Terbit *online* pada laman : http://teknosi.fti.unand.ac.id/

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |

Artikel Pernelitian

Diagnosis Dini Demam Berdarah Berdasarkan Data Hematologi Menggunakan Algoritma Machine Learning

Yulia Nita [a,*], *Maya Gian Sister* [b], *Gandung Triyono* [c]

[a,b,c] *Program Studi Magister Ilmu Komputer Universitas Budi Luhur, Indonesia*

INFORMASI ARTIKEL

Sejarah Artikel:
Diterima Redaksi: 09 Juli 2025
Revisi Akhir: 02 September 2025
Diterbitkan *Online*: 13 September 2025

KATA KUNCI

Demam Berdarah Dengue,
Deteksi Dini,
Data Hematologi,
Naïve Bayes,
Sistem Pendukung Keputusan

KORESPONDENSI

E-mail: nitayulia23@gmail.com*

A B S T R A C T

Infeksi virus dengue yang dikenal sebagai DBD masih menjadi tantangan serius dalam layanan kesehatan di Indonesia karena sifatnya yang menular dan terus menimbulkan masalah hingga saat ini. Penyebaran DBD yang cepat dan peningkatan angka kejadian memerlukan strategi deteksi dini yang lebih efektif untuk mencegah komplikasi serius. Sayangnya, metode konvensional seperti pemeriksaan NS1, IgM/IgG, dan PCR masih menghadapi keterbatasan dalam ketersediaan serta biaya. Penelitian ini difokuskan pada pengembangan Sistem Pendukung Keputusan (SPK) yang berbasis algoritma Naïve Bayes dengan memanfaatkan data hematologi rutin untuk mengklasifikasikan tingkat risiko infeksi DBD. Dataset yang digunakan berasal dari platform Kaggle dengan 924 data pasien yang telah melalui tahap pembersihan dan normalisasi. Data yang digunakan terdiri dari variabel-variabel seperti usia, gender, tekanan darah, gula darah, suhu tubuh, denyut jantung, dan level risiko. Algoritma Naïve Bayes dipilih untuk membangun model Atas dasar kapasitasnya dalam mengolah data secara optimal dengan asumsi bahwa setiap atribut bersifat independen. Dataset Pembagian data dilakukan ke dalam dua subset, di mana sebagian besar (80%) ditujukan untuk training, dan sisanya (20%) untuk testing. Kinerja model dievaluasi menggunakan metrik seperti akurasi, presisi, recall, serta F1-score. Dari hasil pengujian, model mampu memperoleh tingkat akurasi sebesar 98,03%, dengan performa sangat baik di seluruh kelas risiko, terutama recall sempurna pada kelas risiko tinggi. Hal ini menunjukkan kemampuan model dalam mengidentifikasi kasus-kasus berisiko tinggi tanpa terlewat. Dengan demikian, penelitian ini membuktikan bahwa data hematologi yang sederhana dapat dimanfaatkan secara optimal untuk deteksi dini DBD. Sistem yang dikembangkan berpotensi menjadi alat bantu diagnosis yang cepat, hemat biaya, dan dapat diimplementasikan secara luas untuk mendukung pelayanan kesehatan primer.

1. PENDAHULUAN

DBD merupakan jenis penyakit infeksi tropis, penyakit ini menyebar melalui gigitan nyamuk Aedes aegypti dan Aedes albopictus yang berperan sebagai pembawa virus dengue. Hingga kini, penyakit ini tetap menjadi permasalahan utama dalam Kesejahteraan kesehatan masyarakat, dengan penekanan pada negara-negara yang belum sepenuhnya maju seperti Indonesia [1]. Wilayah tropis dan subtropis, termasuk Indonesia, memiliki kondisi iklim dan lingkungan yang sangat mendukung siklus hidup vektor penular, seperti suhu tinggi, kelembaban, dan pola pemukiman padat penduduk yang mempercepat penyebaran penyakit. Berdasarkan laporan World Health Organization (WHO), virus dengue telah menyebar di lebih dari 65 negara dengan jumlah kasus global yang dilaporkan mencapai rata-rata 925.896 per tahun [2].

Di Indonesia sendiri, angka kejadian DBD menunjukkan tren yang mengkhawatirkan. Kementerian Kesehatan Republik Indonesia melaporkan bahwa pada tahun 2023 terjadi 114.720 kasus DBD dengan 894 kematian. Angka tersebut meningkat drastis hingga mencapai 210.644 kasus dan 1.239 kematian hanya dalam 43 minggu pertama tahun 2024, tersebar di 259 kabupaten/kota di 32 provinsi. Tidak hanya meningkat dari segi jumlah kasus, cakupan wilayah terjangkit pun meluas hingga mencakup 482 kabupaten/kota [3]. Selain itu, siklus epidemiologis DBD yang sebelumnya berkisar 8–10 tahun kini

https://doi.org/10.25077/TEKNOSI.v11i2.2025.185-191

bergeser menjadi 2–3 tahun sekali, menunjukkan adanya peningkatan frekuensi dan intensitas kejadian yang memerlukan penanganan strategis dan berbasis data [3].

