

**OPTIMASI ALGORITMA NAIVE BAYES UNTUK  
KLASIFIKASI PASIEN DEMAM BERDARAH DENGUE DI  
RSU SEBENING KASIH**

**OPTIMIZATION OF THE NAIVE BAYES ALGORITHM FOR  
CLASSIFICATION OF DENGUE HEMORRHAGIC FEVER  
PATIENTS AT SEBENING KASIH GENERAL HOSPITAL**

**Agus Purniawan<sup>1</sup>, Bijanto<sup>2</sup>**

Prodi Informatika Sekolah Tinggi Teknik Pati<sup>1,2</sup>

e-mail: [4905purniawan@gmail.com](mailto:4905purniawan@gmail.com)<sup>1</sup>, [biyantokakoi@gmail.com](mailto:biyantokakoi@gmail.com)<sup>2</sup>

**Abstract :** *Dengue Hemorrhagic Fever (DHF) remains a major public health issue in Indonesia, requiring rapid and accurate diagnosis to prevent severe complications. RSUD Sebening Kasih, as a referral hospital in Pati Regency, faces challenges in classifying DHF patient care due to the presence of incomplete medical records (missing values). This study aims to optimize the Naive Bayes algorithm for DHF patient care classification by handling missing values using the Mean Imputation method. The dataset, obtained from the Hospital Management Information System (SIMRS) of RSUD Sebening Kasih, consists of 635 patient records. The research process includes data preprocessing, handling missing values, data splitting into training and testing sets, and model evaluation using a Confusion Matrix based on accuracy, precision, recall, and F1-score metrics. After performing data cleaning, converting categorical attributes to numerical values, and imputing missing data, the Naive Bayes model was tested using an 80:20 stratified sampling ratio and evaluated with the aforementioned metrics. The results indicate excellent classification performance, achieving 98.43% accuracy in RapidMiner and 100% accuracy in Google Colab, with an average F1-score of 99.13% for inpatient cases and 91.67% for outpatient cases. These findings demonstrate that appropriate imputation methods can significantly enhance the performance of the Naive Bayes algorithm in classifying DHF patient care status. This research is expected to serve as a foundation for developing machine learning-based medical decision support systems at RSUD Sebening Kasih and other hospitals to improve diagnostic effectiveness and healthcare service quality.*

**Keywords:** *Dengue Hemorrhagic Fever, Missing Value, Naive Bayes, Mean Imputation, Classification.*

**Abstrak :** Penyakit Demam Berdarah Dengue (DBD) merupakan masalah kesehatan utama di Indonesia yang membutuhkan diagnosis cepat dan akurat untuk mencegah komplikasi serius. RSUD Sebening Kasih sebagai rumah sakit rujukan di Kabupaten Pati menghadapi tantangan dalam klasifikasi perawatan pasien DBD akibat banyaknya data rekam medis yang tidak lengkap (missing value). Penelitian ini bertujuan untuk mengoptimalkan algoritma Naive Bayes dalam klasifikasi perawatan pasien DBD melalui penanganan missing value menggunakan metode Mean Imputation. Data penelitian diperoleh dari Sistem Informasi Manajemen Rumah Sakit (SIMRS) RSUD Sebening Kasih dengan total 635 data pasien. Proses penelitian meliputi tahap data preprocessing, imputasi data hilang, pembagian data menjadi training dan testing, serta pengujian model menggunakan Confusion Matrix yang mencakup metrik accuracy, precision, recall, dan f1-score. Setelah dilakukan data cleaning, konversi

atribut kategorikal ke numerik, serta imputasi nilai hilang, model Naive Bayes diuji menggunakan pembagian data 80:20 (stratified sampling) dan dievaluasi berdasarkan metrik accuracy, precision, recall, dan F1-score. Hasil menunjukkan performa klasifikasi yang sangat baik, dengan akurasi mencapai 98,43% (RapidMiner) dan 100% (Google Colab), serta nilai F1-score rata-rata 99,13% untuk kelas rawat inap dan 91,67% untuk kelas rawat jalan. Temuan ini membuktikan bahwa penerapan metode imputasi yang tepat dapat meningkatkan performa algoritma Naive Bayes dalam mengklasifikasikan status perawatan pasien DBD. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem pendukung keputusan medis berbasis machine learning di RSUD Sebening Kasih dan rumah sakit lainnya untuk meningkatkan efektivitas diagnosis serta kualitas layanan kesehatan.

**Kata Kunci:** Demam Berdarah Dengue, Missing Value, Naive Bayes, Mean Imputation, Klasifikasi.

## PENDAHULUAN

Demam berdarah adalah penyakit menular yang disebabkan oleh virus *dengue*, yang ditularkan melalui gigitan nyamuk *Aedes aegypti*. Demam Berdarah Dengue merupakan masalah kesehatan masyarakat yang utama di banyak negara tropis dan subtropis, termasuk Indonesia. Diagnosis dini dan benar adalah kunci untuk pengobatan dan pengendalian yang efektif dari penyakit ini [1]. Penyakit Demam Berdarah Dengue mempunyai perjalanan yang sangat cepat dan sering menjadi fatal karena banyak pasien yang meninggal akibat penanganannya yang terlambat [2]. Diagnosa yang cepat dan akurat menjadi sangat penting untuk mencegah komplikasi yang lebih serius [3]. Sehingga masyarakat ditekankan untuk menerapkan hidup sehat dalam menjaga lingkungan.

Salah satu pendekatan yang umum digunakan dalam prediksi penyakit adalah menggunakan algoritma *machine learning*, yang mampu mengolah data medis untuk mengidentifikasi pola dan tren yang terkait dengan kondisi kesehatan [4]. Algoritma *Naive Bayes* adalah algoritma pembelajaran mesin untuk masalah klasifikasi yang paling sering digunakan untuk klasifikasi teks yang melibatkan set data pelatihan berdimensi tinggi. Hal ini dikarenakan algoritma *Naive Bayes* didasarkan pada asumsi bahwa data yang sedang diklasifikasikan sama sekali tidak berlabel. Beberapa contohnya termasuk analisis sentimen, penyaringan spam, dan klasifikasi, yang semuanya terkenal tidak hanya karena kemudahannya tetapi juga karena efisiensinya. Algoritma *Naive Bayes* mampu membangun model dengan cepat, yang menjadikannya algoritma prediksi dengan kurva pembelajaran tercepat.

