Vol. 10, No. 2, Oktober 2025, Page. 424-436
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Penerapan Algoritma Naive Bayes, Recursive Feature Elimination, dan Adaptive Synthetic Sampling Pada Klasifikasi Penyakit Dermatitis

Wahyu Hidayat [1], Taghfirul Azhima Yoga Siswa[*2], Rofilde Hasudungan[3]

[1,2,3]Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur, Indonesia
Email: [1]2111102441154@umkt.ac.id, [2]tay756@umkt.ac.id, [3]rh219@umkt.ac.id

Abstrak

Dermatitis merupakan salah satu penyakit kulit yang umum terjadi dan menyerang sekitar 5,7 juta orang setiap tahunnya. Di Indonesia, penyakit ini tergolong sebagai salah satu dari tiga besar faktor risiko yang berkontribusi terhadap peningkatan kasus kanker kulit. Penelitian ini bertujuan untuk mengklasifikasikan penyakit dermatitis menggunakan algoritma *Naive Bayes* dengan penerapan teknik seleksi fitur *Recursive Feature Elimination (RFE)* serta penyeimbangan data *Adaptive Synthetic Sampling (ADASYN).* Data penelitian terdiri atas 392 kasus dermatitis dari UPT Puskesmas Bontang Barat tahun 2024, berdasarkan surat persetujuan izin penelitian Nomor B/000.9.2.4/393/PUS-BB/2025, dengan izin etik dan persetujuan dari pihak terkait untuk penggunaan data dalam kegiatan penelitian dan publikasi ilmiah. Validasi model dilakukan menggunakan metode *5-fold cross-validation*, sedangkan evaluasi kinerja model menggunakan *confusion matrix* untuk mengukur akurasi. Hasil penelitian menunjukkan bahwa fitur sistolik, diastolik, umur, berat badan, dan tinggi badan berkontribusi signifikan terhadap proses klasifikasi. Model awal menghasilkan akurasi sebesar 60,15%, meningkat menjadi 66,52% setelah penerapan *ADASYN*, dan mencapai 90,89% ketika *RFE* dan *ADASYN* diterapkan secara bersamaan. Peningkatan akurasi sebesar 24,37% dibandingkan model awal ini membuktikan bahwa penerapan teknik seleksi fitur dan penyeimbangan data dapat meningkatkan kinerja model klasifikasi penyakit dermatitis.

Kata kunci: *ADASYN*, Dermatitis, Klasifikasi, *Naive Bayes, Recursi Feature Elimination (RFE)*

Abstract

Dermatitis is a common skin disease affecting approximately 5.7 million people annually. In Indonesia, this disease is classified as one of the three major risk factors contributing to the increase in skin cancer cases. This study aims to classify dermatitis using the Naive Bayes algorithm with the application of the Recursive Feature Elimination (RFE) feature selection technique and Adaptive Synthetic Sampling (ADASYN) data balancing. The research data consisted of 392 dermatitis cases from the Bontang Barat Community Health Center (UPT) in 2024, based on the research permit approval letter Number B/000.9.2.4/393/PUS-BB/2025, with ethical clearance and approval from related parties for data use in research activities and scientific publications. Model validation was carried out using the 5-fold cross-validation method, while model performance evaluation used a confusion matrix to measure accuracy. The results showed that systolic, diastolic, age, weight, and height features contributed significantly to the classification process. The initial model produced an accuracy of 60.15%, which increased to 66.52% after ADASYN was applied, and reached 90.89% when RFE and ADASYN were applied simultaneously. This 24.37% increase in accuracy compared to the initial model demonstrates that the application of feature selection and data balancing techniques can improve the performance of the dermatitis disease classification model.

Keywords: *ADASYN, Classification, Dermatitis, Naive Bayes, Recursive Feature Elimination (RFE)*

1. PENDAHULUAN

Dermatitis merupakan suatu kondisi kulit yang rata-rata sudah menyerang sekitar 5,7 juta setiap tahunnya. Secara garis besar dermatitis lebih sering terjadi pada remaja dan orang dewasa namun penyakit dermatitis lebih cenderung membaik pada orang paruh baya di atas umur 30 tahun [1]. Penyakit dermatitis di Indonesia cukup banyak ditemui dan penyakit kulit ini telah menduduki tiga besar

penyebab peningkatan jumblah kanker kulit di indonesia yaitu sekitar 192.414 tercatat dan 122.076 kasus baru dan juga 70.338 termasuk kasus lama. Berdasarkan data epidemiologi di Indonesia memperlihatkan bahwa 97% dari 389 kasus penyakit kulit adalah jenis dermatitis kontak, sementara sebanyak 66,3% dari kasus tersebut adalah jenis dermatitis kontak iritan dan 33,7% adalah dermatitis kontak alergi [2]. Ada berbagai macam pendekatan yang telah diterapkan untuk meningkatkan akurasi klasifikasi penyakit kulit ini. Salah satunya adalah pendekatan dengan *machine learning* [3]

Di dalam *machine learning* terdapat berbagai macam metode salah satu metodenya yaitu metode klasifikasi. Metode ini mempunyai fungsi untuk mengklasifikasikan data yang ada kedalam suatu kelas tertentu. Banyak penelitian yang sudah menerapkan metode klasifikasi di dalam dunia kesehatan seperti contoh klasifikasi penyakit kulit dengan menggunakan berbagai macam algoritma. Contohnya seperti penggunaan algoritma *Convolutional Neural Network* mencapai akurasi sebesar 92%, kemudian algoritma *Random Forest* mencapai akurasi 85%, dan algoritma *Support Vector Machine* yang mencapai akurasi sebesar 78% [4]. Selain penyakit kulit algoritma klasifikasi juga dapat menangani klasifikasi berbagai penyakit medis lainnya, seperti penyakit diabetes, menggunakan algorima *Naive Bayes* yang memiliki akurasi sebesar 91,56% dan juga algoritma *Decision Tree* yang memiliki akurasi sebesar 87,01% [5]. Selanjutnya klasifikasi penyakit ginjal kronis yang membandingkan 4 algoritma, yaitu algoritma C4.5 yang mempunyai akurasi sebesar 90,45%, algoritma *K-NN* mempunyai akurasi 91,50%, algortima *Naive Bayes* mempunyai akurasi sebesar 92,92%, dan algoritma *Regresi Logistik* mempunyai akurasi sebesar 80,09% [6].

