

Klasifikasi Kondisi Penyakit Asma Menggunakan Algoritma *Naïve Bayes*

Relin Pramudiya¹, Ery Hartati^{2*}

¹⁻²Program Studi Informatika, Fakultas Ilmu Komputer dan Rekayasa, Universitas Multi Data

¹⁻²Jl. Rajawali 14, 30113 Palembang - Sumatera Selatan

Email: relinrp@mhs.mdp.ac.id¹, ery_hartati@mdp.ac.id²

Abstract

Asthma is a chronic respiratory disease characterized by symptoms such as shortness of breath, coughing, and wheezing, and can be life-threatening if not properly treated. Delayed diagnosis and difficulty in assessing symptom severity are often the main causes of serious complications. This study aims to develop a classification model for asthma conditions using the Naïve Bayes algorithm based on clinical symptoms and patient demographic data. The dataset used was obtained from the public Kaggle platform, consisting of 316,800 samples with 19 attributes. The research stages included data pre-processing, model training, application of the Synthetic Minority Oversampling Technique (SMOTE), and model performance evaluation using accuracy, precision, recall, and F1-score metrics. The initial classification results showed an accuracy of 75.19%, which increased to 83.59% after applying SMOTE. The F1-score also improved from 77% to 83%. These findings indicate that the Naïve Bayes algorithm is capable of classifying asthma conditions reliably, quickly, and efficiently. Therefore, the model is considered suitable to be implemented as an initial classification system to support timely and accurate clinical diagnosis of asthma.

Key Words: *Asthma, Classification, Clinical Diagnosis, Naïve Bayes, SMOTE*

Abstrak

Asma merupakan penyakit kronis pada saluran pernapasan yang ditandai dengan gejala sesak napas, batuk, dan mengi, serta berpotensi mengancam nyawa jika tidak ditangani secara tepat. Keterlambatan *diagnosis* dan kesulitan dalam menilai kondisi gejala sering menjadi penyebab utama komplikasi serius. Penelitian ini bertujuan untuk mengembangkan model klasifikasi kondisi penyakit asma menggunakan algoritma *Naïve Bayes* berbasis data gejala klinis dan demografis pasien. *Dataset* yang digunakan bersumber dari *platform* publik *Kaggle*, terdiri dari 316.800 sampel dengan 19 atribut. Tahapan penelitian meliputi *pre-processing* data, pelatihan model, penerapan teknik *Synthetic Minority Oversampling Technique* (SMOTE), serta evaluasi kinerja model menggunakan metrik akurasi, presisi, recall, dan *F1-score*. Hasil klasifikasi awal menunjukkan akurasi sebesar 75,19%, yang kemudian meningkat menjadi 83,59% setelah penerapan SMOTE. Nilai *F1-score* juga meningkat dari 77% menjadi 83%. Temuan ini menunjukkan bahwa algoritma *Naïve Bayes* mampu melakukan klasifikasi kondisi asma secara andal, cepat, dan efisien. Dengan demikian, model ini layak digunakan sebagai sistem klasifikasi awal untuk mendukung proses diagnosis klinis penyakit asma secara tepat waktu dan akurat.

Kata Kunci: *Asma, Diagnosis Klinis, Klasifikasi, Naïve Bayes, SMOTE*

Pendahuluan

Asma merupakan penyakit kronis pada saluran pernapasan yang menimbulkan gejala berulang seperti sesak napas, batuk, dan mengi, serta dapat mengancam nyawa apabila tidak ditangani dengan tepat [1]. Menurut *WHO*, lebih dari 235 juta orang di seluruh dunia menderita asma, dan pada tahun 2025 jumlah ini diperkirakan meningkat hingga 400 juta orang [2]. Di Indonesia, jumlah kasus asma mencapai lebih dari 12 juta jiwa, atau sekitar 4,5% dari total populasi. Faktor utama penyebab kematian akibat asma adalah keterlambatan *diagnosis*, kesulitan dalam menilai kondisi gejala, dan pengobatan yang tidak tepat [3].

Proses *diagnosis* asma yang bergantung pada ingatan dan pengalaman dokter, serta keterbatasan waktu dan biaya pelayanan kesehatan, membuat pentingnya dukungan sistem yang mampu membantu identifikasi gejala asma secara efektif [4]. Beberapa penelitian menunjukkan bahwa faktor pemicu asma dapat bervariasi, mulai dari infeksi saluran napas, alergen, asap rokok, hingga stres dan perubahan cuaca

*Corresponding Author: Ery Hartati

[5], sehingga diperlukan sistem yang mampu mengenali pola kompleks dari gejala dan pemicu tersebut [6].

Beberapa penelitian sebelumnya menunjukkan bahwa teknologi kecerdasan buatan dapat membantu dalam meningkatkan akurasi *diagnosis* dan klasifikasi penyakit pernapasan, termasuk asma. Penelitian menggunakan jaringan saraf tiruan (JST) dengan metode *backpropagation* dalam memklasifikasi asma berdasarkan data klinis pasien dan berhasil mencapai tingkat akurasi yang baik dengan nilai *MSE* yang rendah, menunjukkan potensi metode ini dalam klasifikasi tingkat kondisi asma [7]. Penelitian lainnya juga merancang sistem deteksi asma, *bronkitis*, dan *pneumonia* melalui analisis suara pernapasan menggunakan metode *long short-term memory* (LSTM). Meskipun akurasi model masih terbatas pada 60%, penelitian ini memberikan kontribusi dalam pengembangan sistem deteksi dini berbasis suara [8].