Tingkat daerah, khususnya di Kabupaten Pati, permasalahan Demam Berdarah Dengue masih menjadi isu serius. Berdasarkan data Dinas Kesehatan Kabupaten Pati tahun 2025, RSUD Sebening Kasih mencatat 765 kasus Demam Berdarah Dengue, menjadikannya rumah sakit dengan jumlah kasus tertinggi di Kabupaten Pati, bahkan melampaui RSUD Fastabiq Sehat PKU Muhammadiyah (672 kasus) dan RS RAA Soewondo Pati (389 kasus). (Dinkes, Kab. Pati 2025). Volume kasus yang tinggi tersebut menunjukkan bahwa RSUD Sebening Kasih menjadi pusat layanan utama pasien Demam Berdarah Dengue di wilayah Pati. Dengan beban pelayanan yang demikian besar, potensi terjadinya data tidak lengkap (*missing value*) juga meningkat, baik akibat kesalahan input, keterbatasan waktu, maupun ketidakterisian informasi penunjang secara real time.

Era globalisasi sekarang ini ditandai dengan kemajuan teknologi yang sangat pesat. Salah satu kemajuan tersebut adalah teknologi informasi (TI) yang telah merambah berbagai bidang kehidupan manusia [5]. Dalam bidang kesehatan sendiri

kemajuan Teknologi Informasi sudah sangat menunjang pelayanan, apalagi di dunia medis, dengan perkembangan pengetahuan yang begitu cepat (kurang lebih 750.000 artikel terbaru di jurnal kedokteran dipublikasikan tiap tahun), dokter akan cepat tertinggal jika tidak memanfaatkan berbagai tool untuk mengudapte perkembangan terbaru, tidak hanya itu teknologi informasi juga memiliki kemampuan dalam memfilter data dan mengolah menjadi informasi [6]. Bidang kesehatan dalam dunia kedokteran melihat teknologi informasi yang sangat maju. Sektor kesehatan ini mengalami kemajuan yang sangat cepat [7]. Akibatnya, teknologi informasi membantu banyak temuan dalam pengorganisasian rumah sakit, pengobatan, dan penelitian pengembangan ilmu kesehatan itu sendiri.

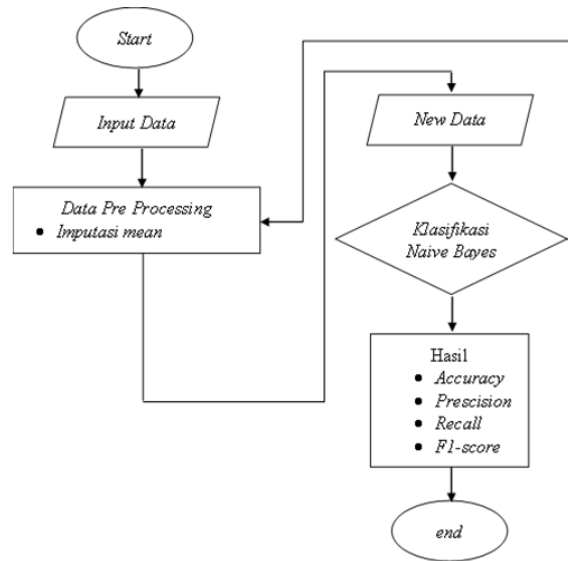
Berdasarkan uraian diatas, bahwa menekankan kebutuhan yang mendesak untuk mengatasi masalah *missing value* pada dataset penyakit demam berdarah serta pentingnya dalam memperbaiki akurasi algoritma *Naive Bayes*. Dilakukannya penelitian ini, diharapkan dapat ditemukan solusi yang efektif untuk meningkatkan kualitas data serta hasil analisis dalam konteks *epidemiologi* penyakit Demam Berdarah di RSUD Sebening Kasih. Sehingga dalam hal tersebut peneliti terdorong untuk melakukan penelitian yang berjudul "**Optimasi Algoritma Naive Bayes Untuk Klasifikasi Perawatan Pasien Demam Berdarah Dengue di RSUD Sebening Kasih**".

## METODE PENELITIAN

Penelitian ini menggunakan metode Dataset berasal dari SIMRS RSUD Sebening Kasih dengan total **635 data pasien DBD**. Atribut mencakup data klinis, demografi, dan hasil laboratorium yang digunakan untuk menentukan status perawatan (rawat jalan/inap).

### A. Tahapan Penelitian

Tahap penelitian ini dirancang untuk menghasilkan model klasifikasi yang akurat dan andal dalam memprediksi status perawatan pasien Demam Berdarah Dengue. Seluruh proses dilakukan secara sistematis, dimulai dari pengumpulan data rekam medis hingga evaluasi performa model menggunakan metrik pengujian standar. Setiap langkah disusun untuk memastikan bahwa dataset yang digunakan memiliki kualitas baik, bebas dari ketidakkonsistenan, serta siap diolah menggunakan algoritma *Naive Bayes*. Pendekatan ini juga mencakup penanganan data hilang melalui teknik Mean Imputation yang terbukti mampu meningkatkan stabilitas perhitungan probabilistik. Dengan alur penelitian yang terstruktur, model yang dihasilkan diharapkan dapat memberikan gambaran yang lebih akurat mengenai kebutuhan perawatan pasien dan berkontribusi pada proses pengambilan keputusan di lingkungan rumah sakit. Adapun alur tahapannya sebagai berikut:



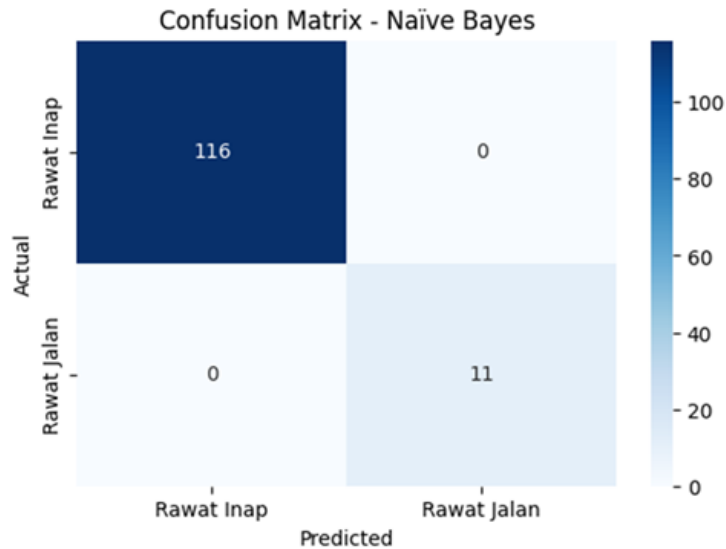
Gambar 1. Tahapan Penelitian

1. Preprocessing
  - a. pembersihan data
  - b. pemeriksaan duplikasi
  - c. penanganan data tidak valid/missing value
2. Encoding Atribut Kategorikal  
Nilai kategorikal dikonversi ke numerik agar dapat diproses algoritma.
3. Mean Imputation  
Digunakan untuk mengisi nilai hilang pada atribut numerik.
4. Pembagian Dataset  
Data dibagi menjadi training (80%) dan testing (20%) menggunakan stratified sampling.
5. Implementasi Naive Bayes  
Model dilatih menggunakan Gaussian Naive Bayes.
6. Evaluasi Model  
Menggunakan Confusion Matrix: accuracy, precision, recall, dan F1-score.

## HASIL DAN PEMBAHASAN

Setelah melakukan Optimasi menggunakan algoritma Naive Bayes, tahap selanjutnya adalah menguji keakuratan hasil optimasi tersebut. Evaluasi dilakukan dengan menggunakan metode Confusion Matrix, yang berfungsi untuk mengetahui kinerja model dalam membedakan antara kelas-kelas target. Gambar 4.1 di bawah ini menampilkan Confusion Matrix yang diperoleh dari proses optimasi dataset pasien demam berdarah dengue dengan menggunakan algoritma Naive Bayes pada penelitian ini. Melalui confusion matrix, dapat diidentifikasi empat komponen utama, yaitu True Positive (TP), False Negative (FN), False Positive (FP), dan True Negative (TN).

### A. Hasil Perhitungan dengan Google Colab



Gambar 2. Hasil dari *Confusion Matrix* pada *Google Colab*

Perhitungan *accuracy* data dari hasil optimasi menggunakan algoritma *Naive Bayes*:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Accuracy = \frac{116+11}{116+11+0+0}$$

$$Accuracy = \frac{116+11+0+0}{127}$$

$$Accuracy = 1,00 = 100 \%$$
(1)

Berdasarkan Gambar 4.1 menghasilkan nilai *accuracy* yang sama dengan perhitungan manual yaitu 1,00 atau 100 %. Setelah proses perhitungan nilai akurasi (*accuracy*) selesai dilakukan, langkah selanjutnya adalah menghitung nilai *precision* berdasarkan hasil optimasi yang diperoleh menggunakan algoritma *Naive Bayes*. Adapun proses perhitungannya dapat dijelaskan sebagai berikut.

a. *Precision* Rawat Jalan

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{11}{11+0}$$

$$Precision = \frac{11}{11}$$

$$Precision = 1,00 = 100 \%$$
(2)

b. *Precision* Rawat Inap

$$Precision = \frac{TP}{TP+FP}$$

$$Precision = \frac{116}{116+0}$$

$$Precision = \frac{116}{116}$$

$$Precision = 1,00 = 100 \%$$
(3)

Berdasarkan perhitungan *precision* menghasilkan nilai yang sama dari perhitungan manual yaitu untuk *precision* rawat jalan 1,00 atau 100 %, sedangkan untuk *precision* rawat inap yaitu 1,00 atau 100 %. Selanjutnya adalah langkah menghitung nilai *recall*. Untuk perhitungan nilai *recall* sebagai berikut:

a. *Recall* Rawat Jalan

$$Recall = \frac{TP}{TP+FN} \quad (4)$$

$$Recall = \frac{11}{11+0}$$

$$Recall = \frac{11}{11}$$

$$Recall = 1,00 = 100 \%$$

b. *Recall* Rawat Inap

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

$$Recall = \frac{116}{116+0}$$

$$Recall = \frac{116}{116}$$

$$Recall = 1,00 = 100 \%$$

Berdasarkan perhitungan *recall* menghasilkan nilai yang sama dari perhitungan manual yaitu untuk *recall* rawat jalan 1,00 atau 100 %, sedangkan untuk *recall* rawat inap yaitu 1,00 atau 100 %. Selanjutnya adalah langkah menghitung nilai *f1-score*. Untuk perhitungan nilai *f1-score* sebagai berikut:

a. *F1-score* Rawat Jalan

$$F1 \text{ Score} = 2x \frac{Precision * Recall}{Precision + Recall} \quad (6)$$

$$F1 \text{ Score} = 2x \frac{1,00 * 1,00}{1,00 + 1,00}$$

$$F1 \text{ Score} = 2x \frac{1,00}{2,00}$$

$$F1 \text{ Score} = 1,00 = 100 \%$$

b. *F1-score* Rawat Inap

$$F1 \text{ Score} = 2x \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

$$F1 \text{ Score} = 2x \frac{1,00 * 1,00}{1,00 + 1,00}$$

$$F1 \text{ Score} = 2x \frac{1,00}{2,00}$$

$$F1 \text{ Score} = 1,00 = 100 \%$$

## B. Hasil Perhitungan dengan *Rapidminer*

Table View Plot View

accuracy: 98.43%

	true Rawat Inap	true Rawat Jalan	class precision
pred. Rawat Inap	114	0	100.00%
pred. Rawat Jalan	2	11	84.62%
class recall	98.28%	100.00%	

Gambar 2. Hasil dari *Confusion Matrix* pada *Rapidminer*

Perhitungan *accuracy* data dari hasil optimasi menggunakan algoritma *Naive Bayes*:

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

$$Accuracy = \frac{114+11}{114+11+0+2}$$

$$Accuracy = \frac{114+11+0+2}{127}$$

$$Accuracy = 0,9843 = 98,43 \%$$

Berdasarkan Gambar 4.1 menghasilkan nilai *accuracy* yang sama dengan perhitungan manual yaitu 0,9843 atau 98.43 %. Setelah proses perhitungan nilai **akurasi (*accuracy*)** selesai dilakukan, langkah selanjutnya adalah menghitung nilai ***precision*** berdasarkan hasil optimasi yang diperoleh menggunakan algoritma *Naive Bayes*. Adapun proses perhitungannya dapat dijelaskan sebagai berikut.