Naive Bayes merupakan suatu algoritma pengklasifikasian probabilitas yang cukup sederhana yang berfungsi untuk menghitung sekumpulan probabilitas dengan cara menjumlahkan antara frekuensi dan kombinasi nilai dari dataset yang telah diberikan. Dengan menggunakan banyak data *training* maka akurasi yang didapatkan akan semakin tinggi. Contoh hasil pengujian akurasi diperoleh secara sempurna yaitu 100% dengan penggunaan data *training* yang cukup banyak yaitu sekitar 140 [7]. Algoritma *Naive Bayes* dianggap lebih baik untuk melakukan klasifikasi dalam bidang kesehatan dibandingkan dengan algoritma lainnya. Dalam kasus penyakit kanker paru-paru dengan menggunakan perbandingan algoritma *Naive Bayes* dan algoritma *Decision Tree* menunjukkan bahwa algoritma *Naive Bayes* mempunyai akurasi yang cukup tinggi yaitu sebesar 92,47%, sementara itu algoritma *Decision Tree* hanya mempunyai akurasi sekitar 88,17%. Berdasarkan hasil yang ada algoritma *Naive Bayes* lebih cocok untuk penggunaan klasifikasi yang membutuhkan deteksi cepat dan efisien [8] . Meskipun algoritma *Naive Bayes* dapat menghasilkan akurasi yang cukup tinggi namun dari beberapa penelitian yang dilakukan cenderung rata rata masih menggunakan jumblah atribut atau fitur yang terbatas dalam proses klasifikasinya, hal ini bisa membatasi kemampuan dari *Naive Bayes* untuk menangkap kompleksitas data yang lebih besar. jika algoritma dihadapkan dengan data berdimensi tinggi atau yang mempunyai atribut banyak, maka akan menyebabkan terjadinya *overfitting.* Untuk mengatasi masalah tersebut dibutuhkan metode seleksi fitur yang berfungsi untuk mengurangi fitur-fitur yang kurang penting agar model dapat bekerja lebih efektif dan dapat memberikan hasil yang lebih maksimal saat proses klasifikasi. Salah satu seleksi fitur yang paling banyak digunakan yaitu *Recursive Feature Elimination* (*RFE*).

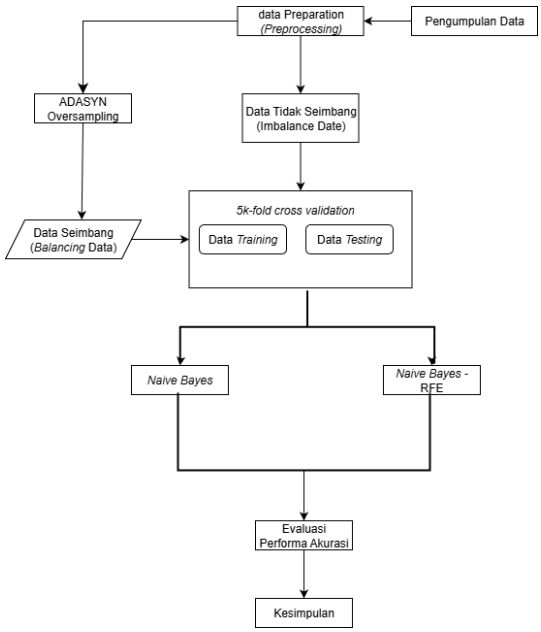
Recursive Feature Elimination merupakan sebuah algoritma seleksi fitur yang sangat baik dibandingkan dengan metode seleksi fitur yang lainnya, *recursive elimination* mempunyai kelebihan dalam hal mempertimbangkan fitur dataset. Cara kerjanya yaitu dengan menghapus fitur-fitur yang sangat tidak penting secara rekursif, recursive elimination bisa secara efektif meredukasi dimensi dataset dan akan tetap mempertahankan fitur-fitur yang paling informatif [9]. Penggunaan metode *Recursive Feature Elimination* (*RFE*) dapat memperkuat proses analisis data yang ada dengan membantu untuk mengidentifikasi fitur-fitur penting yang ada di dalam dataset [10]. sehingga nanti nya fokus di alihkan ke dalam variabel yang paling signifikan dalam menentukan diagnosa penyakit itu sendiri. Penggunaan metode *Recursive Elimination* dapat meningkatkan akurasi, seperti penelitian yang dilakukan oleh

Vol. 10, No. 2, Oktober 2025, Page. 424-436
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

Jaddoa et al,[11] menunjukkan bahwa algoritma *RFE* dapat meningkatkan akurasi model seperti contoh dalam data penyakit hati dengan menggunakan algoritma *Artificial Neural Network* (ANN) dan *AdaBoost*. Model adaBoost sebelum menggunakan *RFE* hanya mempunyai akurasi sebesar 86.74%, sebaliknya dengan menggunakan *RFE* akurasinya meningkat sebesar 89.15%. dan algoritma ANN sebelum menggunakan *RFE* akurasinya hanya sebesar 90.36%, lalu setelah menggunakan *RFE* akurasinya meningkat menjadi 92.77%.