Pencegahan dan pengendalian DBD tidak dapat dilepaskan dari peran deteksi dini, yang krusial dalam mencegah perkembangan Komplikasi serius seperti Syok Dengue (Dengue Shock Syndrome) dan pendarahan akut. Namun, hingga saat ini, fasilitas layanan kesehatan di berbagai daerah masih mengandalkan uji diagnostik serologis seperti NS1 antigen, IgM/IgG, dan PCR yang tidak selalu tersedia secara merata, memerlukan waktu tunggu, serta menimbulkan beban biaya tambahan bagi pasien [4]. Dengan demikian, pengembangan pendekatan pengenalan secara lebih awal yang lebih cepat, efisien, dan dapat diakses secara luas.

Salah satu alternatif yang menjanjikan adalah pemanfaatan data pemeriksaan hematologi rutin seperti kadar trombosit, hematokrit, leukosit, dan hemoglobin, yang umumnya tersedia di hampir seluruh laboratorium klinik. Penelitian sebelumnya [5] menunjukkan bahwa trombositopenia dan peningkatan hematokrit dapat menjadi indikator penting dalam fase awal infeksi DBD dan dapat dimanfaatkan sebagai alat skrining awal sebelum hasil uji serologis diperoleh. Dengan bantuan teknologi informasi, data ini dapat diolah lebih lanjut untuk menghasilkan sistem cerdas yang mendukung pengambilan keputusan medis secara cepat dan akurat.

Sistem Pendukung Keputusan (SPK) berbasis data mining dan algoritma klasifikasi menawarkan solusi inovatif dalam mendeteksi dini DBD. Dengan melatih sistem menggunakan dataset hematologi pasien, Algoritma seperti Pohon Keputusan, Naïve Bayes, Tetangga Terdekat (KNN), dan Hutan Acak (Random Forest) dapat mengenali pola klinis tertentu yang berkorelasi dengan infeksi dengue [6]. SPK semacam ini berpotensi besar untuk mempercepat proses triase, memudahkan tenaga medis dalam menentukan prioritas penanganan, serta mengurangi beban sistem kesehatan saat terjadi lonjakan kasus .

Sebagai respon Dalam rangka menjawab isu yang diidentifikasi, fokus utama penelitian ini adalah pada perancangan dan pengembangan Sistem Pendukung Keputusan berbasis algoritma klasifikasi terhadap parameter hematologi pasien, guna meningkatkan akurasi deteksi dini infeksi DBD. Hasil dari sistem ini diharapkan dapat menjadi alat bantu diagnostik yang mendukung pelayanan kesehatan primer, mempercepat penanganan pasien, serta menjadi kontribusi nyata dalam upaya pengendalian penyakit menular di Indonesia secara holistik dan berkelanjutan.

Terdapat 4 Penelitian sebelumnya yang dijadikan acuan dalam studi ini. Studi pertama [7], menunjukkan bahwa optimasi algoritma Naïve Bayes untuk deteksi dini Demam Berdarah Dengue (DBD) mampu meningkatkan akurasi klasifikasi hingga 92%, setelah dilakukan feature selection dan parameter smoothing. Penelitian kedua [8] yang mengimplementasikan Naïve Bayes untuk klasifikasi pasien DBD di Puskesmas Taman Krocok dan memperoleh akurasi sebesar 90%, dengan f1-score 89% untuk kasus negatif dan 92% untuk kasus positif. Penelitian ketiga [9], berfokus pada diagnosa DBD di Kelurahan Antasan Besar menggunakan metode Naïve Bayes berbasis sistem web, dan menghasilkan tingkat akurasi yang konsisten antara 96% hingga 97% pada tiga kali percobaan, menunjukkan nilai kesalahan klasifikasi yang kecil. Penelitian keempat [10], menggunakan gabungan metode K-Means untuk klasterisasi, Naïve Bayes untuk prediksi, dan Linear Regression untuk peramalan jumlah kasus DBD di Jawa Barat. Dengan 486 data pasien tahun 2014–2022, hasil prediksi menggunakan Naïve Bayes mencapai akurasi 88,27%.

Penelitian ini penting karena menawarkan solusi deteksi dini DBD yang cepat, terjangkau, dan akurat dengan memanfaatkan data hematologi rutin serta algoritma klasifikasi. Kontribusinya terletak pada pengembangan sistem pendukung keputusan yang berperan dalam mendukung tenaga medis saat melakukan diagnosis awal, mempercepat penanganan pasien, serta mendukung pengendalian DBD secara efektif di layanan kesehatan primer.

2. METODE

Dalam studi ini, proses penelitian dibagi ke dalam beberapa langkah yang sistematis. Pada tahap awal, dilakukan penelusuran literatur untuk memahami dasar teori dan penelitian sebelumnya guna memperoleh pemahaman mengenai konsep penyakit Demam Berdarah Dengue (DBD) dan mendalami algoritma klasifikasi yang digunakan, yakni Naïve Bayes. Selanjutnya, dilakukan pengumpulan data dari sumber terbuka, yaitu dataset DBD yang diambil dari platform Kaggle.

