

# Implementasi Algoritma *Naive Bayes* Untuk Klasifikasi Penyakit Ispa

Badriyah Ani Kusuma Putri<sup>1</sup>, Taghfirul Azhima Yoga Siswa<sup>2\*</sup>, Fendy Yulianto<sup>3</sup>

<sup>1,2,3</sup>Sains dan Teknologi, Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur, Kota Samarinda, Indonesia

Email: <sup>1</sup>2011102441009@umkt.ac.id, <sup>2\*</sup>tay758@umkt.ac.id, <sup>3</sup>fy415@umkt.ac.id

\*) Penulis Korespondensi

**Abstrak**— Infeksi Saluran Pernapasan Akut (ISPA) ialah salah satu penyakit pada permasalahan kesehatan serius di Indonesia, terutama pada balita. Proses klasifikasi ISPA secara tradisional sering kali tidak cukup efektif untuk menangani variasi gejala yang kompleks. Oleh karena itu, penelitian ini bertujuan untuk menerapkan algoritma *Naive Bayes* guna mengklasifikasikan tingkat keparahan penyakit ISPA secara lebih akurat dan efisien. Dataset yang digunakan berasal dari UPT Puskesmas Bontang Barat dengan total 1.501 data pasien pada periode 2024-2025, yang telah melalui proses *pre-processing* berupa data *selection*, *cleaning*, dan *transformation*. Proses klasifikasi menggunakan *Gaussian Naive Bayes* dan divalidasi menggunakan metode *K-Fold Cross Validation* dengan  $K=10$ . Hasil evaluasi menunjukkan bahwa algoritma ini mampu mencapai rata-rata nilai *Accuracy* sebesar 75,00%, *Precision* 78,12%, *Recall* 75,00%, dan *F1-Score* 76,36%. Nilai ini menunjukkan bahwa *Naive Bayes* cukup andal dalam mendukung proses diagnosis ISPA dan memiliki kinerja yang relatif lebih baik dibandingkan beberapa algoritma lain dalam penelitian terdahulu.

**Kata Kunci:** *Naive Bayes*, ISPA, Klasifikasi, *Machine Learning*, Data Mining

**Abstract**— *Acute Respiratory Infection (ARI)* is a disease that remains a serious health problem in Indonesia, especially in toddlers. The traditional *ARI* classification process is often not effective enough to handle complex symptom variations. Therefore, this study aims to apply the *Naive Bayes* algorithm to classify the severity of *ARI* disease more accurately and efficiently. The data set used comes from the West Bontang Community Health Center (UPT Puskesmas) with a total of 1,501 patient data in the 2024-2025 period, which has gone through a *pre-processing* process in the form of data *selection*, *cleaning*, and *transformation*. The classification process uses *Gaussian Naive Bayes* and is validated using the *K-Fold Cross Validation* method with  $K=10$ . The evaluation results show that this algorithm can achieve an average *Accuracy* value of 75.00%, *Precision* 78.12%, *Recall* 75.00%, and *F1-Score* 76.36%. This value shows that *Naive Bayes* is quite reliable in supporting the *ISPA* diagnosis process and has relatively better performance compared to several other algorithms in previous studies.

**Keywords:** *Naive Bayes*, *ARI*, *Classification*, *Machine Learning*, Data Mining

## 1. PENDAHULUAN

Infeksi Saluran Pernapasan Akut (ISPA) ialah gangguan pada sistem pernapasan yang muncul secara tiba-tiba penyakit ini disebabkan oleh beberapa faktor yaitu virus atau bakteri [1]. Sehingga gejala yang dialami penderita bisa beragam seperti Ketika seseorang menderita ISPA sejumlah organ pernapasan seperti hidung, tenggorokan, laring, saluran bronkus, paru-paru, dan trakea menjadi sasaran infeksi [2]. Penyakit ISPA mudah menular baik melalui udara maupun melalui benda yang telah terpapar kuman, penularannya bisa terjadi antara individu dengan cepat serta gejala yang ditimbulkan bisa berbeda-beda, mulai dari kondisi ringan hingga cukup serius dimana hal ini bisa berdampak negatif pada aktivitas harian serta menurunkan kualitas hidup orang yang terinfeksi [3].

Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Indonesia masih menjadi suatu permasalahan kesehatan yang cukup besar dan cukup serius khususnya pada balita [4]. Hingga saat ini kasus ISPA masih sering ditemukan dan tercatat sebanyak 86,364 kasus yang telah terjadi tingginya angka kejadian ini menunjukkan bahwa ISPA masih menjadi salah satu penyakit yang berbahaya [5]. Menurut Badan Pusat Statistik (BPS) Kota Bontang tercatat 5.820 kasus di tahun 2024 yang menjadikannya sebagai kasus terbanyak ketiga di kota tersebut. Jumlah ini menunjukkan penyakit ISPA tergolong salah satu penyakit cukup serius, beberapa cara tradisional yang umumnya digunakan untuk mengklasifikasikan penyakit ISPA seperti *anamnesis* (wawancara dengan pasien/orang tua pasien), pemeriksaan fisik, kriteria diagnosis sederhana dari *World Health Organization (WHO)* atau Departemen Kesehatan, pemeriksaan penunjang manual, observasi klinis berulang, penggunaan buku Kesehatan Ibu dan Anak (KIA) [6]. Namun untuk menangani permasalahan di bidang kesehatan secara lebih akurat diperlukan adanya pendekatan yang lebih moderen seperti penggunaan Kecerdasan Buatan dalam melakukan proses klasifikasi [7].

*Artificial Intelligence (AI)* ialah cabang dari ilmu komputer yang memiliki fokus utama untuk membuat mesin belajar, memahami situasi, membuat perencanaan, dan beradaptasi agar mampu menyelesaikan berbagai pekerjaan secara mandiri, kemudian dikembangkan untuk meniru cara berfikir manusia dalam hal pengambilan keputusan, penalaran logis, dan aspek kecerdasan lainnya [8]. Dalam kecerdasan buatan ada beberapa pendekatan

seperti *Machine Learning*, Sistem Pakar, Visi Komputer, dan *Natural Language Processing*, dalam penerapan kecerdasan buatan tersebut melalui proses pengolahan data informasi dengan pembelajaran mesin yang dikenal *Machine Learning* pada pembelajaran tersebut menggunakan teknik-teknik atau metode yang dikenal dengan algoritma [9].

*Machine Learning* ialah bagian dari kecerdasan buatan yang memungkinkan sistem mempelajari pola dari data secara otomatis, *Machine Learning* memiliki beberapa teknik utama, seperti *Supervised Learning* dan *Unsupervised Learning* yang digunakan sesuai dengan karakteristik dan tujuan analisis data [10]. Pendekatan *Supervised Learning* umumnya digunakan untuk menyelesaikan permasalahan seperti klasifikasi, sementara Pendekatan *Unsepervised Learning* digunakan untuk permasalahan klusterisasi [11]. Data yang digunakan pada penelitian ini sudah dilengkapi dengan label (*Class*) dengan tujuan yang berfokus pada proses klasifikasi, pada proses klasifikasi metode *Machine Learning* memiliki beberapa algoritma yang sering digunakan seperti algoritma *Naive Bayes*, *Decision Tree*, *K-Nearest Neighbor (KNN)*, *Support Vector Machine (SVM)* maupun *Linier Regression* [12].