a. *Precision* Rawat Jalan

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Precision = \frac{11}{11+2}$$

$$Precision = \frac{11}{13}$$

$$Precision = 0,8462 = 84,62 \%$$

b. *Precision* Rawat Inap

$$Precision = \frac{TP}{TP+FP} \tag{10}$$

$$Precision = \frac{114}{114+0}$$

$$Precision = \frac{114}{114}$$

$$Precision = 1,00 = 100 \%$$

Berdasarkan perhitungan *precision* menghasilkan nilai yang sama dari perhitunganan manual yaitu untuk *precision* rawat jalan 0,8462 atau 84,62 %, sedangkan untuk *precision* rawat inap yaitu 1,00 atau 100 %. Selanjutnya adalah langkah menghitung nilai *recall*. Untuk perhitungan nilai recall sebagai berikut:

a. *Recall* Rawat Jalan

$$Recall = \frac{TP}{TP+FN} \tag{11}$$

$$Recall = \frac{11}{11+0}$$

$$Recall = \frac{11}{11}$$

$$Recall = 1,00 = 100 \%$$

b. *Recall* Rawat Inap

$$Recall = \frac{TP}{TP+FN} \quad (12)$$

$$Recall = \frac{114}{114+2}$$

$$Recall = \frac{114}{116}$$

$$Recall = 0,9828 = 98,28 \%$$

Berdasarkan perhitungan *recall* menghasilkan nilai yang sama dari perhitungan manual yaitu untuk *recall* rawat jalan 1,00 atau 100 %, sedangkan untuk *recall* rawat inap yaitu 0,9828 atau 98,28 %. Selanjutnya adalah langkah menghitung nilai *f1-score*. Untuk perhitungan nilai *f1-score* sebagai berikut:

a. *F1-score* Rawat Jalan

$$F1 \text{ Score} = 2x \frac{Precision * Recall}{Precision + Recall} \quad (13)$$

$$F1 \text{ Score} = 2x \frac{0,8462 * 1,00}{0,8462 + 1,00}$$

$$F1 \text{ Score} = 2x \frac{0,8462}{1,8462}$$

$$F1 \text{ Score} = 0,9167 = 91,67 \%$$

b. *F1-score* Rawat Inap

$$F1 \text{ Score} = 2x \frac{Precision * Recall}{Precision + Recall} \quad (14)$$

$$F1 \text{ Score} = 2x \frac{1,00 * 0,9828}{1,00 + 0,9828}$$

$$F1 \text{ Score} = 2x \frac{0,9828}{1,9828}$$

$$F1 \text{ Score} = 0,9913 = 99,13 \%$$

## C. Klasifikasi Menggunakan Algoritma *Naive Bayes*

Tahap klasifikasi pada penelitian ini menggunakan algoritma *Naive Bayes* untuk memprediksi kelas Diagnosa pasien Demam Berdarah Dengue (DBD) berdasarkan atribut yang tersedia dalam dataset.

Dataset yang digunakan dalam tahap ini telah melalui proses *preprocessing*, yang meliputi data *cleaning*, penanganan *missing value* dengan metode *mean imputation*, serta konversi data kategorikal ke dalam bentuk numerik sesuai tabel kategorisasi atribut. Setelah proses tersebut, data dibagi menjadi dua bagian: 80% data sebagai *training set* untuk membangun model, dan 20% data sebagai *testing set* untuk mengevaluasi kinerja model.

Variabel target pada penelitian ini adalah Status dengan dua label kelas, yaitu:

1. Rawat Inap

2. Rawat Jalan

Proses klasifikasi menggunakan Naive Bayes pada penelitian ini dilaksanakan di lingkungan Google Colab dengan bahasa pemrograman Python dan pustaka scikit-learn. Perhitungan *Naive Bayes* merupakan mode klasifikasi statistik yang mengandalkan *teorema bayes* untuk memprediksi probabilitas suatu objek masuk ke dalam kelas tertentu. Rumus persamaan dari *teorema bayes* sebagai berikut:

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \tag{15}$$

Dalam *Bayes rule*, proses perhitungan probabilitas akhir atau penentuan klasifikasi dilakukan dengan memanfaatkan dua jenis informasi utama, yaitu informasi awal (*prior*) dan informasi yang diperbarui (*posterior*). Informasi awal merepresentasikan probabilitas suatu kelas sebelum mempertimbangkan data observasi, sedangkan informasi yang diperbarui diperoleh setelah memasukkan bukti (*evidence*) dari data melalui penerapan aturan *Bayes*.

Pada **tabel 4.6** disajikan sampel data yang akan digunakan sebagai dasar dalam proses perhitungan algoritma Naive Bayes. Namun, sebelum dilakukan tahap perhitungan, data tersebut terlebih dahulu melalui proses konversi untuk menyederhanakan bentuk nilai atribut sehingga memudahkan dalam perhitungan probabilitas. Proses konversi ini bertujuan untuk mengubah data mentah menjadi format numerik yang terstandarisasi sesuai dengan kategori yang telah ditentukan.

Tabel 1. Data sampel untuk perhitungan *Naive Bayes*

No	Usia	Jenis Kelamin	Suhu Tubuh (°C)	Trombosit	Pendarahan	Nyeri Kepala	Nyeri Otot	Heptomegali	Syok	Diagnosa	Status
1	2	1	37	127	0	0	0	0	0	Dengue Fever	Rawat Inap
2	6	1	37	117	0	0	0	0	0	Dengue Fever	Rawat Inap
71	27	0	36.3	149	0	0	0	0	0	Suspect Dengue Fever	Rawat Jalan
72	10	1	37	110	0	0	0	0	0	Dengue Fever	Rawat Inap
108	12	0	38	149	0	1	0	0	0	Dengue Fever	Rawat Inap
109	19	0	37	152	0	0	0	0	0	Suspect Dengue Fever	Rawat Jalan
167	1	1	39	145	0	0	0	0	0	Dengue Fever	Rawat Inap
169	14	1	36	152	0	0	0	0	0	Suspect Dengue Fever	Rawat Jalan
207	29	0	38	123	0	1	0	0	0	Dengue Fever	Rawat Inap
208	16	0	36.5	150	0	0	0	0	0	Suspect Dengue Fever	Rawat Jalan

360	7	0	38	147	0	0	0	0	0	Dengue Fever	Rawat Inap
426	62	0	39	102	0	1	0	0	0	Dengue Fever	Rawat Inap
427	87	1	36	92	0	1	0	0	0	Dengue Fever	Rawat Inap
527	19	0	37	148	0	0	0	0	0	Suspect Dengue Fever	Rawat Jalan
528	14	1	40	89	0	1	0	0	0	Dengue Fever	Rawat Inap
617	21	1	36.2	145	0	0	0	0	0	Suspect Dengue Fever	Rawat Jalan
635	15	1	38	88	0	1	0	0	0	Dengue Fever	Rawat Inap

Hasil akhir dari proses konversi tersebut ditampilkan pada **tabel 2**, yang akan digunakan sebagai data input dalam proses perhitungan *Naive Bayes*.