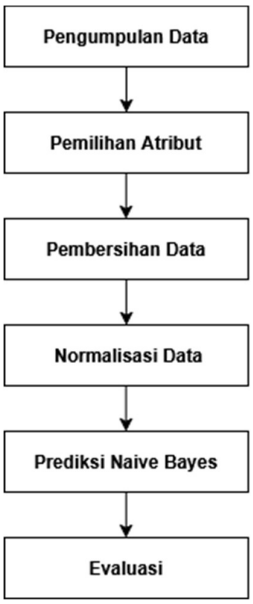
Tahap berikutnya adalah pemilihan atribut, dilanjutkan pembersihan data (data cleaning) untuk menghapus duplikasi dan data tidak lengkap. Setelah itu, dilakukan pengelompokan dan normalisasi atribut numerik agar data berada pada skala yang seragam dan lebih mudah dianalisis oleh model klasifikasi.

Setelah data siap, dilakukan pemodelan dan analisis menggunakan algoritma Naïve Bayes untuk mengklasifikasikan tingkat risiko DBD (Low, Medium, High) untuk mengukur seberapa baik model dalam memprediksi risiko infeksi DBD secara dini. Langkah selanjutnya adalah mengevaluasi kinerja model Dengan memanfaatkan indikator performa seperti tingkat akurasi, ketepatan (presisi), dan daya tangkap (recall), guna

https://doi.org/10.25077/TEKNOSI.v11i2.2025.185-191

menilai efektivitas model dalam melakukan prediksi dini terhadap risiko infeksi DBD.

Tahapan deteksi dini DBD menggunakan algoritma Naive Bayes mulai dari pengumpulan data hingga evaluasi Terlihat dengan jelas dari gambar 1.



Gambar 1. Metode Penelitian

2.1. Pengumpulan Data

Proses pengumpulan data merupakan tahapan terstruktur dalam penelitian yang bertujuan memperoleh data atau informasi penting yang diperlukan untuk merespons pertanyaan penelitian, membuktikan hipotesis, serta merealisasikan sasaran penelitian yang telah ditetapkan. Pentingnya hal ini terletak pada fakta bahwa kualitas data yang dikumpulkan secara langsung memengaruhi keakuratan hasil analisis dan kesimpulan penelitian. Dalam konteks penelitian kuantitatif, data biasanya dikumpulkan dalam bentuk angka atau hasil pengukuran yang dapat diolah secara statistik. Sedangkan dalam penelitian kualitatif, data lebih banyak berupa deskripsi, opini, atau narasi yang dianalisis secara tematik.

2.2. Pemilihan Atribut

Setelah proses pengambilan data dilakukan, langkah berikutnya adalah seleksi atribut. Tahap ini bertujuan untuk mengidentifikasi atribut-atribut yang paling signifikan serta memiliki integritas data yang memadai, yaitu atribut yang tidak memiliki ambiguitas serta jumlah nilai kosong (missing values) yang rendah [11]. Dataset yang digunakan mencakup delapan atribut utama yang merepresentasikan hasil pemeriksaan medis pasien, dengan satu atribut sebagai variabel target klasifikasi.

Seluruh atribut dalam dataset ini dinilai memiliki keterkaitan yang kuat dengan upaya klasifikasi tingkat risiko infeksi Demam Berdarah Dengue (DBD). Oleh karena itu, tidak ada atribut yang dieliminasi dalam tahap ini, karena masing-masing masih berada dalam rentang distribusi data yang dapat diterima dan layak untuk diproses lebih lanjut. Seluruh atribut yang terpilih akan dilibatkan dalam tahap pembersihan data. Adapun daftar Informasi mengenai atribut yang digunakan tersedia pada Tabel 1 berikut ini:

Tabel 1. Pemilihan Atribut

No	Atribut	Deskripsi	Nilai
1	Gender	Jenis kelamin pasien	Pria, Wanita
2	Age	Umur pasien	Bilangan bulat (dalam tahun)
3	SystolicBP	Tekanan darah sistolik	Satuan mmHg
4	DiastolicBP	Tekanan darah diastolik	Satuan mmHg
5	BS	Kadar gula darah (Blood Sugar)	Nilai numerik
6	BodyTemp	Suhu tubuh pasien	Derajat Celcius (°C)
7	HeartRate	Denyut jantung pasien	Detak per menit (bpm)
8	RiskLevel	Tingkat risiko infeksi Demam Berdarah Dengue (target)	Low, Medium, High

Berikut adalah penjelasan masing-masing atribut sesuai dengan Tabel 1. Yang digunakan pada penelitian ini:

a. Gender
   Atribut ini menunjukkan jenis kelamin pasien, dengan tipe data kategorikal (string). Nilai yang terdapat pada atribut ini adalah *Pria* atau *Wanita*. Atribut ini tidak memiliki satuan karena berupa kategori, contoh: *Pria*, *Wanita.*
b. Age
   Atribut ini menunjukkan umur pasien yang dinyatakan dalam satuan tahun dan bertipe data integer. Contoh nilai pada atribut ini antara lain 16, 32, dan seterusnya.
c. SystolicBP
   Atribut ini menunjukkan tekanan darah sistolik pasien, yaitu tekanan saat jantung berkontraksi. Tipe data pada atribut ini adalah *integer*, dengan satuan mmHg (milimeter air raksa). Contoh: 110 mmHg, 120 mmHg, dan sebagainya.
d. DiastolicBP
   Atribut ini merepresentasikan tekanan darah diastolik pasien, yaitu tekanan saat jantung berelaksasi. Tipe data pada atribut ini adalah *integer*, dengan satuan mmHg. Contoh: 70 mmHg, 85 mmHg, dan sebagainya.
e. BS (Blood Sugar)
   Atribut ini menunjukkan kadar gula darah pasien, dengan tipe data *float* atau *numerik desimal*. Satuan yang digunakan adalah mg/dL. Contoh: 90.5 mg/dL, 135.0 mg/dL, dan sebagainya.
f. BodyTemp
   Atribut ini merepresentasikan suhu tubuh pasien, dengan tipe data *float* (desimal). Satuan pada atribut ini adalah derajat Celcius (°C). Contoh: 36.8°C, 38.2°C, dan sebagainya [1].
g. HeartRate
   Atribut ini menunjukkan jumlah denyut jantung pasien per menit, dengan tipe data *integer*. Satuan pada atribut ini adalah bpm (beats per minute). Contoh: 78 bpm, 102 bpm, dan sebagainya.
h. RiskLevel
   Atribut ini merupakan label target untuk mengklasifikasikan tingkat risiko infeksi DBD pada