Terkait penelitian sebelumnya yang dilakukan oleh [13] dalam mengklasifikasikan penyakit gagal jantung memakai algoritma *K-Nearest Neighbor* dengan akurasi 76%, dan *Naive Bayes* 90%. Sementara pada kasus penyakit lain dalam mengklasifikasikan penyakit diabetes yang membandingkan beberapa algoritma yaitu, *K-Nearest Neighbor* dengan akurasi yang di dapat 74,48%, dan *Naive Bayes* 75,78% [14]. Kemudian penelitian yang melakukan perbandingan algoritma *Decision Tree* dan *Naive Bayes* pada kasus penyakit kanker paru-paru dengan akurasi *Decision Tree* 88,17% maupun *Naive Bayes* 92,47% [15]. Pada kasus penyakit serangan jantung yang membandingkan algoritma *Logistic Regression* dengan akurasi 81.7%, *Neural Network* 78.2%, dan *Naive Bayes* 84.3% [16]. Kemudian pada penelitian [17] dalam mengklasifikasikan penyakit ISPA menggunakan algoritma *K-Nearest Neighbor (KNN)* dengan akurasi 71%, *Decision Tree* 68%, *Random Forest* 72%, dan *Artificial Neural Network (ANN)* 67%. Berdasarkan penelitian sebelumnya algoritma *Naive Bayes* menunjukkan performa yang konsisten dan unggul dalam beberapa kasus, namun pada penelitian terkait klasifikasi penyakit ISPA yang telah dilakukan sebelumnya belum menerapkan algoritma *Naive Bayes*. Oleh karena itu, pada penelitian ini ingin mencoba pendekatan baru memakai algoritma *Naive Bayes* menjadi metode klasifikasi penyakit ISPA.

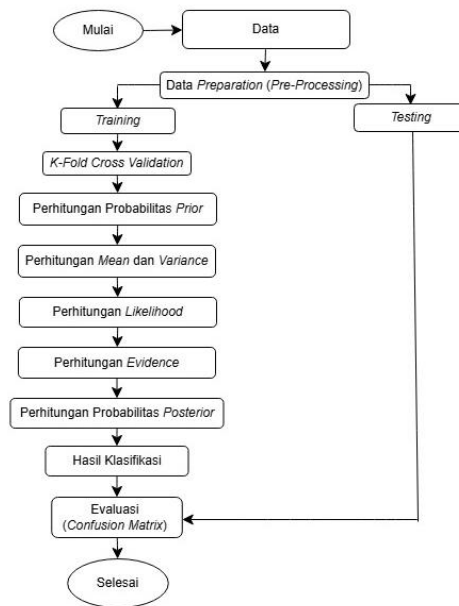
*Naive Bayes* merupakan algoritma klasifikasi yang dikenal cepat dan efisien, algoritma ini juga tetap bekerja optimal meskipun jumlah data latih terbatas, karena mengandalkan prinsip probabilitas *bayes* yang sederhana namun akurat [18]. Seperti penelitian [19] pada kasus klasifikasi penyakit diabetes yang membandingkan beberapa algoritma ialah *Decision Tree* dengan akurasi 69%, *Support Vector Machine* 76%, dan *Naive Bayes* 78%. Pada kasus serupa yang membandingkan algoritma *K-Nearest Neighbor* dengan akurasi 71%, dan *Naive Bayes* 77% [20].

Dari permasalahan untuk mengklasifikasikan penyakit Infeksi Saluran Pernapasan Akut (ISPA) secara lebih akurat dan efisien dengan menerapkan algoritma *Naive Bayes*. Mengingat bahwa algoritma *Naive Bayes* telah terbukti memberikan hasil klasifikasi yang konsisten dan unggul dalam berbagai kasus penyakit dari penelitian terdahulu untuk mengeksplorasi penerapan algoritma tersebut pada kasus penyakit ISPA yang sebelumnya belum dilakukan. Melalui pendekatan berbasis *Machine Learning*, penelitian ini diharapkan dapat memberikan kontribusi dalam mendukung proses diagnosis dini penyakit ISPA.

## 2. METODE PENELITIAN

### 2.1 Prosedur Penelitian

Prosedur penelitian ini terdiri dari beberapa tahap pada tujuan dengan tahapannya dimulai dengan Identifikasi Masalah, Pengumpulan Data, Data *Preparation (Pre-Processing)* akan dilakukan 3 langkah utama yaitu data *selection*, data *cleaning*, data *trasformation* untuk memastikan bahwa data pada kondisi yang sesuai untuk diklasifikasikan., Pembagian Data, *K-Fold Cross Validation*, *Naive Bayes* maupun diakhiri Evaluasi memakai *confusion matrix*.



Gambar 1 Alur Algoritma Naive Bayes

### 2.1.1 K-Fold Cross Validation

Pada tahap ini peneliti menerapkan metode *K-Fold Cross Validation* dengan  $K=10$  mengukur performa model, yang dianggap mampu memberikan hasil akurasi cukup baik [21]. Berdasarkan pada penelitian sebelumnya yang dilakukan [14] melakukan perbandingan  $K=10$  dan  $K=25$  dengan hasil *Accuracy* yang didapat pada  $K=10$  sebesar 75,78% sedangkan pada pengujian  $K=25$  hanya mendapatkan *accuracy* 74,48%. Oleh sebab itu pada penelitian memakai  $K=10$  dalam penelitian ini, yang bertujuan untuk menghasilkan penilaian yang lebih tepat terkait performa model dalam mengklasifikasikan data penyakit ISPA di Kota Bontang di Tabel 1.

Tabel 1. 10 K-Fold Cross Validation

1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10

Pada tahap ini peneliti menerapkan teknik *k-fold cross validation*  $k=10$  mengukur performa model *machine learning*. Teknik *k-fold cross validation* dianggap mampu hasil akurasi baik [21].

