Dan K-Nearest Neighbor Pada Analisis Data Status Kerja Di Kabupaten Demak Tahun 2012
- [3] Kurniawan, A. (2015).Penentuan Calon Pendonor Darah Menggunakan Algoritma Naïve Bayes Classification (Studi Kasus Pmi Semarang)
- [4] Coastera, A. F., Yusa, M., Wara, N. L., & Sari, J. P.(2020). Analisis Performa Algoritma Naïve Bayes Untuk Penentuan Kelayakan Pendonor Darah
- [5] Wahono, H., & Riana, D. (2020). Prediksi Calon Pendonor Darah Potensial Dengan Algoritma Naïve Bayes, KnearestNeighbors dan Decision Tree C4.5
- [6] Rudiantoro, A.(2019). Penerapan Algoritma Naïve Bayes Classifier Untuk Klasifikasi Donor Darah Berbasis Web (Studi Kasus Palang Merah Indonesia Kota Kebumen)
- [7] A P Irfan et al.2020. Parking Slot detection using GLCM and Similarity Measure. IOP Conf. Ser.: Mater. Sci. Eng. 875 012094
- [8] S Cokrowibowo et al. 2020. Web Page Extraction and Classification Using JSOUP and Naïve Bayes. IOP Conf. Ser.: Mater. Sci. Eng. 875 012089