pasien. Tipe datanya adalah *kategorikal* dengan tiga nilai kelas, yaitu: *Low*, *Medium*, dan *High*. Atribut ini tidak memiliki satuan karena bersifat klasifikasi.

2.3. Pembersihan Data

Penyusunan ulang data agar layak untuk dianalisis dilakukan dengan menghapus Sejumlah baris yang mengandung nilai hilang serta data yang muncul lebih dari satu kali. Ketidaklengkapan dan redundansi data dapat menyebabkan ketidakakuratan dalam proses prediksi serta meningkatkan potensi ambiguitas hasil klasifikasi [6].

2.4. Normalisasi Data

Langkah selanjutnya dalam penelitian ini adalah melakukan pengelompokan dan normalisasi data. Pengelompokan dilakukan untuk memudahkan pemahaman data dengan mengklasifikasikan nilai-nilai numerik ke dalam kategori tertentu yang lebih jelas dan terstruktur [12]. Hal ini penting karena sebagian besar atribut dalam dataset memiliki nilai yang sangat bervariasi.

2.5. Prediksi Algritma Naïve Bayes Data

Setelah tahapan prapemrosesan data selesai dilakukan, data yang telah bersih dan siap pakai Dimanfaatkan dalam proses analisis dengan pendekatan algoritma Naïve Bayes. Analisis ini diawali dengan mengintegrasikan seluruh atribut data klinis pasien, menetapkan sasaran klasifikasi—yaitu tingkat risiko DBD (Low, Medium, High)—secara spesifik, dan memilih pendekatan analisis yang paling tepat untuk mendukung proses pengambilan keputusan berbasis data .

Tahap analisis ini bertujuan untuk mengenali pola-pola yang muncul dalam data klinis, sehingga dapat digunakan untuk menarik kesimpulan berupa prediksi tingkat risiko infeksi DBD pada pasien. Metode Naïve Bayes digunakan karena kemampuannya dalam menghitung probabilitas suatu kejadian berdasarkan informasi dari atribut-atribut yang saling independen, dan memberikan hasil klasifikasi yang cepat serta efisien [9]. Naïve Bayes menggunakan rumus teorema Bayes sebagai dasar perhitungannya, yang dirumuskan sebagai berikut:

$$P(T|X) = \frac{P(X|T).P(T)}{P(X)} \quad (1)$$

2.6. Evaluasi

Model yang dibangun Penilaian dilakukan dengan berbagai metrik kinerja seperti akurasi, presisi, recall, dan F1-score digunakan sebagai metrik evaluasi yang diterapkan pada data pengujian untuk mengukur sejauh mana kemampuan model dalam mengklasifikasikan tingkat risiko infeksi Demam Berdarah Dengue (DBD) secara dini dan tepat. Akurasi menunjukkan jumlah prediksi yang tepat dibandingkan dengan total keseluruhan data pengujian. Precision mengukur tingkat ketepatan model dalam memprediksi satu kelas tertentu, sedangkan recall menunjukkan Sejauh mana model mampu mengidentifikasi semua data yang memang berasal dari kelas yang dimaksud. Sementara itu, F1-score adalah nilai rata-rata harmonik antara precision dan recall, yang menjadi metrik penting khususnya saat distribusi kelas dalam dataset tidak seimbang [13].

$$\text{Accuracy} = \frac{(\text{True Positive} + \quad \text{Negative})}{\text{Total Data}} \times 100\% \quad (2)$$

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positive (TP)} + \text{False Positive (FP)}} \times 100\% \quad (3)$$

$$\text{Recall} = \frac{\text{True Positive (TP}}{\text{True Positive (TP)} + \text{False Negative (FN)}} \times 100\% \quad (4)$$

Tabel 2. Matriks Kebingungan

Data Aktual	Prediksi Low		Prediksi Medium		Prediksi High	
Low	True (Low)	Positive	False (Medium)	Positive	False (High)	Positive
Medium	False (Low)	Positive	True (Medium)	Positive	False (High)	Positive
High	False (Low)	Positive	False (Medium)	Positive	True (High)	Positive

3. HASIL DAN PEMBAHASAN

3.1. Pengumpulan Data

Dalam proses pengumpulan data untuk riset ini, digunakan dataset yang diperoleh dari platform Kaggle sebagai sumber utama, yang telah dimanfaatkan dalam berbagai studi terkait klasifikasi risiko penyakit berbasis data klinis. Dataset yang digunakan dapat diunduh melalui tautan: https://www.kaggle.com/datasets. Dataset ini terdiri dari sekitar 1014 data pasien, dengan 8 atribut serta disajikan dalam format .csv (Comma Separated Values) yang kompatibel dengan berbagai perangkat lunak analisis data.