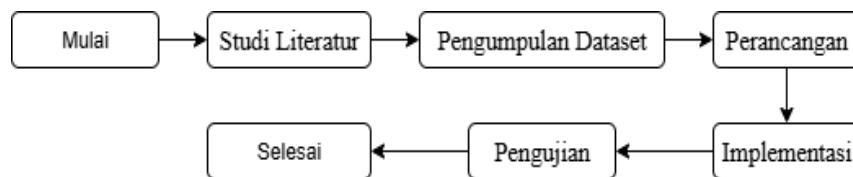
Penelitian yang mengembangkan sistem klasifikasi suara pernapasan asma dengan metode *k-Nearest Neighbors* (*k*-NN), yang berhasil mencapai akurasi klasifikasi hingga 86,6%, menunjukkan efektivitas pendekatan berbasis audio dalam identifikasi gejala asma [9]. Sementara itu, penelitian yang membandingkan regresi decision tree dan linear berganda untuk klasifikasi *body mass index* (BMI) pada pasien asma, dengan hasil bahwa regresi linear berganda menunjukkan performa lebih baik dalam klasifikasi berbasis data klinis [10]. Penelitian lainnya menerapkan algoritma *Random Forest* dalam klasifikasi penyakit paru-paru, termasuk asma, dan memperoleh nilai *AUC* sebesar 0,993, yang menandakan tingkat akurasi sangat tinggi [11].

Salah satu solusi yang dapat diterapkan adalah pemanfaatan algoritma *machine learning* untuk membangun sistem klasifikasi kondisi asma berdasarkan data klinis dan gejala pasien [12]. Dalam hal ini, metode *Naïve Bayes* merupakan salah satu pendekatan efisien karena keterampilannya dalam mengklasifikasikan data dengan jumlah fitur yang banyak namun tetap mempertahankan kesederhanaan model [13]. *Naïve Bayes* juga dikenal efektif dalam menangani masalah klasifikasi medis karena mampu mengestimasi probabilitas dari setiap kelas berdasarkan data *historis* gejala pasien, meskipun asumsi independensi antar fitur seringkali tidak terpenuhi secara sempurna [14].

Penelitian ini mengusulkan pengembangan model klasifikasi kondisi penyakit asma menggunakan metode *Naïve Bayes*. Metode ini digunakan karena keterampilannya dalam mengklasifikasikan data dengan efisien serta memberikan estimasi probabilitas berdasarkan data historis gejala pasien. Dengan memanfaatkan pendekatan ini, diharapkan hasil klasifikasi dapat membantu dalam proses penilaian kondisi asma secara lebih cepat dan akurat. Hal ini penting untuk diterapkan sebagai sistem klasifikasi awal dalam diagnosis penyakit asma secara tepat waktu, sehingga dapat menurunkan risiko keterlambatan penanganan dan mengurangi angka kematian akibat serangan asma yang parah.

Metode Penelitian

Bagian ini menjelaskan langkah-langkah sistematis yang dilakukan dalam penelitian, mulai dari studi literatur hingga pengujian model. Penelitian ini menerapkan algoritma *Naïve Bayes* dalam klasifikasi kondisi penyakit asma berdasarkan data gejala klinis. Pada Gambar 1 dapat dilihat tahapan metodologi penelitian ini.



Gambar 1 Tahapan Metodologi

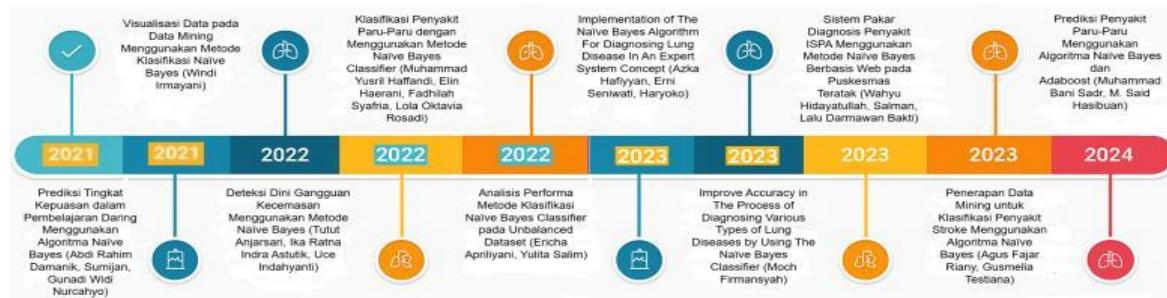
A. Studi Literatur

Tahapan ini diawali dengan pengumpulan dan kajian terhadap literatur seperti jurnal ilmiah, artikel, dan buku yang berkaitan dengan penyakit asma, klasifikasi kondisi asma, serta penerapan algoritma

Naïve Bayes dalam bidang medis. Studi ini bertujuan untuk memahami konsep dasar dari algoritma yang digunakan, metode yang relevan, serta kelebihan dan keterbatasannya dalam konteks klasifikasi medis [15].

Beberapa penelitian sebelumnya menunjukkan hasil penggunaan algoritma *Naïve Bayes* untuk klasifikasi penyakit, khususnya di bidang paru-paru. Penelitian lainnya mengembangkan model klasifikasi berbagai jenis penyakit paru seperti asma, *bronkitis*, dan *pneumonia* menggunakan algoritma *Naïve Bayes*, dan memperoleh akurasi sebesar 90,22% dalam proses diagnosis [16]. Penelitian menggunakan metode serupa untuk klasifikasi enam jenis penyakit paru di RSUD Mayjen H.A. Thalib dan berhasil mencapai akurasi hingga 97,06%, yang menunjukkan efektivitas metode ini bahkan pada jumlah data yang relatif kecil [17]. Penelitian lain oleh mengombinasikan algoritma *Naïve Bayes* dan metode *ensemble Adaboost* untuk menaikkan akurasi klasifikasi penyakit paru, dengan hasil akurasi mencapai 94,66%, lebih tinggi dibandingkan penggunaan *Naïve Bayes* tunggal [18]. Penelitian yang menerapkan metode ini dalam sistem pakar berbasis web untuk *diagnosis ISPA* dan memperoleh akurasi sebesar 92,3%, menunjukkan bahwa metode ini juga efektif dalam sistem berbasis keputusan [19]. Secara keseluruhan, berbagai studi tersebut mengindikasikan bahwa algoritma *Naïve Bayes* merupakan pendekatan klasifikasi yang andal, sederhana, dan efisien dalam konteks *diagnosis* penyakit berbasis gejala.