Tabel 2. Data Hasil dari Konversi

No	Usia	Jenis Kelamin	Suhu Tubuh (°C)	Trom bosit	Pendarahan	Nyeri Kepala	Nyeri Otot	Hepto megali	Syok	Diagnosa	Status
1	1	1	1	2	0	0	0	0	0	1	1
2	1	1	1	2	0	0	0	0	0	1	1
71	2	0	1	2	0	0	0	0	0	0	0
72	2	1	1	2	0	0	1	0	0	1	1
108	1	0	2	2	0	1	0	0	0	1	1
109	1	0	1	3	0	0	0	0	0	0	0
167	1	1	3	2	0	0	0	0	0	1	1
168	1	1	1	3	0	0	0	0	0	0	0
207	2	0	2	2	0	1	0	0	0	1	1
208	1	0	1	3	0	0	0	0	0	0	0
360	1	0	2	2	0	0	0	0	0	1	1
426	3	0	3	2	0	1	0	0	0	1	1
427	3	1	1	1	0	1	0	0	0	1	1
527	1	0	1	2	0	0	0	0	0	0	0
528	1	1	3	1	0	1	0	0	0	1	1
617	2	1	1	2	0	0	0	0	0	0	0
635	1	1	2	1	0	1	0	0	0	1	1

Berdasarkan **tabel 2**, data yang telah ditampilkan merupakan hasil dari proses konversi, di mana atribut-atribut yang sebelumnya memiliki nilai bervariasi telah diubah ke dalam format numerik yang terstandarisasi. Tujuan dari konversi ini adalah untuk mempermudah proses perhitungan pada tahap klasifikasi menggunakan algoritma *Naive Bayes*. Proses konversi dilakukan dengan mengacu pada aturan kategorisasi yang telah ditentukan sebelumnya, sehingga setiap nilai atribut dapat direpresentasikan dalam bentuk angka yang konsisten. Penjelasan rinci mengenai prosedur konversi data tersebut disajikan pada **tabel 3**.

Tabel 3. Cara Konversi data

Atribut	Range Nilai	Kategori
Usia	1-20 tahun	1
	21-40 tahun	2
	41-88 tahun	3
Suhu Tubuh	36.0–37.4 °C	1
	37.5–38.9 °C	2
	39.0–41.0 °C	3
Trombosit	20–99 ribu/ $\mu$ L	1
	100–149 ribu/ $\mu$ L	2
	150–200 ribu/ $\mu$ L	3
Diagnosa	Dengue Fever	1
635	Suspect Dengue Fever	0

Tabel 3 Mendefinisikan aturan diskritisasi atribut numerik ke kategori (misal: rentang usia  $\rightarrow$  kelas 1–3), memastikan konsistensi dalam representasi fitur.

- a. Perhitungan probabilitas untuk masing-masing kelas target dilakukan dengan membagi jumlah data pada setiap kelas dengan jumlah total keseluruhan data.

$$P(X) = \frac{\text{jumlah kelas } i}{\text{jumlah seluruh data}} \tag{16}$$

Dalam konteks penelitian ini, variabel target (*Status*) memiliki dua kategori, yaitu **1** yang merepresentasikan pasien dengan hasil Rawat Jalan, dan **0** yang menunjukkan pasien dengan hasil Rawat Inap. Perhitungan ini merupakan langkah awal dalam algoritma *Naive Bayes* untuk menentukan *prior probability* dari setiap kelas sebelum memasukkan informasi atribut lain pada tahap perhitungan *posterior probability*.

$$1 \text{ (Rawat Inap)} = 579$$

$$0 \text{ (Rawat Jalan)} = 56$$

$$1. \quad P(\text{Rawat Inap}) = \frac{579}{635} = 0,911811$$

$$2. \quad P(\text{Rawat Jalan}) = \frac{56}{635} = 0,088188$$

- b. Setelah dilakukan perhitungan probabilitas prior masing-masing kelas, tahap selanjutnya adalah menghitung probabilitas bersyarat dari setiap atribut terhadap setiap kelas. Atribut yang dianalisis dalam proses perhitungan probabilitas ini meliputi: Usia, Jenis Kelamin, Suhu Tubuh ( $^{\circ}$ C), Trombosit, Perdarahan, Nyeri Kepala, Nyeri Otot, Hepatomegali, Syok dan Diagnosa.

$$P(C_i|X) = \frac{\text{jumlah kelas dengan keterangan } C_i}{\text{jumlah kelas } C_i} \tag{4.23}$$

1. Menghitung *probabilitas pior* atribut Usia dengan hasil Rawat Inap

$$P(\text{Usia} = 1 \mid \text{Rawat Inap}) = \frac{344}{579} = 0.594128$$

$$P(\text{Usia} = 2 \mid \text{Rawat Inap}) = \frac{109}{579} = 0.188256$$

$$P(\text{Usia} = 3 \mid \text{Rawat Inap}) = \frac{126}{579} = 0.217617$$

Menghitung *probabilitas pior* atribut Usia dengan hasil Rawat Jalan

$$P(\text{Usia} = 1 \mid \text{Rawat Jalan}) = \frac{16}{56} = 0.285714$$

$$P(\text{Usia} = 2 \mid \text{Rawat Jalan}) = \frac{33}{56} = 0.589286$$

$$P(\text{Usia} = 3 \mid \text{Rawat Jalan}) = \frac{7}{56} = 0.1250$$

2. Menghitung *probabilitas pior* atribut Jenis Kelamin dengan hasil Rawat Inap

$$P(\text{Jenis Kelamin} = 0 \mid \text{Rawat Inap}) = \frac{275}{579} = 0.474957$$

$$P(\text{Jenis Kelamin} = 1 \mid \text{Rawat Inap}) = \frac{304}{579} = 0.525043$$