Tantangan yang harus dihadapi dalam penerapan klasifikasi yaitu adalah ketidakseimbangan data [12]. Ketidakseimbangan data bisa terjadi saat jumlah sampel untuk setiap kelas tidak merata, dalam konteks klasifikasi penyakit seperti dermatitis, ketidakseimbangan data pada jumblah kasus dapat menyebabkan algoritma tersebut mengabaikan kelas minoritas, sehingga dengan kondisi ini maka akurasi model bisa menurun [13] masalah ini bisa diatasi dengan penggunaan teknik *oversampling* seperti contoh metode *ADASYN* (*Adaptive Synthetic Sampling*). Cara kerja *ADASYN* adalah dengan cara membuat sampel sintetis dari kelas kelas minoritas berdasarkan distribusi data yang ada dan nantinya akan berfungsi untuk memperbaiki ketidakseimbangan data dengan sangat efektif, Penelitian yang dilakukan oleh [14] yang menganalisis penyakit kanker paru paru dengan menggunakan teknik *oversampling* dan tidak. dalam penelitian tersebut menggunakan teknik *oversampling* dengan *ADASYN* dan *SMOTE* Hasilnya model yang tidak menggunakan teknik *oversampling* hanya mempunyai akurasi sebesar 59%. Sebaliknya ketika model menggunakan teknik *oversampling* seperti *ADASYN*, akurasinya menjadi sangat tinggi, yaitu mencapai 99%, dan penggunaan *oversampling* lainnya seperti *SMOTE* hanya mempunyai akurasi 96%, walaupun *SMOTE* mempunyai dampak yang signifikan, namun nilai akurasinya tidak cukup tinggi dibandingkan dengan *ADASYN*. Oleh karena itu penggunaan *ADASYN* akan sangat bagus untuk menangani ketidakseimbangan data.

2. METODE PENELITIAN



Gambar 1. Alur Penelitian

Gambar 1 menunjukkan alur penelitian yang dimulai dari pengumpulan dan persiapan data. Dataset penyakit dermatitis tahun 2024 dari UPT Puskesmas Bontang Barat digunakan dalam penelitian ini. Karena data bersifat tidak seimbang, dilakukan *oversampling* dengan metode *ADASYN* untuk menyeimbangkan kelas. Selanjutnya, data dibagi menggunakan 5-fold cross-validation dan dianalisis dengan algoritma *Naive Bayes* serta seleksi fitur *RFE*

2.1. Pengumpulan Data

Data yang di gunakan dalam penelitian ini adalah data penyakit dermatitis yang di peroleh dari UPT.Puskesmas Bontang Barat pada tahun 2024. Data terdiri dari 23 kolom dengan 22 atribut dan 1 atribut sebagai class yang dapat berkontribusi untuk proses pengklasifikasikan data penyakit dermatitis.

2.2. Data *Pre-Processing*

Data yang diperoleh dari UPT.Puskesmas Bontang Barat perlu pengolahan data sebelum dimasukan kedalam proses permodelan, untuk menghindari data yang tidak relevan, maka dilakukan data pre-processing dengan berbagai tahap seperti, data *selection*, data *cleaning*, data *transformation*, dengan proses persiapan yang harus dilakukan sebagai berikut :

- **Data Selection**

  Pada tahapan ini proses seleksi dilakukan setelah proses pengumpulan data, setelah data sudah dikumpulkan maka selanjutnya atribut pada dataset akan diseleksi untuk mencari atribut yang diperlukan untuk proses klasifikasi[15]. Proses ini dilakukan secara manual untuk menghapus atribut yang bersifat administratif atau tidak relevan dengan konteks medis, serta atribut dengan kualitas data rendah.

- **Data *Cleaning***

  Data *cleaning* (pembersihan data) adalah proses penghapusan data atau koreksi untuk menghilangkan nilai kosong atau data yang salah, yang dapat menyebabkan terjadinya kesalahan saat membuat model, proses ini penting untuk memastikan nilai keakurasian analisis. Dalam penelitian ini fungsi yang digunakan adalah *library pandas* yang bernama *dropna()* yang fungsinya untuk menghapus baris yang mempunyai unsur nilai *NAN* atau nilai yang tidak terisi dalam suatu baris [16]

- **Data *Transformation***

  Data *Transformation* merupakan sebuah proses mengubah nilai atribut yang bersifat kategorial (berupa data text) menjadi bentuk numerik. Tahap ini sangat penting dilakukan karena *library scikit-learn* hanya dapat memperoses data dengan tipe *numerik*. Beberapa atribut yang perlu di transformasikan meliputi jenis kelamin yang terdiri dari laki laki dan perempuan dan Disease yang terdiri dari kelas ringan dan juga berat. Proses konversi ini menggunakan *library scikit-learn* dengan memanfaatkan fungsi *labelEncorder*. Fungsi tersebut secara otomatis akan merubah data kategorial menjadi nilai numerik dalam satu kolom data.

- **Data Balancing**

  Tahapan dalam pra-pemrosesan data meliputi proses penyeimbangan data. Pada penelitian ini digunakan metode *Adaptive Synthetic Sampling (ADASYN)* untuk mengatasi ketidakseimbangan kelas dengan melakukan *oversampling* pada data *minoritas.* Integritas dan kesesuaian data dijaga dengan memastikan bahwa data sintetis yang dihasilkan tetap mengikuti distribusi dan karakteristik data asli. Proses *ADASYN* diterapkan hanya pada atribut numerik yang relevan dan dengan rasio sampel yang proporsional, sehingga data hasil sintesis tidak menimbulkan bias dan tetap merepresentasikan kondisi sebenarnya dari dataset penyakit dermatitis.