### 2.1.2 Penerapan Naive Bayes

*Naive bayes* adalah algoritma klasifikasi yang menggunakan pendekatan probabilitas yang didasarkan pada teorema bayes, dimana setiap fitur *independent* [22]. Evaluasi kinerja algoritma ini dilakukan dengan menggunakan varian *Gaussian Naive Bayes* kemudian dataset akan dibagi menjadi dua bagian, data pelatihan maupun pengujian, selanjutnya algoritma *Gaussian Naive Bayes* diuji untuk mengukur tingkat akurasi, *precision*, *recall*, dan *F1-score* pada klasifikasi algoritma ini beroperasi berdasarkan Persamaan 1.

$$P(x_i|y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (1)$$

Keterangan:

- $x_i$  : Adalah nilai fitur ke- $i$
- $y$  : Adalah kelas
- $\mu_y$  : Adalah rata-rata fitur  $x_i$  untuk kelas  $y$
- $\sigma_y^2$  : Adalah varians dari fitur  $x_i$  untuk kelas  $y$

Alur metode merupakan susunan langkah-langkah sistematis yang diterapkan untuk menggunakan suatu teknik atau pendekatan guna mencapai tujuan penelitian, dengan memastikan setiap tahapan dilakukan secara terencana dan teratur, tahapan dari algoritma *Naive Bayes* yang digunakan dalam penelitian klasifikasi penyakit ISPA. berikut merupakan tahap-tahap dari algoritma *Naive Bayes* [23].

1. Menghitung probabilitas *prior* untuk setiap kelas  $P(C_i)$ .

Probabilitas prior adalah frekuensi kelas dalam dataset pelatihan, Untuk menghitung prior  $P(C_i)$  kita dapat menggunakan rumus :

$$P(C_i) = \frac{S_i}{s} \quad (2)$$

Keterangan:

- $P(C_i)$  : Probabilitas *prior*
- $S_i$  : Jumlah data dari kelas  $C_i$
- $s$  : Jumlah total data

2. Menghitung probabilitas *Likelihood* menggunakan *Mean*.

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i \quad (3)$$

Keterangan:

- $\mu$  : Rata-rata dari data
- $n$  : Jumlah total data
- $x_i$  : Nilai ke- $i$  dari data

3. Menghitung kuadrat *Mean*

$$(x_i - \mu)^2 \quad (4)$$

Keterangan:

- $x_i$  : Nilai individual dari data ke- $i$  dalam kumpulan data
- $\mu$  : Nilai rata-rata dari data
- $(x_i - \mu)$  : Selisih nilai individual maupun rerata

4. Menghitung nilai selisih kuadrat

$$\sum_{i=1}^n (x_i - \mu)^2 \quad (5)$$

Keterangan:

- $n$  : Jumlah data
- $x_i$  : Nilai data ke- $i$
- $\mu$  : Nilai rata-rata dari data

5. Menghitung nilai probabilitas *Likelihood* menggunakan *Variance*

$$s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1} \quad (6)$$

Keterangan:

- $s^2$  : *Variance* sampel.
- $x_i$  : Nilai individu.
- $\mu$  : Mean.
- $n$  : Total total data.

6. Menghitung probabilitas *Likelihood*

$$P(x_i|y) = \frac{N_{x_i y}}{N_y} \quad (7)$$

Keterangan:

- $P(x_i|y)$  : Probabilitas terjadinya fitur  $x_i$  diberikan kelas  $y$ .
- $N_{x_i y}$  : Jumlah data dengan fitur  $x_i$  dan kelas  $y$ .
- $N_y$  : Jumlah total data dengan kelas  $y$ .

#### 7. Menghitung nilai Evidence

$$P(x) = \sum_{j=i}^m P(X|C_j) \times P(C_j) \quad (8)$$

Keterangan:

- $P(x)$  : Evidence.
- $\sum_{j=i}^m$  : Jumlah total kelas.
- $P(X|C_j)$  : Nilai Likelihood.
- $P(C_j)$  : Probabilitas prior dari nilai  $C_j$ .

#### 8. Menghitung probabilitas Posterior

$$P(C_k|X) = \frac{P(X|C_k) \cdot P(C_k)}{P(X)} \quad (9)$$

Keterangan:

- $P(C_k|X)$  : Probabilitas Posterior dari kelas  $C_k$  terhadap data  $X$
- $P(X|C_k)$  : Probabilitas data  $X$  terhadap kelas  $C_k$  atau Likelihood.
- $P(C_k)$  : Probabilitas Prior dari kelas  $C_k$  sebelum mempertimbangkan data  $X$
- $P(X)$  : Probabilitas total dari data  $X$  atau Evidence.

### 2.1.3 Evaluasi Model

Tahap evaluasi merupakan bagian penting setelah model dibangun, pada tahap ini performa model dianalisis untuk menilai tingkat Accuracy, Precision, Recall maupun F1-score. Kinerja model akan diukur menggunakan teknik Confusion Matrix sebagai metode untuk menghitung Accuracy, Precision, Recall, dan F1-score dalam data mining [24]. Confusion Matrix terdiri dari empat istilah utama merepresentasikan hasil klasifikasi, yaitu True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [25]. Adapun rumusnya adalah :

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (10)$$

Keterangan:

- TP (True Positive) : jumlah prediksi benar untuk kelas positif (kelas minoritas yang benar terklasifikasi sebagai positif).
- TN (True Negative) : jumlah prediksi benar untuk kelas negatif (kelas mayoritas yang benar terklasifikasi sebagai negatif).
- FP (False Positive) : jumlah prediksi salah dimana kelas negatif diklasifikasikan sebagai positif.
- FN (False Negative) : jumlah prediksi salah dimana kelas positif diklasifikasikan sebagai negatif.

$$Precision = \frac{TP}{TP + FP} \times 100\% \quad (11)$$

Keterangan:

- TP (True Positive) : jumlah prediksi benar untuk kelas positif (kelas minoritas yang benar terklasifikasi sebagai positif).
- FP (False Positive) : jumlah prediksi salah dimana kelas negatif diklasifikasikan sebagai positif.

$$Recall = \frac{TP}{TP + FN} \times 100\% \quad (12)$$

Keterangan:

- TP (True Positive) : jumlah prediksi benar untuk kelas positif (kelas minoritas yang benar terklasifikasi sebagai positif).

FN (*False Negative*) : jumlah prediksi salah dimana kelas positif diklasifikasikan sebagai negatif.

$$F1 - score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \quad (13)$$

Keterangan:

- Precision* : Mengukur seberapa akurat prediksi positif model.
- Recall* : Mengukur seberapa baik model menangkap semua data.

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Hasil Penelitian

Pada bagian ini dibahas hasil analisis dan pengujian model, dengan fokus pada evaluasi performa serta kesesuaian hasil terhadap tujuan penelitian. Analisis meliputi pengumpulan dan pengolahan data, hingga penerapan algoritma Naïve Bayes pada dataset penyakit ISPA di Kota Bontang yang memiliki 12 fitur (11 sebagai atribut dan 1 sebagai target class).

#### 3.2 Pengumpulan Data

Data penelitian ialah data penyakit ISPA dari Unit Pelayanan Terpadu Pukesmas Bontang Barat, data yang terkumpulan mencakup 12 fitur untuk mengklasifikasikan data penyakit ISPA. Adapun fitur-fitur yang didapatkan yaitu Jenis.kelamin, Umur, Pendidikan, Sistolik, Diastolik, Nadi, RR (*Respiratory Rate*), Temp (Suhu Tubuh), BB (Berat Badan), TB (Tinggi Badan), LP (*Leukosit Polimorfonuklear*), *Disease* berikut adalah contoh detail di Tabel 2.