Berdasarkan tinjauan literatur, dapat disimpulkan bahwa penggunaan algoritma *Naïve Bayes* dalam klasifikasi penyakit pernapasan terus mengalami peningkatan, baik dari segi akurasi maupun cakupan aplikasi. Metode ini telah digunakan dalam klasifikasi berbagai penyakit paru dan ISPA, serta mulai dikombinasikan dengan teknik lain seperti *Adaboost* dan sistem pakar berbasis web. Dapat dilihat pada gambar 2 berikut menampilkan kronologi penelitian dari tahun 2021 hingga 2024, yang menunjukkan tren positif penggunaan *Naïve Bayes* dalam bidang kesehatan.



Gambar 2 Roadmap Penelitian

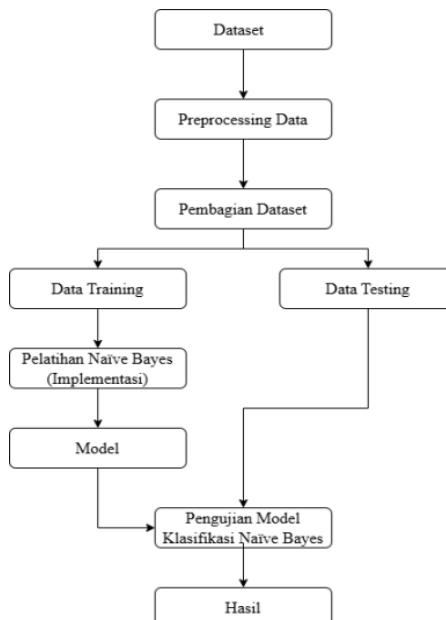
B. Pengumpulan Dataset

Dataset yang dipakai dalam penelitian ini berasal dari website berbagi data publik *Kaggle*, yang dapat diakses melalui link: <https://www.Kaggle.com/Datasets/deepayanthurakur/asthma-disease-prediction/data>. *Dataset* ini dirancang khusus untuk keperluan klasifikasi penyakit asma dan berisi informasi yang komprehensif mengenai kondisi pasien. Secara keseluruhan, *Dataset* ini memuat sebanyak 316.800 sampel data yang terdiri dari berbagai kombinasi variabel klinis dan *demografis*. Terdapat 19 atribut dalam *Dataset* ini, yang mencakup sejumlah parameter penting seperti jenis kelamin, usia, serta berbagai gejala klinis seperti batuk, sesak napas, dan kelelahan. Atribut-atribut ini diharapkan mampu memberikan kontribusi signifikan dalam proses klasifikasi kondisi penyakit asma. Keberagaman dan jumlah data yang besar ini menjadikan *Dataset* ini sangat relevan dan bermanfaat sebagai dasar dalam membangun model klasifikasi berbasis *machine learning*, khususnya dengan pendekatan algoritma *Naïve Bayes* [20].

C. Perancangan

Pada tahap ini, sistem klasifikasi dirancang menggunakan algoritma *Naïve Bayes*. Tahapan diawali dengan *preprocessing* data, termasuk pembersihan data dari duplikasi atau entri tidak valid, penanganan *missing value* dengan metode statistik seperti pengisian berdasarkan nilai modus, serta normalisasi untuk menyamakan skala atribut [21].

Setelah data diproses, kumpulan *Dataset* dibagi secara proporsional menjadi data pelatihan dan data pengujian, dengan komposisi 80% data pelatihan dan 20% data pengujian, untuk melatih dan mengevaluasi kinerja model klasifikasi *Naïve Bayes* yang dikembangkan dalam penelitian ini [22]. Model ini dirancang untuk mengklasifikasikan kondisi penyakit asma berdasarkan fitur-fitur klinis yang tersedia. Pada Gambar 3 dapat dilihat Skema perancangan.



Gambar 3 Skema Perancangan *Naïve Bayes*

D. Implementasi

Tahapan ini mencakup implementasi algoritma *Naïve Bayes* berdasarkan rancangan sebelumnya. Model dilatih untuk memklasifikasi kondisi asma berdasarkan kombinasi fitur input. Perhitungan dilakukan dengan menggunakan teorema bayes untuk menghitung probabilitas posterior setiap kelas (ringan, sedang, dan tidak ada gejala).

Metode *Naïve Bayes* adalah algoritma klasifikasi berbasis probabilistik yang menggunakan Teorema Bayes dengan asumsi independensi antar fitur [23]. Meskipun asumsi ini umumnya tidak sepenuhnya terpenuhi dalam data dunia nyata, metode *Naïve Bayes* telah terbukti efektif dalam berbagai masalah klasifikasi, khususnya di bidang medis.

Adapun Teorema Bayes dapat dilihat pada Persamaan 1.

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

Keterangan:

- $P(C|X)$: Probabilitas kelas C (misanya kondisi asma) diberikan fitur X
- $P(X|C)$: Probabilitas fitur X diberikan kelas C
- $P(C)$: Probabilitas apriori kelas C
- $P(X)$: Probabilitas apriori fitur X

Naïve Bayes merupakan algoritma yang sangat cocok digunakan dalam klasifikasi berbasis gejala klinis karena memiliki beberapa keunggulan. Algoritma ini cepat dan efisien dalam proses pelatihan, sehingga mampu menghasilkan model dalam waktu singkat meskipun dengan jumlah data yang besar [24]. Selain itu, *Naïve Bayes* tidak sensitif terhadap fitur yang tidak relevan, sehingga tetap mampu memberikan hasil klasifikasi yang baik meskipun terdapat fitur yang kurang berkontribusi terhadap klasifikasi [25]. Keunggulan lainnya adalah kemampuannya dalam menangani *Dataset* berukuran besar dengan banyak fitur, yang sering dijumpai dalam data klinis pasien, sehingga menjadikannya pilihan yang tepat dalam konteks klasifikasi medis.