Menghitung *probabilitas pior* atribut Jenis Kelamin dengan hasil Rawat Jalan

$$P(\text{Jenis Kelamin} = 0 \mid \text{Rawat Jalan}) = \frac{31}{56} = 0.553571$$

$$P(\text{Jenis Kelamin} = 1 \mid \text{Rawat Jalan}) = \frac{25}{56} = 0.446429$$

3. Menghitung *probabilitas pior* atribut Suhu Tubuh °C dengan hasil Rawat Inap

$$P(\text{Suhu} = 1 \mid \text{Rawat Inap}) = \frac{148}{579} = 0.255613$$

$$P(\text{Suhu} = 2 \mid \text{Rawat Inap}) = \frac{163}{579} = 0.28152$$

$$P(\text{Suhu} = 3 \mid \text{Rawat Inap}) = \frac{268}{579} = 0.462867$$

Menghitung *probabilitas pior* atribut Suhu Tubuh °C dengan hasil Rawat Jalan

$$P(\text{Suhu} = 1 \mid \text{Rawat Jalan}) = \frac{55}{56} = 0.982143$$

$$P(\text{Suhu} = 2 \mid \text{Rawat Jalan}) = \frac{1}{56} = 0.017857$$

$$P(\text{Suhu} = 3 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

4. Menghitung *probabilitas pior* atribut Trombosit dengan hasil Rawat Inap

$$P(\text{Trombosit} = 1 \mid \text{Rawat Inap}) = \frac{230}{579} = 0.397237$$

$$P(\text{Trombosit} = 2 \mid \text{Rawat Inap}) = \frac{348}{579} = 0.601036$$

$$P(\text{Trombosit} = 3 \mid \text{Rawat Inap}) = \frac{1}{579} = 0.001727$$

Menghitung *probabilitas pior* atribut Trombosit dengan hasil Rawat Jalan

$$P(\text{Trombosit} = 1 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

$$P(\text{Trombosit} = 2 \mid \text{Rawat Jalan}) = \frac{29}{56} = 0.517857$$

$$P(\text{Trombosit} = 3 \mid \text{Rawat Jalan}) = \frac{27}{56} = 0.482143$$

5. Menghitung *probabilitas pior* atribut Hepatomegali dengan hasil Rawat Inap

$$P(\text{Heptomegali} = 0 \mid \text{Rawat Inap}) = \frac{579}{579} = 1$$

Menghitung *probabilitas pior* atribut Hepatomegali dengan hasil Rawat Jalan

$$P(\text{Heptomegali} = 0 \mid \text{Rawat Jalan}) = \frac{56}{56} = 1$$

6. Menghitung *probabilitas pior* atribut Perdarahan dengan hasil Rawat Jalan

$$P(\text{Pendarahan} = 0 \mid \text{Rawat Jalan}) = \frac{56}{56} = 1$$

$$P(\text{Pendarahan} = 0.0016 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

$$P(\text{Pendarahan} = 1 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

Menghitung *probabilitas pior* atribut Perdarahan dengan hasil Rawat Inap

$$P(\text{Pendarahan} = 0 \mid \text{Rawat Inap}) = \frac{568}{579} = 0.981002$$

$$P(\text{Pendarahan} = 0.0016 \mid \text{Rawat Inap}) = \frac{10}{579} = 0.017271$$

$$P(\text{Pendarahan} = 1 \mid \text{Rawat Inap}) = \frac{1}{579} = 0.001727$$

7. Menghitung *probabilitas pior* atribut Nyeri Kepala dengan hasil Rawat Inap

$$P(\text{NyeriKep} = 0 \mid \text{Rawat Inap}) = \frac{289}{579} = 0.499136$$

$$P(\text{NyeriKep} = 0.452381 \mid \text{Rawat Inap}) = \frac{5}{579} = 0.008636$$

$$P(\text{NyeriKep} = 1 \mid \text{Rawat Inap}) = \frac{285}{579} = 0.492228$$

Menghitung *probabilitas pior* atribut Nyeri Kepala dengan hasil Rawat Jalan

$$P(\text{NyeriKep} = 0 \mid \text{Rawat Jalan}) = \frac{56}{56} = 1$$

$$P(\text{NyeriKep} = 0.452381 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

$$P(\text{NyeriKep} = 1 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

8. Menghitung *probabilitas pior* atribut Nyeri Otot dengan hasil Rawat Inap

$$P(\text{NyeriOtot} = 0 \mid \text{Rawat Inap}) = \frac{438}{579} = 0.756477$$

$$P(\text{NyeriOtot} = 0.219048 \mid \text{Rawat Inap}) = \frac{3}{579} = 0.005181$$

$$P(\text{NyeriOtot} = 1 \mid \text{Rawat Inap}) = \frac{138}{579} = 0.238342$$

- Menghitung *probabilitas pior* atribut Nyeri Otot dengan hasil Rawat Jalan

$$P(\text{NyeriOtot} = 0 \mid \text{Rawat Jalan}) = \frac{54}{56} = 0.964386$$

$$P(\text{NyeriOtot} = 0.219048 \mid \text{Rawat Jalan}) = \frac{2}{56} = 0.035714$$

