

# Application Algorithm Naïve Bayes for Classification of Dengue Fever in dr. Achmad Darwis Hospital

Viola Yuniza, Atus Amadi Putra\*, Nonong Amalita, Fadhilah Fitri

<sup>1</sup>Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

\*Corresponding author: [atusamadiputra@fmipa.unp.ac.id](mailto:atusamadiputra@fmipa.unp.ac.id)

Submitted : 01 November 2023

Revised : 07 Januari 2024

Accepted : 16 Februari 2024

## ABSTRACT

Dengue fever is a disease transmitted by the bite of the *Aedes aegypti* mosquito. Central Agency of Statistic of Lima Puluah Kota District reported that the morbidity rate of this disease was 14.40% per 100,000 population, which was higher than the previous year's morbidity rate of 3.30% per 100,000 population. The main symptoms of this disease are fever lasting 2-7 days, muscle and joint pain with or without rash, dizziness, and even vomiting blood. Dengue infection can cause various clinical symptoms ranging from dengue fever, dengue hemorrhagic fever to dengue shock syndrome. Therefore, a classification method is needed to help and facilitate early diagnosis of this disease. The method used is the Naive Bayes algorithm by classifying the positive and negative patients with dengue fever. The purpose of this research is to determine the classification of patients with dengue fever disease and the accuracy of using the Naive Bayes algorithm. The results of the analysis stated that the Naïve Bayes model successfully classified patients into 12 Dengue fever positive patients and 22 Dengue fever negative patients based on 34 testing data. The accuracy of the model is 91,18%, which shows that the model is very good in classifying Dengue fever patients.

**Keywords:** Classification, Confusion Matrix, DBD, Naïve Bayes,



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

## I. PENDAHULUAN

Demam Berdarah *Dengue* (DBD) adalah penyakit yang ditularkan melalui gigitan nyamuk *Aedes Aegypti*. DBD dijadikan sebagai salah satu ancaman kesehatan global di antara 10 penyakit lainnya oleh WHO (*World Health Organization*) pada awal tahun 2020. Pada tahun 2020, WHO menyatakan bahwa DBD telah menjadi masalah kesehatan masyarakat di dunia selama lebih dari lima dekade. DBD berpotensi memicu kejadian luar biasa apabila tidak dilakukan penanganan secara cepat. Infeksi DBD dapat menimbulkan gejala klinis yang beragam, mulai dari demam *dengue*, demam berdarah *dengue*, hingga menimbulkan sindrom syok *dengue*. Menurut Hidayani (2020) gejala klinis dari DBD yaitu demam yang berlangsung selama 2-7 hari, nyeri yang dirasakan pada otot dan sendi yang disertai dengan adanya ruam atau tanpa ruam, jumlah trombosit <100.000 sel/ml, pusing, bahkan muntah darah.

Berdasarkan Laporan Tahunan Kementerian Kesehatan RI tahun 2022 mengenai DBD, Provinsi Sumatera Barat merupakan provinsi dengan kasus DBD tertinggi urutan ke-10 di Indonesia dengan jumlah kasus sebanyak 4.024. Rencana strategi Kementerian Kesehatan 2020-2024 mengenai DBD adalah penanggulangan nasional. Hal ini bertujuan untuk menurunkan beban kesehatan masyarakat dengan menetapkan target 80% kabupaten/kota memiliki IR *dengue* < 10% per 100.000 penduduk. IR (*Incidence Rate*) adalah jumlah kasus DBD di wilayah tertentu dibagi dengan jumlah penduduk dalam waktu yang sama lalu dikalikan 100.000 penduduk. Kabupaten Lima Puluah Kota memiliki IR *dengue* sebesar 14,40% per 100.000 pada tahun 2022. Angka tersebut melonjak tinggi dibandingkan tahun sebelumnya dengan IR *dengue* 3,30% per 100.000 penduduk (BPS Kabupaten Lima Puluah Kota, 2022). Untuk mendukung dan mencapai target yang ditetapkan oleh Kementerian Kesehatan RI, perlu dilakukan kerjasama oleh RSUD dr. Achmad Darwis selaku RSUD di Kabupaten Lima Puluah Kota dibawah koordinasi Pemerintah daerah dan Departemen Kesehatan Provinsi Sumatera Barat.