E. Pengujian

Setelah implementasi model algoritma *Naïve Bayes*, dilakukan proses pengujian untuk mengevaluasi performa model klasifikasi. Evaluasi dilakukan dengan menggunakan metrik akurasi, *presisi*, *recall* (*Sensitivity*), dan *F1-score* [26]. Dapat dilihat pada persamaan 2, 3, 4, dan 5 metrik evaluasi secara umum.

$$\text{Akurasi} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \quad \dots \dots \dots \quad (2)$$

$$\text{Presisi} = \frac{T_p}{T_p + F_p} \quad \dots \dots \dots \quad (3)$$

$$\text{Recall} = \frac{T_p}{T_p + F_n} \quad \dots \dots \dots \quad (4)$$

$$F1 - score = 2 \times \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}} \quad \dots \dots \dots \quad (5)$$

Keterangan:

- T_p : True Positive
- T_n : True Negative
- F_p : False Positive
- F_n : False Negative

Setelah evaluasi dilakukan, hasil performa model dianalisis untuk menarik kesimpulan terkait efektivitas algoritma *Naïve Bayes* dalam mengklasifikasikan kondisi penyakit asma serta dibandingkan dengan literatur relevan lainnya.

Hasil dan Pembahasan

Pada tahap ini menyajikan hasil implementasi sistem klasifikasi kondisi penyakit asma menggunakan algoritma *Naïve Bayes*. Proses yang dilakukan meliputi tahapan persiapan data, *pre-processing*, penerapan algoritma, klasifikasi, serta evaluasi dan validasi model. Validasi dilakukan dengan membagi *dataset* menjadi data latih dan data uji menggunakan skema 80:20. Pengujian model dilakukan menggunakan metrik evaluasi seperti akurasi, *presisi*, *recall*, dan *F1-score*, guna menilai efektivitas metode dalam mengklasifikasikan tingkat kondisi penyakit asma berdasarkan atribut klinis dan demografis pasien [27].

A. Persiapan Data

Tahapan awal dalam penelitian ini adalah proses pengumpulan dan penyiapan *Dataset*. *Dataset* yang dimanfaatkan diambil dari *platform* publik Kaggle dengan total sebanyak 316.800 sampel data. Setiap entri data merepresentasikan satu kasus pasien yang mencakup 19 atribut, terdiri atas kombinasi atribut *demografis* dan klinis. Atribut-atribut ini antara lain jenis kelamin, usia, dan sejumlah gejala seperti batuk, sesak napas, dan kelelahan. Tiga kolom terakhir dalam *Dataset* digunakan sebagai label atau kelas, yaitu kategori kondisi asma Ringan, Sedang, dan Tidak Ada Gejala. *Dataset* ini kemudian

dianalisis untuk menjamin distribusi kelas yang seimbang agar tidak terjadi bias selama proses pelatihan model [28]. Isi dari *Dataset* asma dapat dilihat pada Gambar 4.

Data Head:							
0	Tiredness	Dry-Cough	Difficulty-in-Breathing	Sore-Throat	None_Sympotn	\	0
1	1	1	1	1	1	0	0
2	1	1	1	1	1	0	0
3	1	1	1	1	1	0	0
4	1	1	1	1	1	0	0
Pains Nasal-Congestion Runny-Nose None_Experiencing Age_0-9 Age_10-19 \							
0	1	1	1	0	1	0	0
1	1	1	1	0	1	0	0
2	1	1	1	0	1	0	0
3	1	1	1	0	1	0	0
4	1	1	1	0	1	0	0
Age_20-24 Age_25-59 Age_60+ Gender_Female Gender_Male Severity_Mild \							
0	0	0	0	0	1	1	1
1	0	0	0	0	1	1	1
2	0	0	0	0	1	1	1
3	0	0	0	0	1	0	0
4	0	0	0	0	1	0	0
Severity_Moderate Severity_None \							
0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0
4	1	0	0	0	0	0	0

Gambar 4 Isi *Dataset* Asma

B. Pre-processing Data

Sebelum dilakukan pelatihan model, data perlu melalui tahapan *pre-processing* untuk memastikan kualitas data yang optimal. Beberapa langkah yang dilakukan pada tahap ini meliputi pemeriksaan duplikasi dan data tidak valid, di mana data duplikat dan entri yang tidak konsisten dihapus untuk menghindari gangguan dalam proses pelatihan. Selanjutnya, penanganan *missing value* dilakukan dengan metode imputasi, yaitu mengisi nilai kosong menggunakan modus untuk atribut kategori, serta rata-rata atau median untuk atribut numerik. Proses normalisasi juga diterapkan guna menghindari dominasi nilai numerik tertentu, sehingga semua fitur berada dalam skala yang sebanding, terutama ketika menggunakan teknik pembobotan dalam klasifikasi. Setelah itu, dilakukan pemisahan antara fitur dan label, di mana fitur diambil dari seluruh kolom kecuali tiga kolom terakhir yang digunakan sebagai label kelas. Terakhir, *Dataset* dibagi menjadi dua bagian, dengan komposisi 80% data pelatihan dan 20% data pengujian, dengan pembagian yang dilakukan secara acak namun tetap mempertahankan proporsi masing-masing kelas agar keseimbangan data tetap terjaga. Dapat dilihat pada Gambar 5 informasi *Dataset* dan pada Gambar 6 visualisasi *missing value*.