2.3. Pembagian Data

Proses pembagian data dilakukan dengan memisahkan dataset menjadi dua komponen utama, yaitu data latih (*training set*) dan data uji (*test set*). Data latih digunakan untuk melatih model agar dapat memahami pola dan hubungan antar fitur dalam dataset, sedangkan data uji digunakan untuk menguji kinerja model setelah tahap pelatihan selesai. Pada penelitian ini digunakan metode 5-*fold cross-validation*, di mana dataset dibagi menjadi 5 subset dengan ukuran yang seimbang. Pada setiap iterasi, satu subset digunakan sebagai data uji dan empat subset lainnya sebagai data latih, kemudian hasil evaluasi dirata-ratakan. Pemilihan nilai k = 5 didasarkan pada pertimbangan jumlah data yang tersedia (392 data) agar setiap subset memiliki cukup sampel untuk proses pelatihan dan pengujian, sehingga

hasil evaluasi tetap stabil dan representatif tanpa meningkatkan risiko *overfitting* maupun *underfitting*[17].

2.4. Penerapan *Naive Bayes*

Naive Bayes adalah teknik dalam *machine learning* yang memanfaatkan perhitungan probabilitas dengan mengacu pada pendekatan *bayes*. Dalam algoritma *naive bayes*, teorema *bayes* diterakan dengan menggabungkan prior *probability* dan *conditional probability* [18].

2.5. Penerapan *Naive Bayes* Dengan *RFE*

Penerapan *Naïve Bayes* dengan metode *Recursive Elimination* merupakan teknik untuk seleksi fitur yang bekerja secara *recursive* yang berfungsi untuk menghapus atau menghilangkan fitur fitur yang di anggap kurang penting. Dengan demikian, model dapat berfungsi dengan lebih efisien dan akurat, Sehingga mampu meningkatkan tingkat akurasi secara keseluruhan.

2.6. Evaluasi Model

Tahap evaluasi adalah langkah penting setelah pembentukan model. Pada tahapan ini, performa model akan diukur untuk mengevaluasi akurasi. Dataset yang digunakan memiliki 12 atribut utama dan 1 atribut kelas. Transformasi dilakukan pada atribut Jenis Kelamin dan *Disease* sebagai label kelas. Pengujian dilakukan dengan teknik *Confusion Matriks. Confusion Matriks* adalah metode yang digunakan untuk mengukur akurasi dalam data mining. Metode ini melibatkan dua kelas, yaitu kelas positif dan kelas negatif. *Confusion Matriks* terdiri dari empat komponen, yaitu *true positif (TP), false positif FP), true negatif (TN),* dan *false negatif* (FN) [19]. Nilai akurasi bisa diperoleh dengan menggunakan persamaan sebagai berikut [20].

$$ACCURACY = \frac{TP + TN}{TP + TN + FP + FN} \; X\,100\% \quad (1)$$

Keterangan :

True Positive (TP) : Jumlah titik data dengan label 'ringan' yang berhasil didefinisikan bener.
True Negative (TN) : Jumlah titik data dengan label 'berat' yang berhasil didefinisikan salah.
False Positive (FP) : Jumblah titik data yang sebenarnya salah tetapi di beri label benar.
False Negative (FN) : Jumblah titik data yang sebenarnya benar tetapi di beri label salah.

3. HASIL DAN PEMBAHASAN

3.1. Hasil Penelitian

Tabel 1. Dataset Awal

No	Atribut	Tipe Data	Keterangan
1	Tanggal	*Date*	Tanggal kunjungan pasien
2	Klinik	*String*	Jenis klinik tempat pasien berkunjung
3	No.mr	*string*	Nomer rekam medis pasien
4	Tanggal lahir	*Date*	Tanggal lahir pasien
5	Jenis Kelamin	*String*	Jenis kelamin pasien
6	Umur	*Numeric (Int)*	Umur berdasarkan tanggal pemeriksaan
7	Status perkawinan	*string*	Status pernikahan pasien
8	Pendidikan	*String*	Tingkat pendidikan terakhir pasien
9	Pekerjaan	*String*	Pekerjaan pasien
10	Status	*String*	Status asuransi pasien
11	Sistolik	*Numeric (Int)*	Tekanan darah saat jantung memompa darah mencapai sekitar 120/80 mmHg

12	Diastolik	*Numeric (Int)*	Tekanan darah saat jantung beristirahat berkisar antara 90/60 mmHg
13	Nadi	*Numeric (int)*	Denyut nadi normal berkisar 60-100 denyut per menit (bpm)
14	RR	*Numeric* (int)	Laju pernafasan pasien
15	Temp	*Numeric (Float)*	Suhu tubuh pasien
16	TB	*Numeric (float)*	Tinggi badan pasien
17	BB	*Numeric (Float)*	Berat badan pasien
18	LK	*Numerik (int)*	Lingkar kaki
19	LP	*Numeric (int)*	Lingkar perut pasien
20	LL	*Numeric(int)*	Lingkar lengan pasien
21	Disease	*String (Class)*	target class (Berat dan Ringan)
22	Perawat	*String*	Nama perawat yang menangani pasien
23	Dokter	*String*	Nama dokter yang menangani pasien

Berdasarkan hasil penelitian yang telah dilakukan, penelitian ini bertujuan untuk mengevaluasi kinerja algoritma *Naive Bayes* yang dikombinasikan dengan metode *Adaptive Synthetic Sampling (ADASYN) dan Recursive Feature Elimination (RFE)* dalam proses klasifikasi penyakit dermatitis di UPT Puskesmas Bontang Barat. Dataset yang digunakan terdiri atas 23 kolom, dengan 22 atribut sebagai fitur dan 1 atribut kelas sebagai label *(Disease)*.