Salah satu cara yang dapat dilakukan oleh RSUD dr. Achmad Darwis mencapai target yang ditetapkan oleh Kementerian Kesehatan RI adalah dengan adanya diagnosa awal yang tepat. Untuk dapat membantu dan mempermudah diagnosa penyakit DBD secara dini, diperlukan sebuah metode klasifikasi. Klasifikasi merupakan metode yang bertujuan untuk memilih atau memasukkan pengamatan baru ke dalam kelompok yang telah memiliki label/kelas yang bersifat kategorik. *Naïve Bayes* merupakan salah satu algoritma yang sering digunakan untuk metode

klasifikasi. Menurut Kusri dan Luthfi (2015), *Naïve Bayes* sering digunakan untuk metode klasifikasi karena dapat meminimalkan tingkat kesalahan dibandingkan dengan algoritma lainnya. Penggunaan algoritma *Naïve Bayes* telah banyak dilakukan karena kemampuannya dalam pemodelan prediktif sangat baik (*Science dan Engineering*, 2014). Pada penelitian yang dilakukan oleh Prajarini (2016) dengan *Naïve Bayes* didapatkan nilai akurasi yang cukup tinggi, yaitu sebesar 97,3%. ini berarti bahwa kemampuan algoritma *Naïve Bayes* sangat baik.

Berdasarkan uraian yang telah dijelaskan, maka pada artikel ini akan dibahas mengenai algoritma *Naïve Bayes* untuk klasifikasi pasien penyakit DBD dengan dua kelas, yaitu positif DBD dan negatif DBD di RSUD dr. Achmad Darwis. Tujuan dilakukannya penelitian ini adalah untuk mengetahui hasil klasifikasi pasien penderita penyakit DBD, serta untuk mengetahui tingkat akurasi menggunakan metode *Naïve Bayes*. Hal ini diharapkan dapat membantu RSUD dr. Achmad Darwis untuk meminimalisir penyakit DBD dengan mengklasifikasikan data pasien yang terjangkau penyakit DBD sehingga dapat mengurangi pasien setiap tahunnya.

## II. METODE PENELITIAN

### A. Jenis Penelitian dan Sumber Data

Jenis penelitian ini adalah penelitian terapan dimana memberikan solusi terhadap suatu masalah. Data yang digunakan, yaitu data sekunder berupa data rekam medis pasien DBD di RSUD dr. Achmad Darwis Tahun 2020-2022. Variabel respon (Y) yang digunakan, yaitu hasil diagnosa pasien “positif” DBD dan pasien “negatif” DBD. Sedangkan untuk variabel prediktor (X), yaitu gejala awal yang dialami oleh pasien. Rincian dari variabel Y dan X dapat dilihat pada Tabel 1.

**Tabel 1.** Variabel Penelitian

Variabel	Keterangan	Variabel	Keterangan
Diagnosa DBD (Y)	0 : Negatif 1 : Positif	Nyeri Sendi ( $X_6$ )	0 : Tidak 1 : Ya
Usia ( $X_1$ )	Kontinu	Nafsu Makan ( $X_7$ )	0 : Tidak 1 : Ya
Suhu Badan ( $X_2$ )	Kontinu	Pusing ( $X_8$ )	0 : Tidak 1 : Ya
Demam ( $X_3$ )	Kontinu	Mual, Muntah ( $X_9$ )	0 : Tidak sama sekali 1 : Mual 2 : Muntah
Trombosit ( $X_4$ )	Kontinu	Gusi Berdarah ( $X_{10}$ )	0 : Tidak 1 : Ya
Jenis Kelamin ( $X_5$ )	0 : Laki-laki 1 : Perempuan	Mimisan ( $X_{11}$ )	0 : Tidak 1 : Ya

### B. Teknik Analisis Data

Teknik analisis data pada penelitian ini adalah algoritma *Naïve Bayes* menggunakan *software R studio*. Langkah-langkah yang dilakukan dalam analisis data adalah sebagai berikut.