Tabel 2 Data ISPA

No	Jns.Kelamin	Umur	Pendidikan	.....	TB	LP	<i>Disease</i>
1	Perempuan	31thn	SMA	.....	120	79	Ringan
2	Perempuan	40thn	Diploma	.....	141	87	Ringan
3	Laki-laki	41thn	SMA	.....	60	0	Ringan
4	Perempuan	26thn	Diploma	.....	138	70	Ringan
5	Perempuan	54thn	SMA	.....	134	85	Ringan
.....	.....	.....	.....	.....	.....	.....	.....
1497	Laki-laki	1	Tidak diketahui	.....	73	27	Berat
1498	Perempuan	8	SD	.....	69	67	Ringan
1499	Perempuan	24	Diploma	.....	110	87	Berat
1500	Perempuan	26	SMA	.....	80	86	Ringan
1501	Perempuan	61thn	SD	.....	55	78	Ringan

Pada Tabel 2 menampilkan fitur-fitur dan isi dari setiap fitur yang didapat dari Unit Pelayanan Terpadu Pukesmas Bontang Barat.

#### 3.3 Hasil Data Pre-processing

Pada tahap ini menampilkan hasil dari tahapan *Pre-processing* data setelah selesai melakukan proses pengumpulan data, agar data tersebut dapat dipastikan bahwa sudah dalam keadaan bersih, terstruktur, dan sesuai dengan logika program sehingga siap diproses lebih lanjut ketahap permodelan.

##### 3.3.1 Data Selection

Pada tahap ini, dilakukan proses ekstraksi data dengan memilih fitur-fitur yang dianggap paling relevan, sedangkan fitur yang dianggap kurang relevan dihilangkan. Proses pemilihan ini dilakukan secara manual dengan menetapkan 9 fitur utama serta 1 fitur sebagai class, sebagaimana ditunjukkan pada Tabel 3.

Tabel 3 Tampilan Data Selection

No	Jns.Kelamin	Sistolik	Diastolik	Nadi	RR	Temp	BB	TB	LP	<i>Disease</i>
1	Perempuan	110	70	89	21	36.5	17	120	79	Ringan
2	Perempuan	140	90	80	18	36.0	30	141	87	Ringan
3	Laki-laki	110	70	98	21	36.4	5.7	60	0	Ringan

4	Perempuan	110	60	89	18	36.5	28	138	70	Ringan
5	Perempuan	120	80	89	23	36.7	25	134	85	Ringan
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
1497	Laki-laki	70	78	78	24	36.0	8	73	27	Berat
1498	Perempuan	70	78	78	21	36.7	7	69	67	Ringan
1499	Perempuan	70	80	80	24	36.6	14	110	87	Berat
1500	Perempuan	70	80	80	24	36.6	12	80	86	Ringan
1501	Perempuan	70	80	80	24	36.6	6	55	78	Ringan

Pada Tabel 3, dipilih 10 fitur utama berdasarkan relevansi, kualitas data, dan signifikansi. Fitur tersebut meliputi Jns.kelamin, Sistolik, Diastolik, Nadi, RR, Temp, BB, TB, LP, dan Disease. Proses ini dilakukan untuk memastikan hanya data relevan dan berkualitas yang dianalisis, sehingga meningkatkan efisiensi klasifikasi penyakit ISPA.

### 3.3.2 Data Cleaning

Tahap ini menampilkan hasil pembersihan data pada dataset ISPA berisi 1.501 baris. Pengecekan nilai kosong (*NaN*) dilakukan menggunakan fungsi `data.isna().sum()`, karena nilai *NaN* dapat memengaruhi hasil analisis. Tabel 4 menunjukkan kondisi data sebelum tahap *cleaning*.

**Tabel 4** Data Sebelum Tahap *Cleaning*

No	Fitur	Missing Values
1	Jns.Kelamin	0
2	Sistolik	0
3	Diastolik	0
4	Nadi	0
5	RR	0
6	Temp	0
7	BB	13
8	TB	0
9	LP	4
10	Disease	0

Pada Tabel 3, terdeteksi 13 nilai *NaN* pada fitur Temp dan 4 pada LP. Untuk membersihkannya, digunakan fungsi `drop_duplicate()` untuk menghapus duplikasi dan `dropna()` untuk menghapus nilai *NaN*. Hasil data setelah pembersihan ditampilkan pada Tabel 4.

**Tabel 5** Data Setelah Tahap *Cleaning*

No	Fitur	Missing Values
1	Jns.Kelamin	0
2	Sistolik	0
3	Diastolik	0
4	Nadi	0
5	RR	0
6	Temp	0
7	BB	0
8	TB	0
9	LP	0
10	Disease	0

Pada Tabel 5 terlihat hasil pembersihan data melalui penghapusan duplikat dan nilai *NaN*, sehingga total 17 baris terhapus. Jumlah data akhir menjadi 1.479 baris, membuat dataset lebih bersih dan siap dianalisis tanpa gangguan missing values.

### 3.3.3 Data Transformation

Tahap data transformation mengubah fitur bertipe string menjadi numerik agar dapat diproses oleh algoritma machine learning seperti Naïve Bayes. Transformasi ini penting untuk memastikan data sesuai format yang diperlukan. Tabel 6 menunjukkan data sebelum transformasi.

**Tabel 6** Data Sebelum Tahap *Transformation*

No	Jns.kelamin	Disease
1	Perempuan	Ringan
2	Perempuan	Ringan
3	Laki-laki	Ringan
4	Perempuan	Ringan
5	Perempuan	Ringan
.....	.....	.....
1497	Laki-laki	Berat
1498	Perempuan	Ringan
1499	Perempuan	Berat
1500	Perempuan	Ringan
1501	Perempuan	Ringan

Pada Tabel 6, fitur Jns.Kelamin dan Disease masih bertipe *string* sehingga perlu ditransformasi menggunakan *LabelEncoder* dari *sklearn* menjadi nilai numerik. Hasil transformasi ditampilkan pada Tabel 7.

**Tabel 7** Data Setelah Tahap *Transformation*

No	Jns.kelamin	Disease
1	1	1
2	1	1
3	0	1
4	1	1
5	1	1
.....	.....	.....
1497	0	0
1498	1	1
1499	1	0
1500	1	1
1501	1	1

Pada Tabel 7 menunjukkan hasil dari transformasi pada fitur Jns.Kelamin dan *Disease* yang sebelumnya bernilai *string* kemudian di rubah kedalam bentuk numerik sesuai dengan label yang diberikan oleh *LabelEncoder*.