3.1.1. Data *Selection*

Tabel 2. Data *Selection*

Atribut	Data 1	Data 2	Data 3	*Other* data
Jenis Kelamin	Laki-laki	Laki-laki	Perempuan	…
Sistolik	90	120	120	…
Diastolik	60	80	80	…
Umur	3	68	47	…
Pendidikan	Tidak/belum sekolah	SD	SMA	…
Nadi	100	80	80	…
RR	26	18	18	…
Temp	36.5	36.0	36.0	…
BB	11.0	56.0	61.0	…
TB	100.0	155.0	148.0	…
LP	—	—	—	…
Disease	Ringan	Berat	Berat	…

Tabel 2 menampilkan atribut-atribut yang telah diseleksi secara manual, sehingga tersisa 12 atribut utama dan 1 atribut sebagai kelas. Seleksi manual dilakukan untuk menghapus atribut yang bersifat administratif atau tidak relevan dengan konteks medis, serta atribut dengan kualitas data rendah. Langkah ini bertujuan untuk memastikan data yang digunakan dalam proses klasifikasi lebih bersih dan fokus pada variabel yang berkaitan dengan kondisi pasien. Setelah proses ini, metode *Recursive Feature Elimination (RFE)* tetap diterapkan untuk memilih fitur yang paling optimal secara kuantitatif dalam meningkatkan kinerja model.

p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

3.1.2. Data *Cleaning*

Tabel 3. Data *Cleaning*

Atribut	Data 1	Data 2	Data 3	*Other* data
Jenis Kelamin	Laki-laki	Laki-laki	Perempuan	…
Sistolik	90	120	120	…
Diastolik	60	80	80	…
Umur	3	68	47	…
Nadi	100	80	80	…
RR	26	18	18	…
Temp	36.5	36.0	36.0	…
BB	11.0	56.0	61.0	…
TB	100.0	155.0	148.0	…
Disease	Ringan	Berat	Berat	…

Pada Tabel 3 merupakan tampilan data yang sudah dibersihkan dengan cara penghapusan atribut yang memiliki banyak nilai kosong dan atribut yang tidak digunakan pada saat pengujian nantinya. Atribut yang dihapus yaitu pendidikan dan LP.

3.1.3. Data *Transformation*

Tabel 4. Data *Transformation*

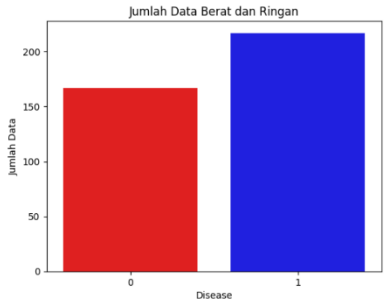
Atribut	Data 1	Data 2	Data 3	*Other* data
Jenis Kelamin	0	0	1	…
Sistolik	90	120	120	…
Diastolik	60	80	80	…
Umur	3	68	47	…
Nadi	100	80	80	…
RR	26	18	18	…
Temp	36.5	36.0	36.0	…
BB	11.0	56.0	61.0	…
TB	100.0	155.0	148.0	…
Disease	0	1	1	…

Pada Tabel 4 semua atribut yang ada pada dataset sudah dirubah menjadi tipe data *Integer*, data yang di transformasi berupa Jns.Kelamin (laki laki dan perempuan menjadi 0 dan 1) dan juga atribut target yaitu *Disease* (ringan dan berat menjadi 0 dan 1).

3.1.4. Data *Balancing*

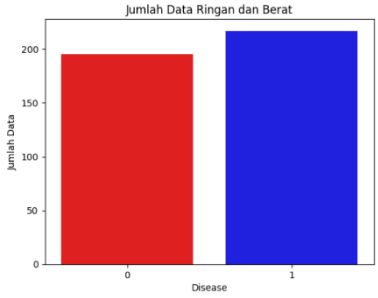
Proses penyeimbangan data dilakukan dengan menggunakan teknik penyeimbang data yaitu *ADASYN*. Yang bertujuan untuk menyeimbangkan jumblah sampel antara kelas miniritas dan juga kelas mayoritas dalam dataset penyakit dermatitis.

Vol. 10, No. 2, Oktober 2025, Page. 424-436
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI



Gambar 2. Jumlah data per kelas sebelum implementasi *ADASYN*

Pada Gambar 2 ada perbedaan jumblah kelas yang mana dari kategori kelas Ringan (0) berjumblah 167 data sedangkan kategori kelas Berat (1) berjumblah 217 data.



Gambar 3. Jumlah data per kelas setelah implementasi *ADASYN*

Pada Gambar 3 terlihat bahwa distribusi kelas antara kategori Ringan dan Berat menjadi lebih seimbang setelah diterapkannya metode *ADASYN*, dengan jumlah data Ringan sebanyak 195 dan kategori Berat sebanyak 217. Meskipun selisih jumlah antar kelas tidak terlalu besar, *ADASYN* tetap digunakan untuk mengurangi bias terhadap kelas mayoritas dan meningkatkan kemampuan model dalam mengenali kelas minoritas. *ADASYN* tidak mensintesis data secara merata, melainkan secara adaptif berdasarkan tingkat kesulitan sampel. Oleh karena itu, hasil akhirnya mendekati seimbang, tetapi tidak identik secara jumlah.

3.1.5. Permodelan *Naive Bayes*

Tabel 5. Hasil Pengujian Model *Naïve Bayes*

Fold	Akurasi	Precision	Recall	F1 - Score
1	63.64%	68.16%	63.64%	62.94%
2	57.14%	61.17%	57.14%	56.15%
3	61.04%	66.65%	61.04%	60.20%
4	61.04%	67.88%	61.04%	59.84%
5	56.58%	60.10%	56.58%	56.01%
Rata Rata	59.89%	64.79%	59.89%	59.03%

Berdasarkan hasil evaluasi model menggunakan *5-fold cross-validation*, diperoleh akurasi rata-rata sebesar 59,89% yang dihitung dari keseluruhan fold, model *Naive Bayes* menunjukkan kinerja yang cukup stabil dalam proses klasifikasi pada dataset yang digunakan. Akurasi yang didapatkan ini masih bisa ditingkatkan lagi dengan menggunakan beberapa penerapan yang akan dilakukan dalam pengujian berikutnya. Dan akan dilakukan juga proses perhitungan manual dengan cara menggunakan *Confusion Matriks* yang berfungsi untuk memastikan ketepatan dari hasil akurasi.