Tabel 3. Data set tabel pasien Resiko DBD

Gender	Age	SystolicBP	DiastolicBP	BS	BodyTemp	HeartRate	RiskLevel
Male	25	130	80	15	39.5	86	high risk
Female	35	140	90	13	39.7	70	high risk
Male	29	90	70	8	39.2	80	high risk
Male	30	140	85	7	38.2	70	high risk
Male	35	120	60	6.1	37	76	low risk
Female	23	140	80	7.01	38.2	70	high risk
Male	23	130	70	7.01	37.7	78	mid risk
Male	35	85	60	11	38.8	86	high risk
Male	32	120	90	6.9	37.6	70	mid risk
Female	42	130	80	18	38.7	70	high risk
----	---	-----	------	---	---	-----	-----
Male	19	120	80	7	37.8	70	mid risk

https://doi.org/10.25077/TEKNOSI.v11i2.2025.185-191

3.2. *Pembersihan Data*

Tahap awal sebelum proses analisis adalah pembersihan data (data cleaning), Dengan sasaran untuk memastikan Mutu data yang dimanfaatkan dalam tahapan klasifikasi. Dari total 1.014 data awal yang diperoleh dari dataset Kaggle, dilakukan pemeriksaan terhadap kemungkinan adanya duplikasi maupun data yang tidak valid. Hasil pemeriksaan menunjukkan bahwa terdapat 90 data duplikat, yang apabila tidak dihapus dapat menyebabkan bias dalam proses pelatihan model dan menghasilkan prediksi yang tidak akurat.

Setelah dilakukan penghapusan terhadap data duplikat tersebut, jumlah data bersih yang siap digunakan dalam tahap selanjutnya adalah sebanyak 924 data, dengan 8 atribut utama yang mencakup Variabel yang digunakan meliputi Age,Gender, SystolicBP, DiastolicBP, Body Temperature, Blood Sugar (BS), Heart Rate, dengan Risk Level sebagai label klasifikasi.

```
Contoh data setelah dibersihkan:
   Gender  Age  SystolicBP  DiastolicBP    BS  BodyTemp  HeartRate  RiskLevel
0    male   25         130           80  15.0      39.5         86  high risk
1  female   35         140           90  13.0      39.7         70  high risk
2    male   29          90           70   8.0      39.2         80  high risk
3    male   30         140           85   7.0      38.2         70  high risk
4    male   35         120           60   6.1      37.0         76   low risk

 Perbandingan jumlah baris:
Sebelum: 1014 baris
Sesudah: 924 baris (berkurang 90 baris)
```

Gambar 1. Pembersihan Data

3.3. *Normalisasi Data*

Setelah pemilihan data, tahap normalisasi dilakukan untuk menghindari dominasi atribut tertentu akibat perbedaan skala, sehingga proses pembelajaran oleh algoritma klasifikasi menjadi lebih adil dan seimbang [14]. Proses ini membantu meningkatkan akurasi model dalam melakukan prediksi. Tabel 4 berikut menyajikan rentang nilai dan kategori pada masing-masing atribut hasil pengelompokan.

Tabel 4. Normalisasi Data

No	Atribut	Nilai
1	Gender	Pria = 0, Wanita = 1
2	Age	0 – 100 tahun (umum)
3	SystolicBP	90 – 120 mmHg
4	DiastolicBP	60 – 80 mmHg
5	BS	70 – 140 mg/dL (non-puasa)
6	BodyTemp	36,5°C – 37,5°C
7	HeartRate	60 – 100 bpm
8	RiskLevel	Low = 0, Medium = 1, High = 2

3.4. *Implementasi Naïve Bayes*

Setelah proses pembersihan dan persiapan data selesai dilakukan, tahap Langkah berikutnya yaitu penerapan algoritma Naïve Bayes untuk melakukan klasifikasi tingkat risiko infeksi Penyakit Dengue Hemorrhagic Fever (DBD). Pemilihan algoritma ini didasarkan pada kesederhanaannya, kecepatan proses, serta kemampuannya dalam menghasilkan klasifikasi yang cukup akurat pada data dengan atribut yang saling bebas.

Naïve Bayes menggunakan teorema Bayes sebagai dasar kerjanya, dengan anggapan bahwa setiap atribut bersifat independen secara statistik. Dalam implementasinya, sistem akan menghitung probabilitas posterior untuk setiap kelas (Low, Medium, High) berdasarkan data uji yang diberikan [15]. Proses klasifikasi dilakukan dengan menghitung probabilitas setiap kelas berdasarkan kombinasi nilai-nilai atribut pada setiap data pasien, Langkah berikutnya adalah menentukan kelas yang memiliki kemungkinan paling besar sebagai output prediksi.

Dataset dalam penelitian ini berisi 924 data pasien dengan 8 atribut, Selanjutnya, data tersebut dipisahkan menjadi data pelatihan dan pengujian dengan rasio 80 banding 20. Proses pelatihan dilakukan pada 739 data latih, sedangkan 185 data sisanya digunakan untuk menguji performa model.