### 3.4 Hasil Pembagian Data Dengan *K-Fold Cross Validation*

Pada penelitian ini, dataset dibagi menjadi data training dan testing menggunakan metode 10-Fold Cross Validation. Seluruh dataset dibagi menjadi sepuluh subset; setiap iterasi menggunakan satu subset sebagai data testing dan sisanya sebagai data training. Proses ini dilakukan secara acak di setiap fold untuk meminimalkan bias dan menghasilkan evaluasi yang lebih akurat. Hasil evaluasi ini digunakan untuk menganalisis performa algoritma Naïve Bayes dalam klasifikasi data.

### 3.5 Hasil Tahapan Algoritma Naive Bayes

Penelitian memakai algoritma Gaussian Naive Bayes untuk mengklasifikasikan penyakit ISPA fokus pada pengukuran Accuracy, Precision, Recall maupun F1-Score memakai metode 10-Fold Cross Validation. Sebelum evaluasi, dilakukan beberapa tahapan penting dalam penerapan Gaussian Naive Bayes, berikut hasil dari setiap tahap tersebut.

#### 3.6.1 Probabilitas *Prior*

Probabilitas *Prior* dihitung berdasarkan jumlah dari yang terkena *class* ringan atau berat kemudian dibagi dengan jumlah dataset, untuk menghitung probabilitas *Prior* dalam menentukan tingkat keparahan pada penderita penyakit ISPA dengan menggunakan Persamaan 2.2.

$$P(C_{ringan}) = \frac{1240}{1479} = 0,8384$$

$$P(C_{berat}) = \frac{239}{1479} = 0,1616$$

hasil implementasi dari program dalam menghitung Probabilitas *Prior* pada tampilan Gambar 2.

=== Probabilitas Prior dari Seluruh Dataset ===  
 Kelas 0: 0.16160  
 Kelas 1: 0.83840

**Gambar 2** Hasil Probabilitas *Prior*

Pada Gambar 2 hasil dari Probabilitas *Prior Class* berat (0) yaitu 0,16160 kemudian Probabilitas *Prior Class* ringan (1) 0,83840.

### 3.6.2 Perhitungan *Likelihood*

Sebelum memasuki tahap *Likelihood* pada proses menganalisis data perhitungan nilai *Mean* dan *Variance* menjadi bagian penting dalam memahami distribusi data, dimana *Mean* (nilai rata-rata) didapatkan dengan menjumlahkan beberapa data sampel kemudian membaginya dengan jumlah data sampel yang di gunakan, dalam melakukan perhitungan *Mean* menggunakan Persamaan 2.3.

$$\mu = \frac{1 + 1 + 0 + 1}{4} = \frac{3}{4} = 0,75$$

Hasil dari perhitungan *Mean* dalam menentukan tingkat keparahan pada penderita penyakit ISPA diTabel 8.

**Tabel 8** Perhitungan *Mean*

Atribut	Ringan	Berat
Jns.kelamin	0,75	0,25
Sistolik	117,5	125
Diastolik	72,5	75
Nadi	89	85
RR	19,5	23,5
Temp	36,35	36,525
BB	20,175	33,925
TB	114,75	127,05
LP	59	65

Pada Tabel 8 merupakan nilai *Mean* dari atribut yang terdapat dalam dataset penyakit ISPA, untuk atribut Jns.kelamin pada *Class* ringan memiliki nilai 0,25 dan *Class* berat 0,75. Sistolik memiliki nilai *Mean Class* ringan 117,5 dan *Class* berat 125. Diastolik memiliki nilai *Mean Class* ringan 72,5 dan *Class* berat 75. Nadi dengan nilai *Mean Class* ringan 89 dan *Class* berat 85. RR dengan nilai *Mean Class* ringan 19,5 dan *Class* berat 23,5. Temp dengan nilai *Mean Class* ringan 36,35 dan *Class* berat 36,525. BB dengan nilai *Mean Class* ringan 20,175 dan *Class* berat 33,925. TB dengan nilai *Mean Class* ringan 114,75 dan *Class* berat 127,05. Kemudian yang terakhir atribut LP dengan nilai *Mean Class* ringan 59 dan *Class* berat 65.

### 3.6.3 Perhitungan Selisih Kuadrat

Selisih kuadrat adalah hasil dari pengurangan setiap nilai data dengan nilai *Mean* setelah itu dikuadratkan tahap ini menjadi tahapan awal yang penting untuk menghitung selisih kuadrat pada masing-masing data dengan nilai *Mean* sebelum melakukan perhitungan *Variance*, perhitungan selisih kuadrat memakai rumus 2.4.

$$\begin{aligned} (x_i - \mu)^2 &= (1 - 0,75)^2 = 0,25^2 = 0,0625 \\ (x_i - \mu)^2 &= (1 - 0,75)^2 = 0,25^2 = 0,0625 \\ (x_i - \mu)^2 &= (0 - 0,75)^2 = 0,75^2 = 0,5625 \\ (x_i - \mu)^2 &= (1 - 0,75)^2 = 0,25^2 = 0,0625 \end{aligned}$$

Setelah itu selisih kuadrat dijumlahkan untuk menentukan tingkat keparahan pada penderita penyakit ISPA dengan menggunakan Persamaan 2.5.

$$\sum_{i=1}^n (x_i - \mu)^2 = 0,0625 + 0,0625 + 0,5625 + 0,0625 = 0,75$$

### 3.6.4 Perhitungan Variance

Variance berfungsi untuk mengetahui seberapa besar penyebaran data terhadap nilai Mean (nilai rata-rata), Untuk mengitung Variance dalam menentukan tingkat keparahan pada penderita penyakit ISPA dapat menggunakan Persamaan 2.6.

$$s^2 = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n - 1} = \frac{0,0625 + 0,0625 + 0,5625 + 0,0625}{4 - 1} = \frac{0,75}{3} = 0,25$$

Hasil keseluruhan dari perhitungan Variance dalam menentukan tingkat keparahan pada penderita penyakit ISPA di Tabel 9.

**Tabel 9** Perhitungan Variance

Atribut	Ringan	Berat
Jns.kelamin	0,25	0,25
Sistolik	225	366,66
Diastolik	158,33	33,33
Nadi	54	28
RR	3	27,66
Temp	0,056	0,7833
BB	128,23	595,95
TB	1.417,91	1.565,47
LP	1.595,33	1.880,66

Pada Tabel 10 merupakan hasil dari perhitungan Variance yang menunjukkan seberapa besar penyebaran data terhadap nilai Mean (nilai rata-rata), kemudian untuk hasil dari implementasi program di Tabel 10.