1. Mengumpulkan data langsung ke RSUD dr. Achmad Darwis
2. Membagi data *training* dan data *testing*.

Pada penelitian ini data set dibagi menjadi data *training* sebesar 80% dan data *testing* sebesar 20% secara acak.

3. Melakukan klasifikasi menggunakan algoritma *Naïve Bayes*

*Naïve Bayes* adalah salah satu metode *machine learning* berdasarkan nilai probabilitas suatu kelas. Menurut Saleh (2015) cara kerja dari metode *Naïve Bayes* yaitu memprediksi peluang di masa yang akan datang berdasarkan pengalaman di masa lampau. Kelebihan dari metode *Naïve Bayes* hanya membutuhkan jumlah data *training* (data latih) yang kecil untuk menentukan estimasi parameter yang diperlukan pada proses pengklasifikasian. *Teorema Bayes* akan ditunjukkan oleh Persamaan (1) serta langkah-langkah untuk menghitung algoritma *Naïve Bayes* (Han, jiawei, 2006:311-313)

$$P(C_j|X_i) = \frac{P(X_i|C_j)P(C_j)}{P(X_i)} \quad (1)$$

Dimana :

- $i$  : indeks atribut
- $j$  : indeks kelas

- $P(C_j|X_i)$  : probabilitas kelas ke- $j$  berdasarkan kondisi  $X_i$  (*posterior probabilitas*)
- $P(C_j)$  : probabilitas masing-masing kelas (*prior probabilitas*)
- $P(X_i|C_j)$  : probabilitas atribut ke- $i$  pada saat kelas ke- $j$
- $P(X_i)$  : probabilitas masing-masing atribut ke- $i$

Untuk menghitung *prior probabilitas* yang mewakili suatu kelas sebelum masuknya sampel digunakan rumus pada Persamaan (2).

$$P(C_j) = \frac{C_{j,D}}{D} \tag{2}$$

Dimana :

$C_{j,D}$  : jumlah data *training* dari kategori  $C_i$  di  $D$

$D$  : jumlah seluruh data *training*

Asumsi naif tentang independensi yang mengasumsikan bahwa nilai-nilai pada variabel secara kondisional independen satu sama lain atau tidak ada ketergantungan di antara variabel maka diperoleh rumus untuk menghitung nilai probabilitas untuk masing-masing atribut pada Persamaan (3).

$$P(C_j) = \prod_{k=1}^n P(C_j) \tag{3}$$

$$= P(C_j)P(C_j) \dots P(x_i|C_j)$$

Dari persamaan diatas  $x_k$  mengacu pada nilai variabel  $X$  bebas ( $x_1, x_2, \dots, x_k$ ) yang dalam variabel  $X$  tersebut ada nilai berturut-turut  $A_k$  adalah ( $A_1, A_2, \dots, A_n$ ). Untuk masing-masing variabel  $X$  ada yang bersifat kategorik dan kontinu, maka untuk menentukan probabilitas pada variabel yang bersifat kategorik digunakan rumus pada Persamaan (4).

$$P(C_j) = \frac{A_k}{C_{j,D}} \tag{4}$$

Untuk menentukan probabilitas pada variabel  $X$  yang bersifat kontinu diasumsikan memiliki distribusi *Gaussian* dengan nilai *mean* ( $\mu$ ) dan standar deviasi ( $\sigma$ ). Berikut rumus untuk distribusi *Gaussian* pada Persamaan (5).

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{5}$$

Sehingga :

$$P(C_j) = g(x_k, \mu_{C_i}, \sigma_{C_i})$$

Adapun rumus dalam menentukan nilai *mean* dan standar deviasi yaitu menggunakan Persamaan (6) dan Persamaan (7) (Suntoro, 2018:15)

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \tag{6}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1}} \tag{7}$$

4. Melihat keakuratan hasil klasifikasi menggunakan tabel *confusion matrix*

*Confusion matrix* adalah tahapan untuk mengevaluasi kinerja suatu metode klasifikasi yaitu membandingkan hasil prediksi model dengan nilai sebenarnya dari data uji (Han, Kember dan Pei, 2011). Keluaran dari *confusion matrix* berbentuk tabel, yang disajikan pada Tabel 2.