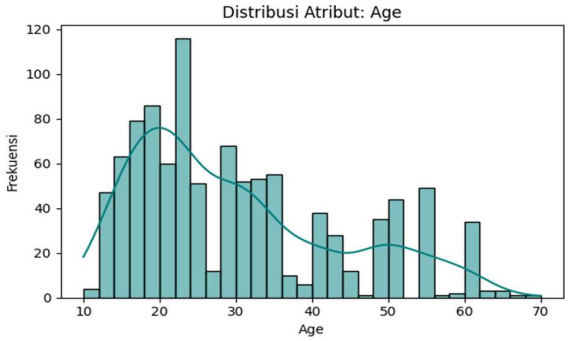
Setiap baris data mewakili satu pasien dan terdiri dari 8 atribut utama, yaitu Age,Gender, SystolicBP, DiastolicBP, Body Temperature, Blood Sugar (BS), Heart Rate, dan Risk Level sebagai target kelas. Nilai-nilai pada atribut tersebut telah dikonversi ke dalam format numerik dan distandardisasi untuk memastikan keseragaman skala.

```
Data Training (10 contoh pertama):
     Gender  Age  SystolicBP  DiastolicBP    BS  BodyTemp  HeartRate  \
847       1   30         120           80   9.0      37.6         76
332       1   23         130           70   6.9      37.7         70
707       1   32         120           90   6.9      37.9         70
218       1   31         120           60   6.1      37.6         76
425       0   35         100           60  15.0      39.5         80
448       0   30         120           75   6.8      37.9         70
722       1   50         140           80   6.7      37.6         70
362       0   40         160          100  19.0      38.8         77
78        0   35         120           80   6.9      37.7         78
29        1   28          90           60   7.2      36.9         82

     RiskLevel
847          1
332          1
707          1
218          1
425          2
448          1
722          1
362          2
78           1
29           0
```

Gambar 2. Konversi Atribut ke Dalam Format Numerik

Selanjutnya, dilakukan analisis awal terhadap distribusi data, khususnya pada atribut Age, yang menjadi salah satu elemen kunci dalam menilai tingkat risiko. Gambar 4. berikut menunjukkan distribusi usia pasien dalam dataset:

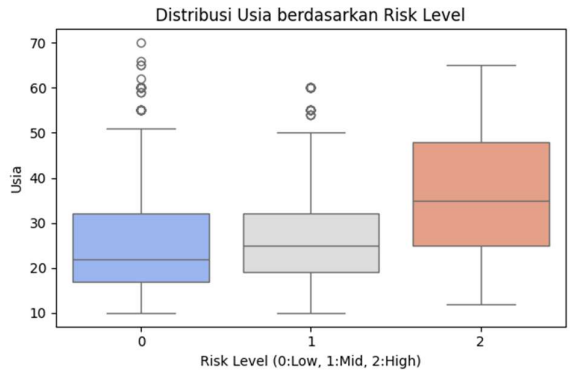


Gambar 3. Distribusi Usia Pasien DBD

Berdasarkan histogram di atas, dapat dilihat bahwa distribusi usia pasien tidak merata dan cenderung terpusat pada kelompok usia muda. Kelompok usia 18 hingga 25 tahun mendominasi dataset dengan frekuensi tertinggi, sedangkan frekuensi mulai menurun setelah usia 30 tahun, dan hanya sedikit data Yang termasuk dalam kategori usia di atas 60 tahun. Ini mengindikasikan bahwa

mayoritas data yang tersedia berasal dari kelompok usia produktif.

Sebagai bagian dari eksplorasi lanjut terhadap data klinis, dilakukan visualisasi distribusi usia pasien terhadap tingkat risiko infeksi DBD. Visualisasi ini bertujuan untuk melihat kecenderungan apakah variabel usia memiliki pengaruh terhadap klasifikasi risiko. Gambar 3 menunjukkan boxplot distribusi usia berdasarkan Risk Level.



Gambar 4. Boxplot distribusi usia berdasarkan Risk Level

Berdasarkan hasil visualisasi tersebut, terlihat bahwa pada kategori Low Risk (0), usia pasien tersebar antara 10 hingga 50 tahun, dengan konsentrasi utama berada di rentang 18–30 tahun. Pada kategori Medium Risk (1), distribusi usia menunjukkan pola yang hampir serupa, namun dengan median usia sedikit lebih tinggi, yaitu sekitar 25 tahun. Sementara itu, pada kategori High Risk (2), distribusi usia tampak lebih tinggi secara keseluruhan, dengan median usia sekitar 35 tahun dan nilai maksimum usia pasien mencapai lebih dari 60 tahun. Hal ini menunjukkan adanya kecenderungan bahwa usia yang lebih tinggi berpotensi masuk dalam kategori risiko infeksi DBD yang lebih serius.