**Tabel 10** Hasil Perhitungan Mean Dan Variance

Atribut	Mean (0)	Mean (1)	Variance (0)	Variance (1)
Jns.kelamin	0,53	0,56	0,25	0,25
Sistolik	115,95	114,36	391,13	374,28
Diastolik	75,4	72,96	116,51	88,21
Nadi	80,62	81,01	22,46	39,72
RR	21,18	21,26	9,49	14,6
Temp	36,54	36,85	0,26	94,57
BB	44,94	38,84	525,48	541,78
TB	139,2	132,19	884,59	1071,14
LP	55,69	46,32	1314,44	1424,66

Pada Tabel 10 Class ringan menunjukkan nilai rata-rata (Mean) dan Variance cenderung lebih tinggi dibandingkan Class berat yang mengindikasikan adanya variasi fitur yang lebih besar. Sebaliknya Class berat memperlihatkan nilai rata-rata (Mean) dan Variance lebih rendah yang menunjukkan bahwa fitur-fitur dalam Class berat memiliki tingkat variasi lebih kecil, selanjutnya hasil perhitungan nilai rata-rata (Mean) dan Variance yang digunakan untuk menentukan tingkat keparahan penderita penyakit ISPA sesuai Tabel 11.

**Tabel 11** Hasil Rumus Mean Variance

TB	139,2	132,19	884,59	1071,14
----	-------	--------	--------	---------

Pada Tabel 10 Class ringan yang terdapat pada atribut TB (Tinggi Badan) memiliki nilai rata-rata (Mean) dan Variance yang lebih besar, yang menunjukkan bahwa nilai rata-rata (Mean) lebih menyebar dibandingkan Class berat kemudian nilai Variance yang besar mengindikasikan adanya variasi yang signifikan dalam dataset.

### 3.6.5 Probabilitas Likelihood

Probabilitas Likelihood adalah peluang munculnya suatu fitur, nilai ini dihitung perbandingan frekuensi kemunculan fitur tersebut terhadap total jumlah data kelas. Probabilitas Likelihood dalam menentukan tingkat keparahan penderita penyakit ISPA dengan memakai Persamaan 2.7.

$$P(x_{perempuan}|y_{ringan}) = \frac{543}{1.240} = 0,4379$$

$$P(x_{laki-laki}|y_{ringan}) = \frac{697}{1.240} = 0,5620$$

$$P(x_{perempuan}|y_{berat}) = \frac{108}{239} = 0,4518$$

$$P(x_{laki-laki}|y_{berat}) = \frac{131}{293} = 0,5481$$

Hasil keseluruhan dari perhitungan *Likelihood* dalam menentukan tingkat keparahan pada penderita penyakit ISPA Tabel 12.

**Tabel 12** Perhitungan *Likelihood*

Atribut	Ringan	Berat
Jns.kelamin		
Perempuan	0,4379	0,4518
Laki-laki	0,5620	0,5481
Sistolik		
110	0,0235	0,01533
Diastolik		
70	0,0311	0,0475
Nadi		
89	0,0543	0,0567
RR		
21	0,1583	0,0678
Temp		
36,5	0,9646	0,4504
BB		
17	0,0339	0,01284
TB		
120	0,0105	0,00992
LP		
79	0,00882	0,00873

Pada Tabel 12 merupakan hasil dari perhitungan *Likelihood* pada *class* ringan dan berat yang terdapat pada atribut Jns.kelamin, Sistolik, Diastolik, Nadi, RR , Temp, BB, TB, LP kemudian untuk hasil dari implementasi program dapat dilihat pada Gambar 3.

```

Category: Temp
36.5 36.0 36.4 36.7 36.8 37.5 37.4 36.6 36.9 37.2 36.3 37.1 37.0 36.1 37.8
Class: 1 0.071774 0.251613 0.037903 0.131452 0.068548 0.009677 0.020161 0.145968 0.046774
Class: 0 0.037657 0.343096 0.012552 0.104603 0.104603 0.008368 0.020921 0.133891 0.046025

Category: Nadi
89.0 80.0 98.0 75.0 79.0 93.0 78.0 104.0 96.0 106.0 88.0 66.0 82.0 84.0 100.0
Class: 1 0.004839 0.294355 0.007258 0.002419 0.004839 0.002419 0.312903 0.000806 0.012903
Class: 0 0.004184 0.355649 0.000000 0.004184 0.000000 0.000000 0.322176 0.000000 0.000000

Category: LP
79.0 87.0 0.0 70.0 85.0 40.0 48.0 81.0 82.0 60.0 45.0 80.0 43.0 90.0 38.0
Class: 1 0.008065 0.012903 0.362097 0.006452 0.095161 0.004839 0.011290 0.007258 0.025806
Class: 0 0.004184 0.012552 0.280335 0.000000 0.133891 0.008368 0.004184 0.004184 0.037657
    
```

**Gambar 3** Hasil Probabilitas *Likelihood*

Pada Gambar 3 merupakan hasil implementasi program dari perhitungan Probabilitas *Likelihood* penyakit ISPA yang paling berpengaruh pada kelas ringan yaitu Temp, Nadi, dan LP.

### 3.6.6 Perhitungan Evidence

Perhitungan Evidence adalah nilai dari data pada probabilitas maupun penentuan kemungkinan suatu data termasuk dalam suatu Class. Untuk perhitungan Evidence sesuai pada 2.8.

$$P(x) = 8,09401866 \times 0,8384 + 6,3340876 \times 0,1616 \\ = 6,78602524 + 1,02358856 = 7,8096138$$

Hasil dari perhitungan probabilitas Evidence dalam menentukan tingkat keparahan pada penderita penyakit ISPA yang menghasilkan nilai 7,8096138 yang menunjukkan bahwa Likelihood untuk semua fitur yaitu 7,8096138, berikut merupakan hasil implementasi program dapat dilihat pada Gambar 4.

$$Evidence (P(X)) = 1.8779259607e - 13$$

**Gambar 4** Hasil Perhitungan Evidence

Pada Gambar 4 menampilkan hasil Evidence dengan nilai sebesar 1.8779259607e-13 yang menunjukkan bahwa semua fitur tersebut sangat jarang ditemukan dalam dataset.

### 3.6.7 Perhitungan Probabilitas Posterior

Probabilitas Posterior menggambarkan tingkat kemungkinan terjadinya suatu kelas setelah data yang diamati dipertimbangkan. Nilai ini diperoleh melalui perhitungan pada Persamaan 2.9.

$$P(C_{ringan}|X) = \frac{8,09401866 \times 0,8384}{7,8096138} = \frac{6,78602524}{7,8096138} = 0,868932243 \\ P(C_{berat}|X) = \frac{6,3340876 \times 0,1616}{7,8096138} = \frac{1,02358856}{7,8096138} = 0,131067757$$

Hasil dari perhitungan probabilitas Posterior dalam menentukan tingkat keparahan pada penderita penyakit ISPA dapat dilihat pada tampilan Gambar 5.

```
=== Posterior Probabilities ===
P(0|x) = 0.8380557376
P(1|x) = 0.1619442624
```

**Gambar 5** Hasil Probabilitas Posterior

Gambar 5 menampilkan dari program dalam proses perhitungan Probabilitas Posterior dari satu sampel data prediksi model memprediksi kelas berat dengan probabilitas tertinggi dibandingan kelas ringan.