**Tabel 2.** *Confusion Matrix* Klasifikasi Dua Kelas

Klasifikasi	Kelas Prediksi	
	Class : Yes	Class : No
Observed Class Class : Yes	True Positive (TP)	False Negative (FN)
Class : No	False Positive (FP)	True Negative (TN)

(Sumber: Gorunescu, 2011:320)

Keterangan :

- a. TP *observed class* benar yang terprediksi sebagai kelas benar.
- b. TN *observed class* salah yang terprediksi sebagai kelas salah.
- c. FP *observed class* salah yang terprediksi sebagai kelas benar.
- d. FN *observed class* benar yang terprediksi sebagai kelas salah.

Menurut Gorunescu (2011) pengukuran performa dalam klasifikasi dapat dilihat dengan nilai *accuracy*, *sensitivity (recall)*, dan *specificity* yang hasil keluarannya dalam bentuk persentase.

1. *Accuracy*

*Accuracy* adalah rasio prediksi (TP dan TN) dengan keseluruhan data, nilai *accuracy* dapat dicari dengan menggunakan rumus pada Persamaan (8).

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{8}$$

2. *Sensitivity (Recall)*

*Sensitivity (recall)* adalah prediksi benar positif dikomparasikan dengan keseluruhan data yang benar positif, nilai *sensitivity* dapat dicari dengan menggunakan rumus pada Persamaan (9).

$$sensitivity = \frac{TP}{TP + FN} \tag{9}$$

3. *Specificity*

*Specificity* adalah nilai keberhasilan memprediksi negatif terhadap keseluruhan negatif, berikut rumus mencari *specificity* pada Persamaan (10).

$$specificity = \frac{TN}{TN + FP} \tag{10}$$

### III. HASIL DAN PEMBAHASAN

Data yang dikumpulkan berupa data harian pengecekan apakah seorang pasien didiagnosa penyakit DBD atau tidak di RSUD dr. Achmad Darwis pada tahun 2020 – 2022. Diperoleh sebanyak 170 data yang kemudian dibagi menjadi data *training* dan data *testing*. Hasil pembagian data disajikan dalam Tabel 3.

**Tabel 3.** Hasil Pembagian Data

Kelompok Data	Label Data	
	Positif	Negatif
Data <i>training</i>	85	51
Data <i>testing</i>	9	25
<b>Total</b>	94	76

Berdasarkan Tabel 3, diketahui bahwa jumlah pasien yang didiagnosa positif DBD lebih banyak daripada yang terdiagnosa negatif DBD. Selanjutnya dilakukan pemodelan menggunakan metode *Naïve Bayes* untuk mengklasifikasikan pasien penyakit DBD berdasarkan variabel yang tersedia. Proses pemodelan klasifikasi dilakukan menggunakan data *training* kemudian dievaluasi menggunakan hasil prediksi pada data *testing*. Dalam memperoleh model klasifikasi dengan menggunakan *Naïve Bayes*, langkah pertama adalah mencari probabilitas prior untuk masing-masing kelas dengan menggunakan Persamaan (2), hasil perhitungannya dapat dilihat pada Tabel 4.

**Tabel 4.** Probabilitas Kategori Y

Positif	Negatif
0,623	0,377

Berdasarkan tabel 4 didapatkan bahwa dari 136 kasus pada data *training*, probabilitas pasien yang positif DBD lebih besar daripada pasien yang negatif DBD yaitu sebesar 0,623. Sedangkan probabilitas pasien yang negatif DBD adalah 0,377. Hal tersebut dikarenakan pada hasil penelitian dalam kasus ini, sebagian besar pasien didiagnosis positif DBD. Selanjutnya menghitung probabilitas masing-masing atribut terhadap masing-masing kelas ( $P(X_i|Y)$ ). Hasil

perhitungan probabilitas setiap atribut katgorik pada masing-masing kelas menggunakan Persamaan (4) dapat dilihat pada Tabel 5.