Model Naïve Bayes yang telah dikembangkan kemudian diuji menggunakan 203 data uji. Selanjutnya, performa Evaluasi model dilakukan melalui pengukuran metrik precision, recall, dan F1-score guna masing-masing kategori risiko: Low Risk, Mid Risk, dan High Risk, Seperti yang diperlihatkan pada Gambar 6:

```
Classification Report:
              precision    recall  f1-score   support

    Low Risk       1.00      0.99      0.99        80
    Mid Risk       1.00      0.96      0.98        76
   High Risk       0.92      1.00      0.96        47

    accuracy                           0.98       203
   macro avg       0.97      0.98      0.98       203
weighted avg       0.98      0.98      0.98       203


Akurasi Model: 98.03%
```

Gambar 5. Hasil Klasifikasi

Model Naïve Bayes yang dibangun memiliki tingkat akurasi yang sangat tinggi sebesar 98.03%, dengan performa yang sangat baik di seluruh kelas, terutama pada kelas Low Risk dan Mid Risk. Walaupun precision pada High Risk sedikit lebih rendah (0.92), recall-nya sempurna (1.00), yang berarti tidak ada kasus berisiko tinggi yang terlewat.

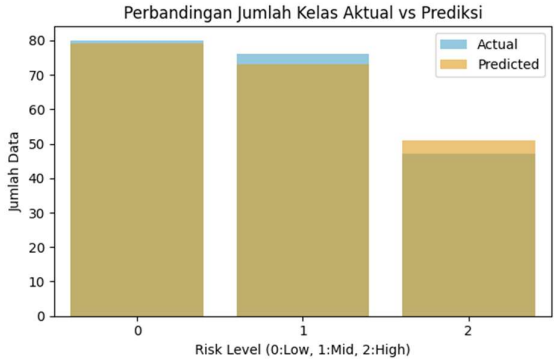
```
Confusion Matrix:
[[79  0  1]
 [ 0 73  3]
 [ 0  0 47]]
```

Gambar 6. Confusion Matrix

Gambar 7 menunjukkan hasil evaluasi model klasifikasi Naïve Bayes terhadap tiga kelas risiko infeksi DBD, yaitu Low, Mid, dan High. Pada baris pertama, tercatat bahwa dari 80 data aktual kelas Low Risk, sebanyak 79 data berhasil diprediksi secara tepat, namun terdapat 1 data yang keliru diklasifikasikan sebagai High Risk. Tidak ada kesalahan klasifikasi ke kelas Mid.

Pada baris kedua, yang mewakili kelas Mid Risk, terdapat 76 data aktual, di mana 73 data berhasil Dikelompokkan secara akurat, dan Tiga entri data keliru diklasifikasikan sebagai High Risk. Kinerja model pada kelas Mid menunjukkan hasil tanpa error diklasifikasikan sebagai Low. Sementara itu, pada baris ketiga untuk kelas High Risk, seluruh 47 data aktual berhasil diprediksi dengan tepat tanpa kesalahan, sehingga menunjukkan akurasi sempurna (100%) pada kelas ini.

Secara keseluruhan, hasil confusion matrix ini memperkuat laporan metrik evaluasi sebelumnya, di mana akurasi model mencapai 98.03%, dan mengindikasikan bahwa model tersebut bekerja dengan performa yang sangat optimal, terutama dalam menghindari kesalahan klasifikasi pada kasus-kasus berisiko tinggi.



Gambar 7. Perbandingan Hasil Prediksi

Gambar 8 menunjukkan perbandingan antara jumlah data aktual dan jumlah data yang diprediksi oleh model untuk masing-masing kelas Risk Level, yaitu Low (0), Mid (1), dan High (2). Warna biru muda mewakili jumlah data aktual, sedangkan warna kuning keemasan menunjukkan jumlah prediksi dari model Naïve Bayes. Berdasarkan visualisasi tersebut, dapat dilihat bahwa hasil prediksi model sangat mendekati jumlah aktual pada setiap kelas. Pada kelas Low Risk, model hampir seluruhnya memprediksi data dengan tepat, dengan selisih yang sangat kecil. Hal serupa terjadi pada kelas Mid Risk, di mana jumlah prediksi hampir sama dengan data sebenarnya. Sementara pada kelas High Risk, terdapat sedikit kelebihan prediksi dibandingkan jumlah aktual. Meskipun demikian, perbedaannya masih dalam batas wajar dan dapat diterima, terutama karena recall pada kelas ini mencapai 100%, yang berarti tidak ada kasus risiko tinggi yang terlewatkan oleh model.

https://doi.org/10.25077/TEKNOSI.v11i2.2025.185-191

4. KESIMPULAN

Penelitian ini berhasil merancang sebuah Sistem Pendukung Keputusan (SPK) untuk deteksi dini DBD, yang dibangun dengan memanfaatkan algoritma Naïve Bayes dan data hematologi sebagai dasar analisisnya.Model yang dibangun mencapai akurasi tinggi sebesar 98,03%, dengan performa precision dan recall yang sangat baik, terutama recall sempurna pada kasus risiko tinggi. Hasil ini membuktikan bahwa data hematologi sederhana dapat dimanfaatkan secara efektif untuk mempercepat diagnosis dini DBD, sehingga dapat membantu tenaga medis dalam pengambilan keputusan cepat dan mendukung pengendalian penyakit secara lebih luas.

DAFTAR PUSTAKA

[1] A. E. L. Putri, "Gambaran Kasus Demam Berdarah Dengue Puskesmas X Kota Malang Tahun 2019-2022," *Media Husada J. Environ. Heal. Sci.*, vol. 3, no. 1, pp. 12–18, 2023, doi: 10.33475/mhjeh.v3i1.38.