Tabel 6. *Confusion Matrix Naive Bayes*

	Predicted Positive (0)	*Predicted Negativ (1)*
Actual Positive (0)	133	34

Actual Negative (1)	120	97

Berdasarkan Tabel 6, model berhasil mengklasifikasikan 133 data positif (kelas 0) dengan benar sebagai positif *(True Positive)* dan 97 data negatif (kelas 1) dengan benar sebagai negatif *(True Negative)*. Namun, model juga menghasilkan 34 kesalahan klasifikasi terhadap data positif yang diprediksi sebagai negatif *(False Negative)*, serta 120 kesalahan klasifikasi terhadap data negatif yang diprediksi sebagai positif *(False Positive)*. Tingginya nilai False Positive menunjukkan bahwa model masih mengalami kesulitan dalam mengenali data dari kelas negatif. Hal ini dapat disebabkan oleh ketidakseimbangan data atau kurang representatifnya fitur untuk kelas negatif.

3.1.6. Permodelan *Naive Bayes* Dengan *ADASYN*

Tabel 7. Hasil Pengujian *Naïve Bayes* dengan *ADASYN*

Fold	Akurasi	Precision	Recall	F1 - Score
1	63.86%	63.83%	63.86%	63.50%
2	67.47%	70.20%	67.47%	65.57%
3	64.63%	64.77%	63.64%	64.22%
4	65.85%	65.95%	65.85%	65.54%
5	70.73%	71.39%	70.73%	70.68%
Rata rata	66.50%	67.22%	66.50%	65.90%

Berdasarkan hasil evaluasi didapatkan performa akurasi yang cukup baik, dengan akurasi rata rata mencapai 66,50%. Hasil ini mengindikasikan bahwa performa dari model algoritma *Naive Bayes* dengan *ADASYN* dalam melakukan proses klasifikasi dengan akurasi yang cukup baik.

Tabel 8. *Confusion Matrix Naive Bayes* dengan *ADASYN*

	Predicted Positive (0)	*Predicted Negativ (1)*
Actual Positive (0)	111	84
Actual Negative (1)	54	163

Tabel 8 menunjukkan *confusion matrix* dari hasil evaluasi model *Naive Bayes* setelah penerapan teknik ADASYN. Berdasarkan tabel tersebut, model berhasil mengklasifikasikan 163 data dari kelas negatif (kelas 1) secara benar sebagai negatif *(True Negative)* dan 111 data dari kelas positif (kelas 0) secara benar sebagai positif *(True Positive)*. Jumlah prediksi benar yang signifikan, khususnya pada kelas negatif, mencerminkan bahwa penerapan *ADASYN* memberikan kontribusi dalam meningkatkan sensitivitas model terhadap data minoritas.

3.1.7. Permodelan *Naive Bayes* Dengan *RFE*

Proses dilakukan dengan mengeliminasi fitur satu per satu berdasarkan kontribusinya terhadap kinerja model secara menyeluruh

Tabel 9. Hasil rangking dari *RFE*

Atribut	Ranking	Mutual Information
Sistolik	1	0.091784
Diastolik	1	0.076598
Umur	1	0.090746
BB	1	0.120435
TB	1	0.094878
Nadi	2	0.000000

Vol. 10, No. 2, Oktober 2025, Page. 424-436
https://journal.uin-alauddin.ac.id/index.php/instek
p-ISSN: 2541-1179
e-ISSN: 2581-1711
JURNAL INSTEK
INFORMATIKA SAINS DAN TEKNOLOGI

RR	3	0.012423
Jns.kelamin	4	0.000000
Temp	5	0.023895

Berdasarkan hasil seleksi fitur menggunakan metode *RFE*, lima fitur yang dipertahankan dengan peringkat 1 adalah Sistolik, Diastolik, Umur, Berat Badan (BB), dan Tinggi Badan (TB). Fitur-fitur ini dipilih karena menunjukkan kontribusi terbesar terhadap kinerja model secara keseluruhan selama proses eliminasi. Sementara itu, fitur lain seperti Nadi, Respiratory Rate (RR), Jenis Kelamin, dan Suhu Tubuh memperoleh peringkat lebih tinggi (≥2) dan dieliminasi karena kontribusinya dinilai lebih rendah. Perlu dicatat bahwa nilai *Mutual Information* yang ditampilkan hanya berfungsi sebagai informasi tambahan dan tidak digunakan sebagai dasar peringkat oleh *RFE*, karena *RFE* melakukan pemilihan fitur berdasarkan dampaknya terhadap akurasi model melalui evaluasi iteratif menggunakan estimator yang mendasarinya.

Tabel 10. Hasil Pengujian *Naïve Bayes* dengan *RFE*

Fold	Akurasi	Precision	Recall	F1 - Score
1	89.61%	89.66%	89.61%	89.57%
2	90.91%	91.08%	90.91%	90.85%
3	93.51%	93.51%	93.51%	93.49%
4	92.21%	92.38%	92.21%	93.23%
5	88.16%	90.70%	88.16%	88.18%
Rata rata	90.88%	91.47%	90.88%	90.86%

Pada Tabel 10 model dengan penambahan metode seleksi fitur menggunakan *Recursive Feature Elimination* memperoleh hasil akurasi yang sangat baik dalam melakukan proses klasifikasi pada dataset penyakit dermatitis. Model algortima *Naive Bayes* mendapatkan akurasi dengan nilai rata rata 90,88%. Penggunaan *RFE* dalam proses klasifikasi dapat membantu model algoritma *Naive Bayes* untuk memilih fitur fitur yang mempunyai peran penting dalam proses klasifikasi pada dataset penyakit dermatitis yang dalam prosesnya dapat meningkatkan tingkat akurasi kinerjanya.