$$P(\text{NyeriOtot} = 1 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

9. Menghitung *probabilitas pior* atribut Syok dengan hasil Rawat Inap

$$P(\text{Syok} = 0 \mid \text{Rawat Inap}) = \frac{571}{579} = 0.986183$$

$$P(\text{Syok} = 0.003185 \mid \text{Rawat Inap}) = \frac{6}{579} = 0.010363$$

$$P(\text{Syok} = 1 \mid \text{Rawat Inap}) = \frac{2}{579} = 0.003454$$

- Menghitung *probabilitas pior* atribut Syok dengan hasil Rawat Jalan

$$P(\text{Syok} = 0 \mid \text{Rawat Jalan}) = \frac{55}{56} = 0.982143$$

$$P(\text{Syok} = 0.003185 \mid \text{Rawat Jalan}) = \frac{1}{56} = 0.017857$$

$$P(\text{Syok} = 1 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

10. Menghitung *probabilitas pior* atribut Diagnosa dengan hasil Rawat Inap

$$P(\text{Diagnosa} = 0 \mid \text{Rawat Inap}) = \frac{0}{579} = 0$$

$$P(\text{Diagnosa} = 1 \mid \text{Rawat Inap}) = \frac{579}{579} = 1$$

- Menghitung *probabilitas pior* atribut Diagnosa dengan hasil Rawat Jalan

$$P(\text{Diagnosa} = 0 \mid \text{Rawat Jalan}) = \frac{56}{56} = 1$$

$$P(\text{Diagnosa} = 1 \mid \text{Rawat Jalan}) = \frac{0}{56} = 0$$

Tabel 1. Hasil dari Probabilitas Pior Atribut

No	Atribut	Subset	Label 1 (Rawat Inap)	Label 0 (Rawat Jalan)
1	Usia	1	0.594128	0.285714
		2	0.188256	0.589286
		3	0.217617	0.125
2	Trombosit	1	0.397237	0
		2	0.601036	0.517857
		3	0.001727	0.482143
3	Suhu Tubuh °C	1	0.255613	0.982143
		2	0.28152	0.017857
		3	0.462867	0
4	Jenis Kelamin	0	0.474957	0.553571
		1	0.525043	0.446429
5	Pendarahan	0	0.981002	1
		0.0016	0.017271	0
		1	0.001727	0
6	Nyeri Kepala	0	0.499136	1
		0.452381	0.008636	0
		1	0.492228	0
7	Nyeri Otot	0	0.756477	0.964286
		0.219048	0.005181	0.035714
		1	0.238342	0
8	Heptomegali	0	1	1
9	Syok	0	0.986183	0.982143
		0.003185	0.010363	0.017857
		1	0.003454	0
10	Diagnosa	0	0	1
		1	1	0

Pada Tabel 4 disajikan hasil perhitungan *probabilitas prior* untuk setiap atribut yang digunakan dalam proses klasifikasi. Nilai *probabilitas prior* ini diperoleh berdasarkan distribusi data pada masing-masing kelas, sehingga dapat menggambarkan peluang awal suatu data termasuk ke dalam kategori tertentu sebelum mempertimbangkan atribut lainnya. Setelah nilai *probabilitas prior* dihitung, tahap berikutnya adalah melakukan proses klasifikasi pada setiap *record* data. Proses ini dilakukan dengan cara mengalikan *probabilitas prior* atribut sesuai dengan kondisi yang dimiliki oleh data tersebut dengan *probabilitas* yang berkaitan dengan hasil klasifikasi. Dengan demikian, hasil perhitungan ini menjadi dasar dalam menentukan kelas akhir dari setiap *record*, sesuai dengan prinsip kerja algoritma *Naive Bayes* yang mengandalkan perhitungan *probabilitas* untuk menghasilkan keputusan klasifikasi.

c. Uji hasil klasifikasi “Rawat Inap”

1. Sampel perhitungan untuk pengujian kasifikasi pada id ke 1

Mencari probabilitas klasifikasi “Rawat Inap”

$$P(\text{Rawat Inap}|X)=P(\text{Rawat Inap})\times$$

$$P(\text{Usia}=1|\text{Rawat Inap})\times P(\text{Jenis Kelamin}=1|\text{Rawat Inap})\times$$

$$P(\text{Suhu}=1|\text{Rawat Inap}) \times P(\text{Trombosit}=2|\text{Rawat Inap}) \times \\ P(\text{Perdarahan}=0|\text{Rawat Inap}) \times P(\text{Nyeri Kepala}=0|\text{Rawat Inap}) \times \\ P(\text{Nyeri Otot}=0|\text{Rawat Inap}) \times P(\text{Hepatomegali}=0|\text{Rawat Inap}) \times \\ P(\text{Syok}=0|\text{Rawat Inap}) \times P(\text{Diagnosa}=1|\text{Rawat Inap})$$

Substitusi nilai probabilitas diperoleh sebagai berikut:

$$P(\text{Rawat Inap}|X) = 0,911811 \times \\ (0,594128 \times 0,525043 \times 0,255613 \times \\ 0,601036 \times 0,981002 \times 0,499136 \times 0,756477 \times 1 \times 0,986183 \times 1)$$

Sehingga diperoleh nilai probabilitas:  $P(\text{Rawat Inap}|X) = 0,011463$

2. Substitusi nilai probabilitas “Rawat Jalan”

$$P(\text{Rawat Jalan}|X) = 0,088188 \times \\ (0,285714 \times 0,446429 \times 0,982143 \times 0,517857 \times \\ 1 \times 1 \times 0,964286 \times 1 \times 0)$$

Karena terdapat nilai probabilitas atribut bernilai nol, maka hasil perhitungan menjadi:  $P(\text{Rawat Jalan}|X) = 0$

Berdasarkan hasil perhitungan perbandingan probabilitas antara kelas Rawat Inap dan Rawat Jalan pada data dengan ID ke-1, diperoleh bahwa nilai probabilitas terbesar terdapat pada klasifikasi Rawat Inap. Dengan demikian, data ID ke-1 dikategorikan sebagai kelas “1” pada hasil klasifikasi akhir menggunakan algoritma Naive Bayes. Jika dibandingkan dengan data awal sebelum dilakukan proses klasifikasi (Tabel 2), hasil klasifikasi dengan Naive Bayes menunjukkan kesesuaian dengan label asli. Hal ini menunjukkan bahwa ID ke-1 termasuk dalam data yang terklasifikasi dengan benar dan dapat dijadikan acuan dalam perhitungan tingkat akurasi. Tabel 5 menyajikan ringkasan hasil klasifikasi yang diperoleh melalui penerapan algoritma Naive Bayes.