**Tabel 5.** Probabilitas Atribut Kategorik pada Masing-Masing Kelas

Atribut	Kategori	Jumlah Data		Probabilitas	
		Positif	Negatif	Positif	Negatif
Jenis Kelamin (X <sub>5</sub> )	Laki – Laki	34	23	0,4000	0,4510
	Perempuan	51	28	0,6000	0,5490
Nyeri Sendi (X <sub>6</sub> )	Tidak	53	44	0,6235	0,8627
	Ya	32	7	0,3765	0,1373
Nafsu Makan (X <sub>7</sub> )	Tidak	30	34	0,3529	0,6667
	Ya	55	17	0,6471	0,3333
Pusing (X <sub>8</sub> )	Tidak	47	31	0,5529	0,6078
	Ya	38	20	0,4471	0,3922
Mual Muntah (X <sub>9</sub> )	Tidak sama sekali	29	19	0,3412	0,3725
	Mual	11	13	0,1294	0,2549
	Muntah	45	19	0,5294	0,3725
Gusi Berdarah (X <sub>10</sub> )	Tidak	77	51	0,9059	1,0000
	Ya	8	0	0,0941	0,0000
Mimisan (X <sub>11</sub> )	Tidak	76	50	0,8941	0,9804
	Ya	9	1	0,1059	0,0196

Sementara itu, untuk menghitung probabilitas atribut yang bertipe numerik dilakukan menggunakan Persamaan (5) berdasarkan kondisi dari masing-masing datanya. Namun sebelum itu dihitung rataan dan standar deviasi dari masing-masing atribut tersebut menggunakan Persamaan (6) dan (7). Adapun hasil perhitungan rataan dan standar deviasi dari atribut yang bertipe numerik dapat dilihat pada Tabel 6.

**Tabel 6.** Rataan dan Standar Deviasi Atribut Numerik

Atribut	Rataan		Standar Deviasi	
	Positif	Negatif	Positif	Negatif
Usia (X <sub>1</sub> )	29	30	20	22
Suhu Badan (X <sub>2</sub> )	37,6	37,9	1,37	0,99
Demam (X <sub>3</sub> )	5	5	2,33	2,62
Trombosit (X <sub>4</sub> )	106094	142902	62355	78632

Selanjutnya adalah menentukan klasifikasi untuk mendiagnosis pasien positif DBD atau negatif DBD dengan menghitung probabilitas posterior menggunakan Persamaan (3). Berikut contoh perhitungan probabilitas posterior pasien dengan kondisi pada Tabel 7.

**Tabel 7.** Data *Testing* Pertama

X <sub>1</sub>	X <sub>2</sub>	X <sub>3</sub>	X <sub>4</sub>	X <sub>5</sub>	X <sub>6</sub>	X <sub>7</sub>	X <sub>8</sub>	X <sub>9</sub>	X <sub>10</sub>	X <sub>11</sub>	Y
14	37,1°C	4	253000	Perempuan	Tidak	Tidak	Tidak	Tidak Sama Sekali	Tidak	Tidak	?

Probabilitas posterior diperoleh dengan mengalikan semua hasil probabilitas kelas positif dan mengalikan semua hasil probabilitas kelas negatif. Setelah dilakukan perkalian dilihat nilai mana yang paling tertinggi. Berikut hasil perkalian seluruh probabilitas kelas positif dan perkalian seluruh probabilitas negatif untuk data *testing* pertama.

$$P(X|\text{Positif}) = 0,623 * 0,067 * 0,319 * 0,238 * 9,96E-05 * 0,60 * 0,6235 * 0,3529 * 0,5529 * 0,3412 * 0,9059 * 0,8941 = 6,412E-09$$

$$P(X|\text{Negatif}) = 0,377 * 0,0653 * 0,2893 * 0,2292 * 0,0005 * 0,549 * 0,8627 * 0,6667 * 0,6078 * 0,3725 * 1 * 0,9804 = 6,111E-08$$