Tabel 11. *Confusion Matrix Naive Bayes dengan RFE*

	Predicted Positive (0)	*Predicted Negativ (1)*
Actual Positive (0)	152	15
Actual Negative (1)	20	192

Tabel 11 menunjukkan *confusion matrix* dari hasil evaluasi model *Naive Bayes* yang telah dioptimalkan menggunakan metode seleksi fitur *RFE*. Berdasarkan hasil tersebut, model berhasil mengklasifikasikan 152 data dari kelas positif (kelas 0) secara benar sebagai positif *(True Positive)* dan 192 data dari kelas negatif (kelas 1) secara benar sebagai negatif *(True Negative)*. Kesalahan klasifikasi yang terjadi juga tergolong kecil, yaitu hanya 15 data positif yang diprediksi sebagai negatif *(False Negative)* dan 20 data negatif yang diprediksi sebagai positif *(False Positive)*. Hasil ini menunjukkan bahwa penerapan metode RFE mampu meningkatkan performa model secara signifikan, dengan akurasi yang tinggi dan tingkat kesalahan yang sangat rendah.

3.1.8. Perbandingan Hasil

Tabel 12. Perbandingan Hasil Akurasi Pengujian

Fold	*Naïve Bayes*	*Naïve Bayes + ADASYN*	*Naïve Bayes + RFE + ADASYN*	Kenaikan *NB* ke *ADASYN*	Kenaikan *ADASYN* ke *RFE*
1	63.64%	63.86%	89.61%	+0.22%	+25.75%
2	57.14%	67.47%	90.91%	+10.33%	+23.44%
3	61.04%	64.63%	93.51%	+3.59%	+28.88%
4	61.04%	65.85%	92.21%	+4.81%	+26.36%
5	56.58%	70.73%	88.16%	+14.15%	+17.43%

Berdasarkan Tabel 12, model *Naive Bayes* dengan teknik oversampling *ADASYN* dan seleksi fitur *RFE* menunjukkan peningkatan akurasi yang konsisten dibandingkan tanpa keduanya. Dari 5-*Fold Cross Validation*, seluruh *fold* mengalami kenaikan akurasi. Akurasi awal sebesar 59,89% meningkat menjadi 66,50% setelah *ADASYN*, dan mencapai 90,88% setelah penerapan *RFE*. Peningkatan bertahap masing-masing sebesar 6,61% dan 24,38%. Hal ini membuktikan bahwa kombinasi *ADASYN* dan *RFE* secara signifikan meningkatkan kinerja model dalam mengklasifikasikan data secara lebih akurat dan efisien.

3.2. Diskusi

Secara medis, fitur tekanan darah, umur, dan BMI (berat badan dan tinggi badan) berkaitan dengan kondisi fisiologis yang dapat memengaruhi tingkat keparahan dermatitis, seperti stres, metabolisme, dan peradangan kulit. Dari sisi metodologis, peningkatan akurasi yang signifikan setelah *RFE* menunjukkan bahwa penghapusan fitur tidak relevan memiliki pengaruh lebih besar dibandingkan hanya menyeimbangkan data dengan *ADASYN*. Hal ini menegaskan pentingnya kualitas fitur dalam meningkatkan performa model klasifikasi.

3.3. Pembahasan

Penelitian ini menggunakan data penyakit dermatitis dari UPT Puskesmas Bontang Barat tahun 2024 dengan tahapan *preprocessing* seperti seleksi, pembersihan, transformasi, dan penyeimbangan data menggunakan *ADASYN* karena ketidakseimbangan kelas. Data kemudian dibagi menggunakan metode 10-*Fold Cross Validation* untuk evaluasi yang merata. Hasil penelitian menunjukkan bahwa lima fitur terbaik (Sistolik, Diastolik, Umur, BB, dan TB) berkontribusi signifikan terhadap performa model *Naive Bayes*. Akurasi awal sebesar 59,89% meningkat menjadi 66,50% setelah *ADASYN* dan mencapai 90,88% setelah ditambah seleksi fitur *RFE*. Hal ini membuktikan bahwa kombinasi *ADASYN* dan *RFE* secara efektif meningkatkan akurasi model. Hasil ini sejalan dengan penelitian Pratama et al. (2022) yang menunjukkan bahwa kombinasi *RFE* dan *ADASYN* mampu meningkatkan akurasi dari 55,5% menjadi 85,2%. Hal ini membuktikan bahwa kombinasi kedua teknik tersebut tidak hanya menyederhanakan fitur, tetapi juga mengatasi ketidakseimbangan data, sehingga meningkatkan performa klasifikasi secara signifikan.

4. KESIMPULAN

Berdasarkan hasil penelitian, diperoleh kesimpulan bahwa metode seleksi *fitur Recursive Feature Elimination (RFE)* berhasil mengidentifikasi lima fitur penting dalam klasifikasi penyakit dermatitis, yaitu Sistolik, Diastolik, Umur, Berat Badan, dan Tinggi Badan. Selain itu, penerapan teknik *oversampling ADASYN* yang dikombinasikan dengan RFE mampu meningkatkan akurasi model Naive Bayes secara signifikan, dari 59,89% menjadi 90,88%. Hasil ini menunjukkan bahwa seleksi fitur dan penyeimbangan data sangat berperan dalam meningkatkan performa model klasifikasi.

DAFTAR PUSTAKA

[1] E. Lisma, A. Arbi, and V. N. Arifin, "Faktor-Faktor Yang Berhubungan Dengan Upaya Pencegahan Dermatitis Kontak," *Jambura Heal. Sport J.*, vol. 6, no. 2, pp. 176–184, 2024, doi: 10.37311/jhsj.v6i2.26823.