Info Dataset:			
<class 'pandas.core.frame.DataFrame'>			
RangeIndex: 316800 entries, 0 to 316799			
Data columns (total 19 columns):			
#	Column	Non-Null Count	Dtype
0	Tiredness	316800	non-null int64
1	Dry-Cough	316800	non-null int64
2	Difficulty-in-Breathing	316800	non-null int64
3	Sore-Throat	316800	non-null int64
4	None_Sympotn	316800	non-null int64
5	Pains	316800	non-null int64
6	Nasal-Congestion	316800	non-null int64
7	Runny-Nose	316800	non-null int64
8	None_Experiencing	316800	non-null int64
9	Age_0-9	316800	non-null int64
10	Age_10-19	316800	non-null int64
11	Age_20-24	316800	non-null int64
12	Age_25-59	316800	non-null int64
13	Age_60+	316800	non-null int64
14	Gender_Female	316800	non-null int64
15	Gender_Male	316800	non-null int64
16	Severity_Mild	316800	non-null int64
17	Severity_Moderate	316800	non-null int64
18	Severity_None	316800	non-null int64

dtypes: int64(19)
memory usage: 45.9 MB
None

Gambar 5 Informasi Umum *Dataset*

Missing Values:	
Tiredness	0
Dry-Cough	0
Difficulty-in-Breathing	0
Sore-Throat	0
None_Sympton	0
Pains	0
Nasal-Congestion	0
Runny-Nose	0
None_Experiencing	0
Age_0-9	0
Age_10-19	0
Age_20-24	0
Age_25-59	0
Age_60+	0
Gender_Female	0
Gender_Male	0
Severity_Mild	0
Severity_Moderate	0
Severity_None	0
dtype: int64	

Gambar 6 Visualisasi Missing Value

Setelah dilakukan klasifikasi awal menggunakan algoritma *Naïve Bayes*, diperoleh hasil yang cukup baik. Namun, analisis distribusi kelas pada dataset menunjukkan adanya ketidakseimbangan data, di mana jumlah sampel pada kelas minoritas jauh lebih sedikit dibandingkan kelas mayoritas. Ketidakseimbangan ini dapat mempengaruhi kinerja model dalam mengklasifikasikan kelas minoritas secara akurat [29]. Oleh karena itu, dilakukan eksperimen tambahan dengan menerapkan teknik *oversampling Synthetic Minority Oversampling Technique* (SMOTE) untuk meningkatkan jumlah sampel pada kelas minoritas sehingga distribusi antar kelas menjadi seimbang sebelum proses klasifikasi dilakukan [30]. Hasil klasifikasi setelah penerapan *SMOTE* menunjukkan peningkatan performa model klasifikasi *Naïve Bayes* yang signifikan dibandingkan dengan hasil sebelumnya.

C. Algoritma *Naïve Bayes*

Algoritma *Naïve Bayes* merupakan metode klasifikasi probabilistik yang sederhana tetapi efektif. Algoritma ini bekerja berdasarkan Teorema Bayes dan memperkirakan bahwa semua fitur atau atribut bersifat independen satu sama lain dalam konteks kelas target tertentu. Meskipun asumsi independensi ini jarang terpenuhi sepenuhnya dalam data klinis nyata, *Naïve Bayes* tetap menjadi pilihan yang baik karena memiliki beberapa keunggulan, antara lain kecepatan dalam proses pelatihan dan klasifikasi, ketangguhan terhadap fitur yang tidak relevan, kinerja yang baik pada *Dataset* berukuran besar, serta kemudahan dalam menginterpretasikan hasil model. Dapat dilihat pada Gambar 7 distribusi kelas sebelum *SMOTE* dan pada Gambar 8 distribusi kelas setelah *SMOTE*.

Distribusi Label:	
Severity_None	
0	237600
1	79200
Name: count, dtype: int64	

Gambar 7 Distribusi Kelas Sebelum *SMOTE*

Distribusi label setelah SMOTE:	
Severity_None	
0	237600
1	237600
Name: count, dtype: int64	

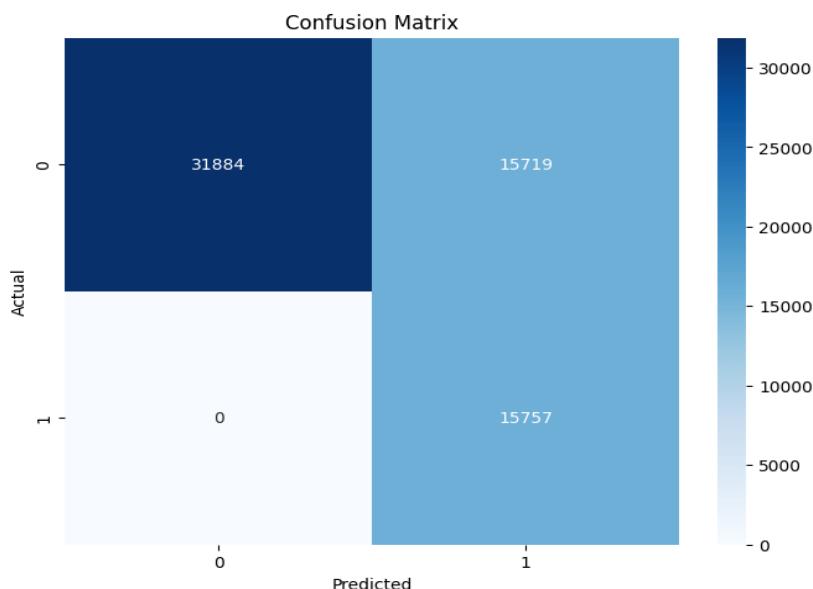
Gambar 8 Distribusi Kelas Setelah *SMOTE*

D. Klasifikasi Menggunakan *Naïve Bayes*

Setelah data siap dan algoritma *Naïve Bayes* diterapkan, dilakukan pelatihan model menggunakan data latih. Proses klasifikasi dilakukan untuk memetakan kombinasi fitur klinis ke dalam salah satu dari tiga kelas tingkat kondisi asma. Model yang dibangun kemudian diterapkan pada data uji untuk mengukur kemampuan klasifikasinya. Setiap entri pada data uji diklasifikasi kelasnya, dan hasil klasifikasi dibandingkan dengan label asli untuk menentukan tingkat akurasi model. Dapat dilihat akurasi *Naïve Bayes* sebelum *SMOTE* pada Gambar 9 dan pada Gambar 10 *confusion matrix* sebelum *SMOTE*.