Jadi untuk pasien berusia 14 Tahun dengan suhu tubuh 37,1°C mengalami demam selama 4 hari, jumlah trombosit 253000, berjenis kelamin perempuan, tidak merasakan gejala nyeri sendi, nafsu makan menurun, serta tidak merasakan pusing, tidak mengalami gejala mual atau muntah, tidak mengalami gejala gusi berdarah dan tidak

mengalami mimisian diklasifikasikan ke pasien negatif DBD dibuktikan dengan perolehan nilai probabilitas pada kategori negatif yang lebih tinggi yaitu sebesar 6,111E-08 sedangkan untuk kategori positif diperoleh nilai probabilitas 6,412E-09.

Hasil klasifikasi model *Naïve Bayes* untuk keseluruhan data *testing* disajikan dalam Tabel 8.

**Tabel 8.** Hasil Prediksi Model terhadap Data *Testing*

No	Prob. Positif	Prob. Negatif	Prediksi	Aktual
1	6,41189E-09	6,11083E-08	NEGATIF	NEGATIF
2	7,4649E-09	2,40258E-08	NEGATIF	NEGATIF
3	5,40948E-09	5,26617E-08	NEGATIF	NEGATIF
4	1,49491E-07	6,43285E-07	NEGATIF	NEGATIF
5	5,84444E-10	7,23161E-10	NEGATIF	POSITIF
6	8,78993E-09	1,75151E-08	NEGATIF	NEGATIF
7	2,19691E-08	4,90066E-08	NEGATIF	NEGATIF
8	2,47558E-08	5,19771E-08	NEGATIF	NEGATIF
9	1,40653E-08	2,12435E-08	NEGATIF	POSITIF
10	2,92701E-08	5,34503E-08	NEGATIF	NEGATIF
11	1,64564E-08	3,48801E-08	NEGATIF	NEGATIF
12	5,76674E-08	1,8515E-08	POSITIF	NEGATIF
13	1,4375E-09	1,01909E-11	POSITIF	NEGATIF
14	1,32872E-10	9,16208E-09	NEGATIF	POSITIF
15	5,8109E-09	3,00042E-09	POSITIF	POSITIF
16	2,46156E-08	8,79808E-08	NEGATIF	NEGATIF
17	0	2,75387E-09	NEGATIF	POSITIF
18	7,83746E-10	7,84009E-9	NEGATIF	NEGATIF
19	2,19127E-07	1,01204E-07	POSITIF	NEGATIF
20	1,72216E-07	1,8767E-07	NEGATIF	NEGATIF
21	5,83725E-13	9,98973E-12	NEGATIF	NEGATIF
22	4,20192E-08	2,00801E-08	POSITIF	POSITIF
23	0	7,03957E-09	NEGATIF	POSITIF
24	1,9613E-07	6,82717E-07	NEGATIF	NEGATIF
25	7,03839E-10	1,53995E-09	NEGATIF	NEGATIF
26	4,83431E-09	2,35494E-08	NEGATIF	NEGATIF
27	9,83841E-09	1,239E-09	POSITIF	NEGATIF
28	7,12644E-10	0	POSITIF	POSITIF
29	1,16038E-08	2,45769E-09	POSITIF	NEGATIF
30	8,33852E-08	2,38291E-08	POSITIF	NEGATIF
31	7,14168E-17	9,53074E-12	NEGATIF	POSITIF
32	3,61604E-08	2,09099E-08	POSITIF	NEGATIF
33	5,48837E-08	4,05906E-08	POSITIF	NEGATIF
34	2,10545E-09	2,45598E-10	POSITIF	NEGATIF

Hasil analisis menyatakan bahwa model *Naïve Bayes* berhasil mengklasifikasikan pasien menjadi 12 pasien positif DBD dan 22 pasien negatif DBD berdasarkan 34 data *testing*. Untuk mengetahui performa dari model yang

telah dibangun, evaluasi model dilakukan menggunakan hasil prediksi pada data *testing* yang direpresentasikan dalam *Confusion Matrix*. Hasil *Confusion Matrix* menggunakan *software R Studio* dapat dilihat pada Tabel 9.