## 3.7 Hasil Evaluasi Algoritma Naive Bayes

Pada penelitian memakai algoritma Naive Bayes dengan varian Gaussian Naive Bayes yang berfokus pada pengukuran performa model menggunakan Confusion Matrix yaitu Accuracy, Precision, Recall, dan F1-Score dalam proses klasifikasi penyakit ISPA. Kemudian dilakukan menggunakan metode 10-Fold Cross Validation untuk memastikan hasil yang lebih stabil, selanjutnya dalam proses klasifikasi penyakit ISPA dilakukan secara otomatis menggunakan program untuk memastikan evaluasi model yang lebih efektif. berikut adalah hasil akurasi Naive Bayes dalam mengklasifikasikan penyakit ISPA dapat dilihat pada Tabel 13.

**Tabel 13** Hasil Rata-Rata Akurasi

Total Value	TP	FP	FN	TN	Average Accuracy	Precision	Recall	F1 Score
Each Fold	19	29	45	203	75,00%	78,12%	75,00%	76,36%

Pada Tabel 13 menunjukkan hasil evaluasi yang didapatkan dari masing-masing Fold, Nilai 19, 29, 45, dan 203 merupakan hasil penjumlahan dari seluruh TP, FP, FN dan TN pada masing-masing Fold, sebagaimana terlihat pada Tabel 13, setelah semua dihitung akurasi yang diperoleh menggunakan rumus pada Persamaan 2.2 sebesar 75,00%, Precision 78,12%, Recall 75,00%, dan F1-Score 76,36%.

### 3.8 Pembahasan

Penelitian ini dilakukan untuk mengevaluasi kinerja algoritma *Naive Bayes* pada klasifikasi penyakit ISPA, dengan menggunakan dataset yang diperoleh dari UPT Puskesmas Bontang Barat periode 2024-2025 proses penelitian diawali dengan pengumpulan data kemudian tahap selanjutnya adalah *Pre-Processing* yang mencakup tahapan data *Selection*, *Cleaning*, dan *Transformation*. Selanjutnya untuk memvalidasi kinerja model dengan memakai metode *10-Fold Cross Validation* menjadi sepuluh subset untuk memastikan hasil yang stabil. Setelah proses validasi tersebut, dilakukan evaluasi performa model menggunakan *Confusion Matrix* untuk melihat detail jumlah prediksi tiap kelas, sehingga dapat diperoleh *Matrix* evaluasi seperti *Accuracy*, *Presisi*, *Recall* maupun *F1-Score* sebagai gambaran terhadap kemampuan model dalam melakukan klasifikasi. Setelah melalui serangkaian proses dan memperoleh hasil akhir, pada tahap akhir ini akan membahas hasil yang diperoleh dan dikaitkan kembali dengan rumusan masalah yang ditetapkan diawal. Pembahasan ini menjawab pertanyaan mengenai bagaimana nilai evaluasi yang didapatkan dari penerapan *Naive Bayes* dalam klasifikasi penyakit ISPA. Hasil penerapan algoritma *Naive Bayes* yang menunjukkan rata-rata *Accuracy* 75,00%, *Precision* 78,12%, *Recall* 75,00%, dan *F1-Score* 76,36% nilai ini relatif lebih baik dibandingkan penelitian [17] yang mengklasifikasikan penyakit ISPA menggunakan algoritma *KNN* yang memperoleh *Accuracy* 71%, *Decision Tree* 68%, *Random Forest* 72%, dan *Artificial Neural Network* (ANN) 67%. Temuan ini membuktikan penerapan algoritma *Naive Bayes* dalam penelitian ini dinilai cukup baik digunakan untuk klasifikasi penyakit ISPA dibandingkan penelitian sebelumnya.

## 4. KESIMPULAN

Penerapan algoritma *Naive Bayes* dalam klasifikasi penyakit ISPA menunjukkan performa yang cukup baik. Berdasarkan hasil evaluasi menggunakan metode *10-Fold Cross Validation*, algoritma ini menghasilkan nilai rata-rata *Accuracy* sebesar 75,00%, *Precision* sejumlah 78,12%, *Recall* sebesar 75,00% maupun *F1-Score* sejumlah 76,36%. Nilai evaluasi tersebut menunjukkan bahwa model mampu mengklasifikasikan tingkat keparahan penyakit ISPA secara cukup akurat dan stabil. Dibandingkan dengan penelitian sebelumnya memakai algoritma lain seperti *KNN*, *Decision Tree*, dan *ANN*, hasil ini relatif lebih unggul. Oleh karena itu, disimpulkan algoritma *Naive Bayes* layak digunakan untuk mendukung proses klasifikasi penyakit ISPA, khususnya dalam konteks prediksi keparahan berdasarkan data rekam medis pasien.

## REFERENCES

- [1] M. Napiah, R. A. Purnama, M. Raharjo, And W. Bismi, "Komparasi Algoritma Untuk Klasifikasi Penyakit Ispa (Infeksi Saluran Pernapasan Akut)," *J. Infortech*, Vol. 4, No. 2, Pp. 105–110, 2022.
- [2] R. E. M. Sormin, M. B. Ria, And M. S. Nuwa, "Hubungan Tingkat Pengetahuan Ibu Dengan Perilaku Pencegahan Ispa Pada Balita," *J. Ilm. Kesehat. Media Husada*, Vol. 12, No. 1, Pp. 74–80, 2023, Doi: 10.33475/Jikmh.V12i1.316.
- [3] C. J. Rahmawati, W. C. Rachmawati, And S. Adi, "Gambaran Literasi Kesehatan Dan Pola Pencarian Pengobatan Mahasiswa Ilmu Kesehatan Masyarakat Universitas Negeri Malang Terhadap Penyakit Ispa," *Sport Sci. Heal.*, Vol. 7, No. 1, Pp. 28–41, 2025, Doi: 10.17977/Um062v7i12025p28-41.
- [4] A. M. Oktafani, I. T. Utami, And P. Kartikasari, "Optimasi Backward Elimination Pada Klasifikasi Penyakit Ispa Menggunakan Algoritma *Naive Bayes Classifier* 1,2,3," *J. Gaussian*, Vol. 14, Pp. 23–30, 2025, Doi: 10.14710/J.Gauss.14.1.23-30.
- [5] D. Puspitasari, "Survei Kesehatan Indonesia (Ski) 2023 Dalam Angka," *Ski (Survei Kesehat. Indones.)*, Pp. 1–68, 2023.
- [6] J. R. Saputri, M. G. Sholih, And R. R. Dwiputra, "Identifikasi Faktor Resiko Dan Terapi Infeksi Saluran Pernafasan Akut (Ispa) Pada Balita: Literature Review," *J. Surya Med.*, Vol. 11, No. 2, Pp. 145–151, 2025, Doi: 10.33084/Jsm.V11i2.9739.
- [7] B. Sulistiyo, B. Surarso, And W. A. Syafei, "Sistem Pakar Identifikasi Dan Alternatif Solusi Terhadap Permasalahan Yang Dihadapi Oleh Peserta Didik Sekolah Menengah Menggunakan Rule-Based Machine Learning," *Suparyanto Dan Rosad (2015)*, Vol. 5, No. 3, Pp. 248–253, 2022, Doi: 10.14710/Jtsiskom.2022.Xxxxx.
- [8] B. Karyadi, "Pemanfaatan Kecerdasan Buatan Dalam Mendukung Pembelajaran Mandiri," *Educ. J. Teknol. Pendidik.*, Vol. 8, No. 2, Pp. 253–258, 2023, Doi: 10.32832/Educate.V8i02.14843.
- [9] A. Surahmat *Et Al.*, *Kecerdasan Buatan Dalam Data Mining*. Widina Bhakti Persada Bandung, 2023.
- [10] H. Abijono, P. Santoso, And N. L. Anggreini, "Algoritma Supervised Learning Dan Unsupervised Learning Dalam