Accuracy: 75.19%					
Classification Report:					
	precision	recall	f1-score	support	
0	1.00	0.67	0.80	47603	
1	0.50	1.00	0.67	15757	
accuracy			0.75	63360	
macro avg	0.75	0.83	0.73	63360	
weighted avg	0.88	0.75	0.77	63360	

Gambar 9 Akurasi *Naïve Bayes* Sebelum *SMOTE*



Gambar 10 Confusion Matrix Sebelum *SMOTE*

E. Pengujian dan Evaluasi

Pengujian dilakukan untuk mengevaluasi kinerja model klasifikasi menggunakan berbagai metrik evaluasi, yaitu akurasi, *presisi*, *recall*, dan *F1-score*. Akurasi mengukur persentase klasifikasi yang benar dibandingkan dengan seluruh data pengujian, sedangkan *presisi* menunjukkan kemampuan model dalam mengklasifikasikan data positif secara tepat. *Recall*, atau *sensitivitas*, mengukur sejauh mana model mampu mendekripsi seluruh kasus positif yang sebenarnya. Sementara itu, *F1-score* merupakan rata-rata harmonik dari *presisi* dan *recall* yang memberikan gambaran kinerja model secara seimbang. Berdasarkan hasil pengujian terhadap data uji, model *Naïve Bayes* menunjukkan performa yang cukup baik, hasil pengujian awal (sebelum *SMOTE*) dilihat pada Tabel 1.

Tabel 1 Evaluasi Kinerja *Naïve Bayes* Sebelum *SMOTE*

Metrik	Nilai (%)
Akurasi	75.19
<i>Presisi</i>	88.00 (<i>weighted</i>) / 75.00 (<i>macro</i>)
<i>Recall</i>	75.00 (<i>weighted</i>) / 83.00 (<i>macro</i>)
<i>F1-score</i>	77.00 (<i>weighted</i>) / 73.00 (<i>macro</i>)

Hasil evaluasi menghasilkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasi kondisi penyakit asma dengan kinerja yang cukup baik. Akurasi sebesar 75,19% menunjukkan tingkat klasifikasi yang tinggi. *Presisi* yang cukup tinggi (88% secara *weighted*) mengindikasikan bahwa model mampu meminimalkan kesalahan klasifikasi positif.

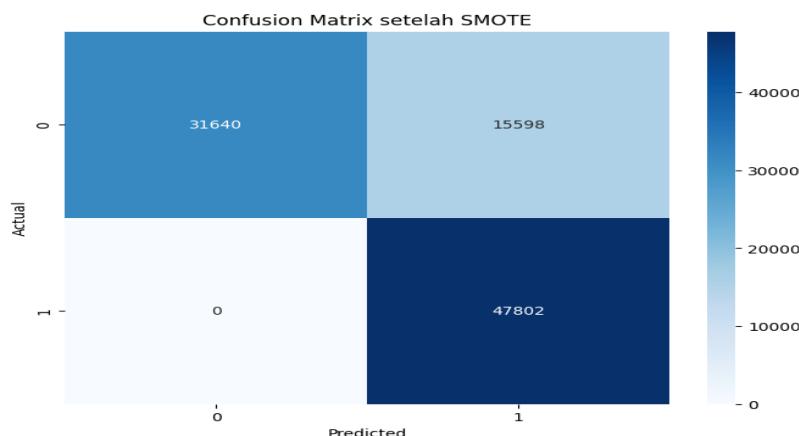
Selanjutnya, setelah menerapkan teknik *SMOTE*, performa model mengalami peningkatan yang signifikan. Hasil evaluasi setelah *SMOTE* ditunjukkan pada Table 2. Sedangkan akurasi *Naïve Bayes* setelah *SMOTE* dapat dilihat pada gambar 11 dan *confusion matrix* setelah *SMOTE* dapat dilihat pada Gambar 12.

Tabel 2 Evaluasi Kinerja *Naïve Bayes* setelah *SMOTE*

Metrik	Nilai (%)
Akurasi	83.59
<i>Presisi</i>	88.00 (<i>weighted</i>) / 88.00 (<i>macro</i>)
<i>Recall</i>	84.00 (<i>weighted</i>) / 83.00 (<i>macro</i>)
<i>F1-score</i>	83.00 (<i>weighted</i>) / 83.00 (<i>macro</i>)

```
Accuracy setelah SMOTE: 83.59%
Classification Report:
precision    recall    f1-score   support
          0       1.00      0.67      0.80     47238
          1       0.75      1.00      0.86     47802
   accuracy                           0.84     95040
    macro avg       0.88      0.83      0.83     95040
  weighted avg       0.88      0.84      0.83     95040
```

Gambar 11 Akurasi *Naïve Bayes* Setelah *SMOTE*



Gambar 12 Confusion Matrix Setelah *SMOTE*

Perbandingan hasil sebelum dan sesudah *SMOTE* menunjukkan bahwa teknik oversampling ini efektif dalam mengatasi ketidakseimbangan data dan meningkatkan performa klasifikasi. Setelah distribusi data diperbaiki menggunakan *SMOTE*, model menunjukkan kemampuan klasifikasi yang lebih stabil dan akurat. Hal ini tercermin dari peningkatan nilai akurasi menjadi 83,59%, *presisi* sebesar 88,00% (baik *weighted* maupun *macro*), *recall* sebesar 84,00% (*weighted*) dan 83,00% (*macro*), serta *F1-score* sebesar 83,00% (*weighted* dan *macro*). Hasil ini menunjukkan bahwa model tidak hanya memiliki akurasi yang lebih tinggi, tetapi juga konsisten dalam mengidentifikasi masing-masing kelas kondisi asma. Dengan demikian, model ini dinilai andal dan layak diterapkan sebagai sistem klasifikasi awal dalam diagnosis klinis penyakit asma.