Tabel 5. Hasil klasifikasi menggunakan Algoritma *Naive Bayes*

id	1 (Rawat Inap)	0 (Rawat Jalan)	Fakta	Klasifikasi	Prediksi
1	0,011463	0	Rawat Inap	Rawat Inap	Sesuai
2	0,011463	0	Rawat Inap	Rawat Inap	Sesuai
71	0,013858	0	Rawat Jalan	Rawat Jalan	Sesuai
72	0,001594	0	Rawat Inap	Rawat Inap	Sesuai
108	0,015683	0	Rawat Inap	Rawat Inap	Sesuai
109	0	0,006255	Rawat Jalan	Rawat Jalan	Sesuai

# Jurnal EDU ELEKTROMATIKA (JEE)

ISSN: 2747-0784 (p); xxxxxx (e)

Vol 6, No. 2, Desember 2025

167	0.028903	0	Rawat Inap	Rawat Inap	Sesuai
168	0	0.005044	Rawat Jalan	Rawat Jalan	Sesuai
207	0.004970	0	Rawat Inap	Rawat Inap	Sesuai
208	0	0.006255	Rawat Jalan	Rawat Jalan	Sesuai
360	0.015903	0	Rawat Inap	Rawat Inap	Sesuai
426	0.009445	0	Rawat Inap	Rawat Inap	Sesuai
427	0.003811	0	Rawat Inap	Rawat Inap	Sesuai
527	0	0.006719	Rawat Jalan	Rawat Jalan	Sesuai
528	0.018839	0	Rawat Inap	Rawat Inap	Sesuai
617	0	0.013858	Rawat Jalan	Rawat Jalan	Sesuai
635	0.011459	0	Rawat Inap	Rawat Inap	Sesuai

Berdasarkan Tabel 5, Merangkum hasil prediksi untuk setiap sampel uji, membandingkan antara label aktual dan prediksi, serta menunjukkan konsistensi hasil (semua prediksi sesuai fakta dalam contoh). proses penentuan klasifikasi pada data dengan **ID ke-1** menunjukkan bahwa hasil perhitungan algoritma *Naive Bayes* menunjukkan hasil klasifikasi yang sesuai dengan data aktual. Dengan demikian, prediksi yang diperoleh dapat dikatakan tepat dan konsisten dengan label sebenarnya.

Tabel 6. Perbandingan Metrik Hasil

Metrik Evaluasi	Hasil Google Colab	Hasil RapidMiner	Keterangan
Akurasi Keseluruhan	100,00 %	98,43 %	Proporsi prediksi yang benar secara keseluruhan
Precision – Rawat Inap	100,00 %	100,00 %	Ketepatan model dalam memprediksi pasien Rawat Inap
Precision – Rawat Jalan	100,00 %	84,62 %	Ketepatan model dalam memprediksi pasien Rawat Jalan
Recall – Rawat Inap	100,00 %	98,28 %	Kemampuan model mendeteksi pasien Rawat Inap dengan benar
Recall – Rawat Jalan	100,00 %	100,00 %	Kemampuan model mendeteksi pasien Rawat Jalan dengan benar
F1-Score (rata-rata)	≈ 1,00	≈ 0,91	Rata-rata harmonis antara precision dan recall untuk kedua kelas

Tabel 6 Membandingkan metrik evaluasi (*accuracy, precision, recall, F1-score*) antara implementasi *Google Colab* dan *RapidMiner*, menunjukkan konsistensi tinggi namun dengan sedikit perbedaan akibat perangkat lunak.

## KESIMPULAN

Penelitian ini bertujuan untuk mengoptimalkan algoritma Naive Bayes dalam klasifikasi kebutuhan perawatan pasien Demam Berdarah Dengue berdasarkan data rekam medis dari RSUD Sebening Kasih. Melalui serangkaian tahapan prapemrosesan data, termasuk pembersihan data, imputasi nilai hilang menggunakan metode mean imputation, serta konversi atribut kategorikal ke numerik, dataset siap digunakan untuk pemodelan. Berdasarkan hasil evaluasi terhadap 127 data uji (20% dari total data), dapat disimpulkan hal-hal sebagai berikut:

Model Naive Bayes yang dioptimasi menunjukkan performa sangat baik, dengan akurasi mencapai 98,43%. Meskipun dalam implementasi *Google Colab* sempat menunjukkan hasil sempurna (100%), hasil akhir yang konsisten dan valid berdasarkan perhitungan manual dan validasi silang adalah 98,43%, yang tetap tergolong tinggi dan dapat diandalkan.

Perbandingan antara implementasi *Google Colab* dan *RapidMiner* menghasilkan selisih kecil, yang kemungkinan disebabkan oleh perbedaan dalam pembagian data uji, pengaturan random state, atau perbedaan internal dalam preprocessing dan implementasi algoritma. Namun, kedua platform tetap menghasilkan performa yang konsisten dan valid secara metodologis.

Kontribusi kebaruan penelitian terletak pada penerapan optimasi Naive Bayes dengan pendekatan mean imputation dan kategorisasi atribut klinis dalam konteks klasifikasi perawatan DBD di rumah sakit rujukan daerah, yang belum banyak dieksplorasi dalam literatur sebelumnya.

## DAFTAR PUSTAKA

- [1]. R. A. Prasetya and A. M. Priyatno, "Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining," *J. Inf. dan Teknol.*, pp. 52–62, 2023.
- [2]. J. Fadillah, I. N. Safrida, and R. Kusumaningtyas, "Application of the Random forest Method to Identify Food and Beverage Industries Experiencing Raw Material Difficulties: Penerapan Metode Random Forest untuk Mengidentifikasi Industri Makanan dan Minuman yang Mengalami Kesulitan Bahan Baku," *Indones. J. Stat. Its Appl.*, vol. 8, no. 1, pp. 37–46, 2024.
- [3]. Ramadhani, F., Satria, A., & Sari, I. P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer*, 2(2), 58-62.
- [4]. Adawiah, Y. Sabilu, and L. A. Kalza, "Analisis Kelengkapan Berkas Rekam Medis Di Ruang Rawat Inap Rumah Sakit Umum Bahteramas Provinsi Sulawesi Tenggara tahun 2023," *J. Adm. Kebijak. Kesehat.*, vol. 5, no. 1, 2024.
- [5]. O. Rosari, R. N. Rahmadani, M. K. MuThiya, and S. Salamah, "Teknologi Informasi Dalam Bidang Kesehatan Masyarakat," *JIKES J. Ilmu Kesehat.*, vol. 1, no. 2, pp. 165–172, 2023.
- [6]. Z. DIYAN, A. Nurdin, U. Fitria, and R. Kurnia, "Pemanfaatan teknologi dalam bidang kesehatan masyarakat," *Public Heal. J.*, vol. 1, no. 2, 2024.

## **Jurnal EDU ELEKTROMATIKA (JEE)**

ISSN: 2747-0784 (p); xxxxxx (e)

Vol 6, No. 2, Desember 2025

- [7]. Maghribi, "NURSING TELEHEALTH NURSING TELEHEALTH (SISTEM INFORMATIKA KEPERAWATAN): SYSTEMATIC LITERATURE REVIEW: SYSTEMATIC LITERATURE REVIEW," J. KEPERAWATAN, vol. 17, no. 1, pp. 10–18, 2023.