- Pengolahan Data,” *J. Teknol. Terap. G-Tech*, Vol. 4, No. 2, Pp. 315–318, 2021, Doi: 10.33379/Gtech.V4i2.635.
- [11] K. Kristiawan And A. Widjaja, “Perbandingan Algoritma Machine Learning Dalam Menilai Sebuah Lokasi Toko Ritel,” *J. Tek. Inform. Dan Sist. Inf.*, Vol. 7, No. 1, Pp. 35–46, 2021, Doi: 10.28932/Jutisi.V7i1.3182.
- [12] M. R. Ahmmed, J. Monir, And S. A. Khushbu, “Analysis Of Flood Risk Prediction Using Different Machine Learning Classifiers: A Study Of Predicting Flood Risk In Rural Areas, Bangladesh,” *2022 13th Int. Conf. Comput. Commun. Netw. Technol. Icccnt 2022*, Pp. 1–6, 2022, Doi: 10.1109/Icccnt54827.2022.9984449.
- [13] F. Zahri, F. Insani, And L. Oktavia, “Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor ( K- Nn ) Untuk Klasifikasi Penyakit Gagal Jantung,” *J. Inf. Syst. Res.*, Vol. 6, No. 2, Pp. 997–1005, 2025, Doi: 10.47065/Josh.V6i2.6480.
- [14] B. Delvika, S. Nurhidayarnis, P. D. Rinanda, N. Abror, And A. Hidayat, “Perbandingan Klasifikasi Antara Naive Bayes Dan K-Nearest Neighbor Terhadap Resiko Diabetes Pada Ibu Hamil,” *Malcom Indones. J. Mach. Learn. Comput. Sci.*, Vol. 2, No. 2, Pp. 68–75, 2022, Doi: 10.57152/Malcom.V2i2.432.
- [15] G. Dwilestari And T. A. Afifah, “Perbandingan Kinerja Algoritma Naive Bayes Dan Decision Tree Dalam Klasifikasi Kanker Paru-Paru,” *Jati*, Vol. 9, No. 1, Pp. 801–807, 2025.
- [16] R. Pranandito And H. Hendry, “Perbandingan Prediksi Penyakit Serangan Jantung Menggunakan Model Machine Learning,” *Jipi (Jurnal Ilm. Penelit. Dan Pembelajaran Inform.)*, Vol. 8, No. 4, Pp. 1228–1237, 2023, Doi: 10.29100/Jipi.V8i4.4165.
- [17] H. M. Fenta, T. T. Zewotir, S. Naidoo, R. N. Naidoo, And H. Mwambi, “Factors Of Acute Respiratory Infection Among Under-Five Children Across Sub-Saharan African Countries Using Machine Learning Approaches,” *Sci. Rep.*, Vol. 14, No. 1, Pp. 1–14, 2024, Doi: 10.1038/S41598-024-65620-1.
- [18] E. Sufarnap And Sudarto, “Seleksi Fitur Information Gain Dalam Meningkatkan Kinerja Naïve Bayes Dalam Prediksi Performa Mahasiswa,” *Digit. Transform. Technol.*, Vol. 4, No. 1, Pp. 1285–1295, 2024, Doi: <https://doi.org/10.47709/Digitech.V4i2.5525> Seleksi.
- [19] T. Syamsudin, T. Handhayani, ) Muhammad, And I. Syaifudin, “Perbandingan Klasifikasi Penyakit Diabetes Menggunakan Metode Machine Learning,” *J. Ilmu Komput. Dan Sist. Inf.*, Pp. 1–7, 2021, Doi: 10.24912/Jiksi.V12i1.28197.
- [20] Q. A. Puteri, T. Sagirani, And J. Lemantara, “Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor (Knn) Untuk Mengetahui Keakuratan Diagnosa Penyakit Diabetes,” *J. Nas. Teknol. Dan Sist. Inf.*, Vol. 9, No. 3, Pp. 247–254, 2023, Doi: 10.25077/Teknosi.V9i3.2023.247-254.
- [21] D. E. Yanti, L. Framesti, And A. Desiani, “Perbandingan Algoritma C4.5 Dan Svm Dalam Klasifikasi Penyakit Anemia,” *Jip (Jurnal Inform. Polinema)*, Pp. 427–434, 2022, Doi: <https://doi.org/10.33795/Jip.V9i4.1381>.
- [22] B. Hirwono, A. Hermawan, And D. Avianto, “Implementasi Metode Naïve Bayes Untuk Klasifikasi Penderita Penyakit Jantung,” *J. Jtik (Jurnal Teknol. Inf. Dan Komunikasi)*, Vol. 7, No. 3, Pp. 450–457, 2023, Doi: 10.35870/Jtik.V7i3.910.
- [23] A. Karima And T. A. Y. Siswa, “Prediksi Kinerja Mahasiswa Dalam Perkuliahan Berbasis Learning Management System Menggunakan Algoritma Naïve Bayes,” *Progresif J. Ilm. Komput.*, Vol. 18, No. 2, P. 211, 2022, Doi: 10.35889/Progresif.V18i2.922.
- [24] C. Salsabila Azra, F. Yulianto, And T. A. Y. Siswa, “Implementasi Metode Naive Bayes Untuk Klasifikasi Kecelakaan Lalu Lintas Di Kota Samarinda,” *Jitet*, Vol. 13, No. 1, 2025, Doi: <http://dx.doi.org/10.23960/Jitet.V13i1.5890> Implementasi.
- [25] A. Mustofa, O. Okfalisa, E. P. Cynthia, Y. Yelfi, And S. K. Gusti, “Klasifikasi Penerima Bantuan Covid-19 Menggunakan Metode Weighted K-Nearest Neighbour,” *J. Nas. Komputasi Dan Teknol. Inf.*, Vol. 5, No. 3, Pp. 415–422, 2022, Doi: 10.32672/Jnkti.V5i3.4399.