Kesimpulan

Penelitian ini bertujuan untuk membangun model klasifikasi kondisi penyakit asma menggunakan algoritma *Naïve Bayes* berdasarkan data gejala klinis dan demografis pasien. *Dataset* yang digunakan terdiri dari 316.800 sampel dengan 19 atribut yang mencerminkan gejala dan kondisi pasien. Proses klasifikasi mencakup tahapan *pre-processing* data, pelatihan model, serta evaluasi performa menggunakan metrik akurasi, *presisi*, *recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan kondisi asma ke dalam tiga kategori (Ringan, Sedang, dan Tidak Ada Gejala) dengan performa yang cukup baik. Sebelum dilakukan penyeimbangan data, model memperoleh akurasi sebesar 75,19% dengan nilai *F1-score* sebesar 77% (*weighted*). Namun, setelah diterapkan teknik *oversampling* menggunakan *SMOTE*, kinerja model meningkat signifikan dengan akurasi mencapai 83,59% dan *F1-score* sebesar 83% (baik *macro* maupun *weighted*). Peningkatan performa setelah penerapan *SMOTE* membuktikan bahwa penanganan terhadap ketidakseimbangan data sangat berpengaruh terhadap efektivitas klasifikasi. Secara keseluruhan, model *Naïve Bayes* terbukti efektif, efisien, dan andal dalam mengklasifikasikan kondisi penyakit asma. Oleh karena itu, model ini layak diterapkan sebagai sistem klasifikasi awal untuk mendukung proses diagnosis klinis penyakit asma secara tepat waktu dan akurat.

Daftar Pustaka

- [1] N. D. Rahmawati, I. L. Hilmi, and S. Salman, “Review Analisis Efektivitas dan Risiko Toksisitas Aminofilin pada Pengobatan Penyakit Asma,” *J. Pharm. Sci.*, vol. 6, no. 1, pp. 95–99, 2023, doi: 10.36490/journal-jps.com.v6i1.14.
- [2] J. Putri, L. N. Sangadah, N. W. Mulyati, and R. Fitriani, “Upaya Peningkatan Pengetahuan tentang Penyakit Asma pada Masyarakat,” *Kolaborasi J. Pengabdi. Masy.*, vol. 2, no. 2, pp. 132–140, 2022, doi: 10.56359/kolaborasi.v2i2.80.
- [3] K. H. Manurung and A. E. Syaputra, “Expert System Diagnosis Penyakit Asma Bronkial dengan Certainty Factor pada Klinik Sari Ramadhan Berbasis Web,” *J. Sistim Inf. dan Teknol.*, vol. 5, no. 2, pp. 139–144, 2023, doi: 10.37034/jsisfotek.v5i1.218.
- [4] N. I. Pradasari and R. L. Atimi, “Pemodelan Bayesian Network untuk Prediksi Penyakit Saluran Pernapasan,” *PETIR J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 12, no. 2, pp. 292–302, 2019, doi: 10.33322/petir.v12i2.637.
- [5] J. G. Dandan, M. B. E. Parhusip, and A. Frethernetty, “Literature Review : Gambaran Faktor-Faktor Pencetus Asma Pada Pasien Asma,” *J. Kedokt. Univ. Palangka Raya*, vol. 10, no. 2, pp. 1–5, 2022, doi: 10.37304/jkupr.v10i2.3492.
- [6] S. W. Audria, I. Farikhah, R. M. Saputra, and N. Purwati, “Prediksi Penyakit Paru-Paru Menggunakan Algoritma Naïve Bayes,” *J. Data Sci. Methods Appl.*, vol. 01, no. 01, pp. 33–41, 2025, doi: 10.30873/jodmapps.v1i1.pp33-41.
- [7] D. H. Tanjung, “Jaringan Saraf Tiruan dengan Backpropagation untuk Memprediksi Penyakit Asma,” *Creat. Inf. Technol. J.*, vol. 2, no. 1, pp. 28–38, 2015, doi: 10.24076/citec.2014v2i1.35.
- [8] M. A. Z. Anam and B. H. Prasetio, “Rancang Bangun Sistem Deteksi Asma, Bronkitis, dan Pneumonia Melalui Suara Pernapasan dengan Metode Long Short-Term Memory,” *J. Pengemb.*

- Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 1–9, 2017, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [9] A. Satriadi, I. M. B. Suksmadana, and B. Kanata, “Analisis Suara Pernapasan Paru-Paru Asma Dengan Tidak Asma Menggunakan Metode K Nearest Neighbors,” *Dielektr. (Jurnal Ilm. Kaji. Teor. dan Apl. Tek. Elektro)*, vol. 8, no. 1, pp. 1–7, 2021, doi: 10.29303/dielektrika.v8i1.251.
- [10] A. T. Nurani, A. Setiawan, and B. Susanto, “Perbandingan Kinerja Regresi Decision Tree dan Regresi Linear Berganda untuk Prediksi BMI pada Dataset Asthma,” *J. Sains dan Edukasi Sains*, vol. 6, no. 1, pp. 34–43, 2023, doi: 10.24246/juses.v6i1p34-43.
- [11] D. Kurniawan, M. Wahyudi, L. Pujiastuti, and Sumanto, “Deteksi dan Prediksi Cerdas Penyakit Paru-Paru dengan Algoritma Random Fores,” *Indones. J. Comput. Sci.*, vol. 3, no. 1, pp. 51–56, 2024.
- [12] M. A. Awal *et al.*, “An Early Detection of Asthma Using BOMLA Detector,” *IEEE Access*, vol. 9, pp. 58403–58420, 2021, doi: 10.1109/ACCESS.2021.3073086.
- [13] A. R. Damanik, Sumijan, and G. W. Nurcahyo, “Prediksi Tingkat Kepuasan dalam Pembelajaran Daring Menggunakan Algoritma Naïve Bayes,” *J. Sistim Inf. dan Teknol.*, vol. 3, no. 3, pp. 88–94, 2021, doi: 10.37034/jsisfotek.v3i3.49.
- [14] A. Hafiyyan, E. Seniwati, and Haryoko, “Implementation of The Naïve Bayes Algorithm For Diagnosing Lung Diseases In An Expert System Concept,” *Int. J. Inf. Syst. Technol.*, vol. 7, no. 1, pp. 23–29, 2023, doi: 10.30645/ijistech.v7i1.300.
- [15] I. I. Spyroglou, G. Spöck, A. G. Rigas, and E. N. Paraskakis, “Evaluation of Bayesian classifiers in asthma exacerbation prediction after medication discontinuation,” *BMC Res. Notes*, vol. 11, no. 522, pp. 18–23, 2018, doi: 10.1186/s13104-018-3621-1.
- [16] M. Firmansyah, “Improve Accuracy in The Process of Diagnosing Various Types of Lung Diseases by Using The Naïve Bayes Classifier,” *Int. J. Inf. Syst. Technol.*, vol. 7, no. 2, pp. 99–107, 2023, doi: 10.30645/ijistech.v7i2.305.
- [17] M. Y. Haffandi, E. Haerani, F. Syafria, and L. Oktavia, “Klasifikasi Penyakit Paru-Paru Dengan Menggunakan Metode Naïve Bayes Classifier,” *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, pp. 176–186, 2022, doi: 10.37600/tekinkom.v5i2.649.
- [18] M. B. Sadr and M. S. Hasibuan, “Prediksi Penyakit Paru-Paru Menggunakan Algoritma Naïve Bayes Dan Adaboost,” *J. Teknol. Inf. Magister Darmajaya*, vol. 1, no. 1, pp. 34–38, 2024.
- [19] W. Hidayatullah, Salman, and L. D. Bakti, “Sistem Pakar Diagnosis Penyakit Ispa Menggunakan Metode Naïve Bayes Berbasis Web Pada Puskesmas Teratak,” *J. Kecerdasan Buatan dan Teknol. Inf.*, vol. 2, no. 1, pp. 32–42, 2023, doi: 10.69916/jkbti.v2i1.13.
- [20] 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.
- [21] A. F. Riany and G. Testiana, “Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes,” *J. Saintekom Sains, Teknol. Komputer, dan Manaj.*, vol. 13, no. 1, pp. 42–54, 2023, doi: 10.33020/saintekom.v13i1.352.
- [22] M. Ikbal, S. Andryana, and R. T. K. Sari, “Visualisasi dan Analisa Data Penyebaran Covid-19 dengan Metode Klasifikasi Naïve Bayes,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 5, no. 4, pp. 389–394, 2021, doi: 10.35870/jtik.v5i4.233.
- [23] M. Syawaludin and M. Khulaimi, “Perancangan Sistem Pakar Prediksi Diagnosis Penyakit Diabetes Menggunakan Algoritma Naive Bayes Berbasis Web,” *JIKOM J. Inform. dan Komput.*, vol. 15, no. 1, pp. 161–171, 2025, doi: 10.55794/jikom.v15i1.280.
- [24] F. Anjeli, Y. Maulita, and H. Khair, “Sistem Pakar Mendiagnosa Penyakit Pada Gangguan Pernafasan Menggunakan Metode Naïve Bayes,” *Bridg. J. Publ. Sist. Inf. dan Telekomun.*, vol. 2, no. 4, pp. 01–18, 2024, doi: 10.62951/bridge.v2i4.198.
- [25] T. Anjarsari, I. R. I. Astutik, and U. Indahyanti, “Deteksi Dini Gangguan Kecemasan Menggunakan Metode Naive Bayes,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 4, pp. 1198–1210, 2022, doi: 10.29100/jipi.v7i4.3197.
- [26] R. Fauzan, A. V. Vitianingsih, D. Cahyono, A. L. Maukar, and Y. A. B. Suprio, “Penerapan

- Algoritma Klasifikasi pada Machine Learning untuk Deteksi Phishing,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 2, pp. 531–540, 2025, doi: 10.57152/malcom.v5i2.1968.
- [27] P. Pratiwi, D. Dwifa, A. Desiani, A. Amran, and B. Suprihatin, “Klasifikasi Penyakit Kanker Paru-Paru Menggunakan Algoritma Naïve Bayes dan Iterative Dichotomizer 3 (ID3),” *Electr. J. Rekayasa dan Teknol. Elektro*, vol. 18, no. 1, pp. 69–80, 2024, doi: 10.23960/elc.v18n1.2519.
- [28] Z. A. Mukharyahya, Y. P. Astuti, and O. N. Cahyani, “Perbandingan Naive Bayes dan Support Vector Machine dalam Klasifikasi Tingkat Kemiskinan di Indonesia,” *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 119–128, 2025, doi: 10.29408/edumatic.v9i1.29512.
- [29] A. U. Dullah, P. Utami, and Jumanto, “Asthma Classification Using an Adaptive Boosting Model with SVM-SMOTE Sampling,” *J. Inf. Syst. Explor. Res.*, vol. 3, no. 1, pp. 1–10, 2025, doi: 10.52465/joser.v3i1.486.
- [30] H. Hairani, K. E. Saputro, and S. Fadli, “K-means-SMOTE for handling class imbalance in the classification of diabetes with C4.5, SVM, and naive Bayes,” *J. Teknol. dan Sist. Komput.*, vol. 8, no. 2, pp. 89–93, 2020, doi: 10.14710/jtsiskom.8.2.2020.89-93.