[2] Kemenkes RI, "Profil Kesehatan Indonesia. Kementerian Kesehatan Republik Indonesia, Jakarta.," 2022.

[3] K. P. Pohan and C. Chairunisah, "Sistem Pakar Mendiagnosa Penyakit Kulit Pada Manusia Menggunakan Metode Naive Bayes Classifier Berbasis Web," *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 23, no. 1, p. 204, 2024, doi: 10.53513/jis.v23i1.9521.

[4] Putri Armilia Prayesy, "Studi Perbandingan Metode Support Vector Machine , Random Forest , Dan Convolutional Neural Network Untuk Klasifikasi A Comparative Study Of Support Vector Machine , Random Forest , And Convolutional Neural Network Methods For Skin Disease Dataset Classif," *JKBTI (Jurnal kecerdasan Buatan Teknol. Informasi)*, vol. 4, no. 1, pp. 70–76, 2025, doi: https://doi.org/10.69916/jkbti.v4i1.214.

[5] R. Maulana, R. Narasati, R. Herdiana, R. Hamonangan, and S. Anwar, "Komparasi Algoritma Decision Tree Dan Naive Bayes Dalam Klasifikasi Penyakit Diabetes," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 7, no. 6, pp. 3865–3870, 2024, doi: 10.36040/jati.v7i6.8265.

[6] M. Rizal, M. Z. Syahaf, S. R. Priyambodo, and Y. Rhamdani, "Optimasi Algoritma Naïve Bayes Menggunakan Forward Selection Untuk Klasifikasi Penyakit Ginjal Kronis," *Naratif J. Nas. Riset, Apl. dan Tek. Inform.*, vol. 5, no. 1, pp. 71–80, 2023, doi: 10.53580/naratif.v5i1.200.

[7] Y. B. Widodo, S. A. Anggraeini, and T. Sutabri, "Perancangan Sistem Pakar Diagnosis Penyakit Diabetes Berbasis Web Menggunakan Algoritma Naive Bayes," *J. Teknol. Inform. dan Komput.*, vol. 7, no. 1, pp. 112–123, 2021, doi: 10.37012/jtik.v7i1.507.

[8] G. Dwilestari and T. A. Afifah, "Perbandingan Kinerja Algoritma Naive Bayes Dan Decision Tree Dalam Klasifikasi Kanker Paru-Paru," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 9, no. 1, pp. 801–807, 2025, doi: https://ejournal.itn.ac.id/index.php/jati/article/view/12463.

[9] Sutarman, R. Siringoringo, D. Arisandi, E. Kurniawan, and E. B. Nababan, "Model Klasifikasi Dengan Logistic Regression Dan Recursive Feature Elimination Pada Data Tidak Seimbang," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 735–742, 2024, doi: 10.25126/jtiik.1148198.

[10] C. Kuzudisli, B. Bakir-Gungor, B. Qaqish, and M. Yousef, "RCE-IFE: Recursive Cluster Elimination with Intra-cluster Feature Elimination," *PeerJ Comput. Sci.*, pp. 2–27, 2024, doi: https://doi.org/10.1101/2024.02.28.580487.

[11] A. sami Jaddoa, S. J. Saba, and E. A.Abd Al-Kareem, "Liver Disease Prediction Model Based on Oversampling Dataset with RFE Feature Selection using ANN and AdaBoost algorithms," *Buana Inf. Technol. Comput. Sci. (BIT CS)*, vol. 4, no. 2, pp. 85–93, 2023, doi: 10.36805/bit-cs.v4i2.5565.

[12] M. E. Özateş, A. Yaman, F. Salami, S. Campos, S. I. Wolf, and U. Schneider, "Identification and interpretation of gait analysis features and foot conditions by explainable AI," *Sci. Rep.*, vol. 14, no. 1, pp. 1–13, 2024, doi: 10.1038/s41598-024-56656-4.

[13] C. Kaope and Y. Pristyanto, "The Effect of Class Imbalance Handling on Datasets Toward Classification Algorithm Performance," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 22, no. 2, pp. 227–238, 2023, doi: 10.30812/matrik.v22i2.2515.

[14] M. Tiara *et al.*, "Pemanfaatan Algoritma Adasyn Dan Support Vector Machine Dalam Meningkatkan Akurasi Prediksi Kanker Paru-Paru," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 5, pp. 8773–8778, 2024, doi: https://doi.org/10.36040/jati.v8i5.10752.

[15] 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.

[16] R. Syaputra, T. A. Y. Siswa, and W. J. Pranoto, "Model Optimasi SVM Dengan PSO-GA dan SMOTE Dalam Menangani High Dimensional dan Imbalance Data Banjir," *Teknika*, vol. 13, no. 2, pp. 273–282, 2024, doi: 10.34148/teknika.v13i2.876.

[17] T. Ridwansyah, "Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier," *KLIK Kaji. Ilm. Inform. dan Komput.*, vol. 2, no. 5, pp. 178–185, 2022, doi: 10.30865/klik.v2i5.362.

[18] W. Yulita, "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," *J. Data Min. dan Sist. Inf.*, vol. 2, no. 2, p. 1, 2021, doi: 10.33365/jdmsi.v2i2.1344.

[19] N. S. Fauziah and R. D. Dana, "Implementasi Algoritma Naive bayes dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.

[20] B. P. Pratiwi, A. S. Handayani, and S. Sarjana, "Pengukuran Kinerja Sistem Kualitas Udara Dengan Teknologi Wsn Menggunakan Confusion Matrix," *J. Inform. Upgris*, vol. 6, no. 2, pp. 66–75, 2021, doi: 10.26877/jiu.v6i2.6552.