**Tabel 9.** *Confusion Matrix Naive Bayes*

Kelas Aktual	Hasil Prediksi	
	Kelas 1	Kelas 0
Kelas 1	10	1
Kelas 0	2	21

Berdasarkan Tabel 9 pasien positif DBD diprediksi positif DBD sebanyak 10 orang, pasien negatif DBD yang terprediksi negatif DBD sebanyak 21 orang, pasien yang positif DBD dan diprediksi negatif DBD sebanyak 1 orang, dan pasien yang negatif DBD dan diprediksi positif DBD sebanyak 2 orang. Menggunakan Persamaan (8), (9), dan (10) diperoleh hasil persentase akurasi sebesar 91,18%, hal ini menunjukkan bahwa terdapat 91,18% amatan yang diprediksi secara benar dari keseluruhan data. Persentase untuk *sensitivity* sebesar 91,30% menunjukkan bahwa terdapat 91,30% amatan yang benar diprediksi positif DBD dari keseluruhan amatan yang diprediksi mengalami positif DBD dan hasil persentase *specificity* sebesar 90,91%. Menunjukkan bahwa terdapat 90,91% amatan yang benar diprediksi negatif DBD terhadap keseluruhan amatan yang diprediksi mengalami negatif DBD. Berdasarkan nilai tersebut dapat dikatakan bahwa model klasifikasi yang dibangun memiliki performa yang sangat baik dalam mengklasifikasikan pasien penyakit DBD di RSUD dr. Achmad Darwis.

#### IV. KESIMPULAN

Hasil analisis menyatakan bahwa model *Naive Bayes* berhasil mengklasifikasikan pasien menjadi 12 pasien positif DBD dan 22 pasien negatif DBD berdasarkan 34 data *testing*. Didapatkan hasil akurasi dari model sebesar 91,18% dimana nilai ini menunjukkan bahwa model yang dibangun sangat baik dalam mengklasifikasikan pasien DBD. Dari hasil penelitian yang sudah diperoleh, diharapkan dapat digunakan di Rumah Sakit lainnya untuk mendiagnosa penyakit DBD. Untuk penelitian selanjutnya diharapkan dapat menggunakan metode klasifikasi lain seperti *Decision Tree*, *Neural Network* dan *Support Vector Machine* dengan data yang lebih banyak agar mendapatkan hasil yang lebih baik.

#### DAFTAR PUSTAKA

- Badan Pusat Statistik Kabupaten Lima Puluh Kota. (2022). *Angka Demografi Kabupaten Lima Puluh Kota 2020-2022*. Lima Puluh Kota: BPS Kabupaten Lima Puluh Kota
- Gorunescu, Florin. (2011). *Data Mining: Concepts and Techniques*. Verlag berlin Heidelberg: Springer
- Han, Jiawei dan Kamber, Micheline. (2006). *Data Mining: Concept and Techniques Second Edition*. Morgan: Kaufmann Publishers.
- Han, Jiawei dan Kamber, Micheline. (2011). *Data Mining: Concept and Techniques Third Edition*. Morgan: Kaufmann Publishers.
- Hidayani. (2020). Demam Berdarah Dengue : Perilaku Rumah Tangga dalam Pemberantasan Sarang Nyamuk dan Program Penanggulangan Demam Berdarah Dengue. *Paper Knowledge . Toward a Media History of Documents*, 1–20.
- Kemendes RI. (2022). *Rencana Kerja Tahunan 2022*. Tarakan: Kementerian Nasional Republik Indonesia.
- Kementerian Kesehatan RI. (2022). *Laporan Tahunan 2022 Demam Berdarah Dengue*. Jakarta: Kemendes RI.
- Kursini, L.T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi.
- Prajarini, Dian. (2016). Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Kulit. *Informatic*

*Jurnal*, 1(3), 137-141.

Science, C., & Engineering, S. (2014). Prediction of Different Dermatological Conditions Using Naïve Bayesian Classification. *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(1), 864–868.

Saleh, A. (2015). Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga. *Citec Jurnal*, 2 (3), 207-217.