[2] WHO, "Dengue and severe dengue," World Health Organization. Accessed: Apr. 24, 2025. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue

[3] Kemenkes, "Strategi Nasional Penaggulangan Dengue 2021-2025," Waspada Penyakit di Musim Hujan. [Online]. Available: https://kemkes.go.id/id/waspada-penyakit-di-musim-hujan

[4] A. D. Putra, D. Nurani, M. M. Dewi, and S. Alfie Nur Rahmi, "Supervised Machine Learning Model untuk Prediksi Penyakit Hepatitis," *Indones. J. Comput. Sci.*, vol. 13, no. 2, pp. 3329–3341, 2024, [Online]. Available: http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

[5] H. F. Husniah and T. Arifin, "Implementasi Algoritma Naïve Bayes Berbasis Particle Swarm Optimization Untuk Memprediksi Penyakit Hepatitis," *J. Ilmu Komput.*, vol. 14, no. 1, p. 36, 2021, doi: 10.24843/jik.2021.v14.i01.p05.

[6] R. G. Gunawan and M. Ilham Pratama, "Analisa Kinerja Algoritma Machine Learning Untuk Prediksi Virus Hepatitis C," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 4, no. 3, pp. 772–777, 2024, doi: 10.37859/coscitech.v4i3.6513.

[7] L. Iryani, N. I. H. Kunio, and A. S. Wati, "Optimasi Metode Naïve Bayes Menggunakan Smoothing dan Feature Selection Untuk Penyakit Demam Berdarah Dengue," *J. Sci. Appl. Informatics*, vol. 7, no. 3, pp. 435–440, 2024.

[8] M. Jasri, A. Wijaya, and R. Sunggara, "Penerapan Data Mining untuk Klasifikasi Penyakit Demam Berdarah Dengue (DBD) Dengan Metode Naïve Bayes (Studi Kasus Puskesmas Taman Krocok)," *SMARTICS J.*, vol. 8, no. 1, pp. 35–42, 2022, [Online]. Available: https://doi.org/10.21067/smartics.v8i1.7375

[9] A. Wahyu Redhani and N. Hidayat, "Implementasi Metode Naïve Bayes untuk Diagnosa Pengidap Demam Berdarah pada Kelurahan Antasan Besar berbasis Web," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 12, pp. 5320–5328, 2021, [Online]. Available: http://j-ptiik.ub.ac.id

[10] A. Ikbal, A. Irma Purnamasari, and I. Ali, "Analisis Klasterisasi Untuk Prediksi Jumlah Kasus Dbd Berdasarkan Jenis Kelamin Dan Kabupaten/Kota Di Jawa Barat," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3789–3796, 2024, doi: 10.36040/jati.v7i6.8296.

[11] R. G. Wardhana, G. Wang, and F. Sibuea, "Penerapan Machine Learning Dalam Prediksi Tingkat Kasus Penyakit Di Indonesia," *J. Inf. Syst. Manag.*, vol. 5, no. 1, pp. 40–45, 2023, doi: 10.24076/joism.2023v5i1.1136.

[12] A. Khusaeri, "Analisis Algoritma K-Means Clustering Dalam Pengelompokan Daerah Penyebaran Penyakit Demam Berdarah Dengue," *J. Inf. Syst. Informatics Comput.*, vol. 8, no. 2, pp. 434–449, 2024, doi: 10.52362/jisicom.v8i2.1795.

[13] Amrin and P. Omar, "Implementasi Algoritma Klasifikasi Logistic Regression dan Naïve Bayes untuk Diagnosa Penyakit Hepatitis," *J. Tek. Komput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.

[14] M. Y. Matdoan, "Penerapan Metode K-Nearest Neighbor untuk Mengklasifikasi Penyebaran Kasus Demam Berdarah Dengue (DBD) di Kabupaten Maluku Tenggara," *Sq. J. Math. Math. Educ.*, vol. 4, no. 2, pp. 75–82, 2022, doi: 10.21580/square.2022.4.2.13056.

[15] A. Damayanti and G. Testiana, "Penerapan Data Mining untuk Prediksi Penyakit Hepatitis C Menggunakan Algoritma Naïve Bayes," *J. Manaj. Inform. Jayakarta*, vol. 3, no. 2, pp. 177–186, 2023, doi: 10.52362/jmijayakarta.v3i2.1098.

NOMENKLATUR

Persamaan 1 :

X : Data yang akan diklasifikasikan.
T : Hipotesis bahwa X termasuk kelas tertentu.
P(T|X) : Probabilitas T benar setelah melihat X (*posterior*).
P(T) : Probabilitas awal T sebelum melihat X (*prior*).
P(X|T) : Probabilitas X muncul jika T benar (*likelihood*).
P(X) : Probabilitas keseluruhan X dalam dataset.

BIODATA PENULIS

Yulia Nita lahir di Halat pada 27 Mei 1996. Merupakan mahasiswa Program Studi Magister Ilmu Komputer di Universitas Budi Luhur, Jakarta.

Maya Gian Sister lahir di Tanah Laut pada 1 Mei 2000. Merupakan mahasiswa Program Studi Magister Ilmu Komputer di Universitas Budi Luhur, Jakarta.

Dr. Ir. Gandung Triyono, S.Kom., M.Kom. Merupakan peneliti dan dosen di Fakultas Teknologi Informasi di Universitas Budi Luhur dengan konsentrasi bidang keilmuan Sistem Pendukung Keputusan.