

PENERAPAN ALGORITMA *DECISION TREE*, *RECURSIVE ELIMINATION* DAN ADASYN PADA DATA KLASIFIKASI PENYAKIT DERMATITIS

Landika Bramasta Erlian Susilo, Taghfirul Azhima Yoga Siswa, Naufal Azmi Verdikha

Teknik Informatika, Universitas Muhammadiyah Kalimantan Timur

Jl. Ir. H. Juanda No.15, Sidodadi, Kec. Samarinda Ulu, Kota Samarinda, Kalimantan Timur 75124

tay758@umkt.ac.id

ABSTRAK

Dermatitis merupakan penyakit kulit yang cukup umum di Indonesia, berdasarkan Riset Kesehatan Dasar (Riskesdas) pada tahun 2018 kasus dermatitis di Indonesia mencapai angka 6,8%. Kemudian di kota Kendari menurut Badan Pusat Statistik (BPS) pada tahun 2021 terdapat 2600 kasus dermatitis menempati peringkat ke-4 dari 10 penyakit terbanyak. Penelitian ini bertujuan untuk mengklasifikasikan penyakit dermatitis menggunakan algoritma *Decision Tree* dengan menerapkan seleksi fitur *Recursive Feature Elimination (RFE)* dan *Adaptive Synthetic Sampling (ADASYN)* untuk mengatasi data yang tidak seimbang. Data yang digunakan mencakup 392 data kasus penyakit dermatitis yang diperoleh dari UPT Puskesmas Bontang Barat periode 2024. Penelitian ini melakukan *preprocessing* data dilanjutkan dengan permodelan *Decision Tree*, seleksi fitur *RFE*, dan *ADASYN* dengan membagi data *training* dan *testing* menggunakan *10 fold cross validation*. Evaluasi model dalam penelitian ini menggunakan *confusion matrix* untuk mencari nilai akurasi. Hasil penelitian ini menunjukkan bahwa seleksi fitur *RFE* pada data dermatitis menghasilkan beberapa fitur yang relevan seperti Umur, BB, RR, Nadi, dan TB. Hasil evaluasi menunjukkan bahwa *Decision Tree* menghasilkan akurasi 60.19%, *Decision Tree* dengan *ADASYN* menghasilkan akurasi 66.26%, dan *Decision Tree* dengan seleksi fitur *RFE* menghasilkan akurasi 92.44% terjadi peningkatan sebesar 26.2%.

Kata Kunci: Dermatitis, *Decision Tree*, *RFE*, *ADASYN*, Klasifikasi

1. PENDAHULUAN

Kulit berperan penting sebagai pelindung tubuh terhadap paparan debu, zat kimia, sinar matahari, serta mikroorganisme seperti virus dan bakteri. Salah satu gangguan yang sering menyerang kulit adalah dermatitis, yaitu peradangan yang ditandai dengan gejala gatal, kemerahan, dan pembengkakan. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) tahun 2018, prevalensi dermatitis di Indonesia mencapai 6,8% (Riskesdas, 2018). Di Kota Kendari, pada tahun 2021 tercatat 2.600 kasus dermatitis yang menjadikannya penyakit keempat terbanyak [1].

Dalam rangka meningkatkan efektivitas penanganan kasus dermatitis, pendekatan berbasis teknologi seperti *machine learning* menjadi solusi yang menjanjikan. *Machine learning* merupakan cabang dari kecerdasan buatan yang mampu menyajikan analisis prediktif untuk mendukung proses pengambilan keputusan [2]. Perkembangan teknologi ini memungkinkan pemanfaatannya dalam klasifikasi penyakit, termasuk dermatitis [3].

2. TINJAUAN PUSTAKA

Sejumlah studi sebelumnya telah menerapkan algoritma *machine learning* untuk klasifikasi penyakit. Penelitian oleh [4] menunjukkan bahwa algoritma *Naive Bayes* menghasilkan akurasi sebesar 80%, sedangkan *K-Nearest Neighbor (KNN)* mencapai 84% pada klasifikasi diabetes mellitus. Sementara itu, [5] melaporkan akurasi sebesar 70% untuk *KNN*, 79% untuk *Naive Bayes*, dan 73% untuk *Logistic Regression* pada klasifikasi penyakit diabetes. Penelitian lain oleh [6] menggunakan algoritma

Decision Tree untuk klasifikasi penyakit jantung dan berhasil mencapai akurasi sebesar 87,50%.

Algoritma *Decision Tree* kerap dipilih karena strukturnya yang terorganisir dan kemudahan dalam pengolahan data [7]. Selain itu, pendekatan ini memungkinkan interpretasi hierarkis terhadap data medis, yang berguna dalam pengambilan keputusan klinis [8]. Keunggulan lainnya adalah kemudahan dalam memahami hasil klasifikasi oleh tenaga kesehatan [9]. Dalam studi oleh [10], algoritma *Decision Tree* menghasilkan akurasi sebesar 95,45%, dibandingkan *KNN* (90,91%) dan *Naive Bayes* (86,36%).

Namun, tantangan seperti *overfitting* dapat terjadi ketika model terlalu kompleks dan hanya cocok dengan data pelatihan [11]. Untuk mengatasi hal ini, diperlukan teknik seleksi fitur agar hanya fitur yang relevan yang digunakan dalam pelatihan model. Salah satu metode yang efektif adalah *Recursive Feature Elimination (RFE)*, yang secara bertahap menghapus fitur dengan kontribusi paling rendah terhadap proses klasifikasi [12]. Teknik ini umum digunakan dalam skenario klasifikasi dan prediksi berbasis *machine learning* [13].

RFE terbukti dapat menyaring fitur yang paling berpengaruh dan meningkatkan akurasi model [14]. [15] menunjukkan bahwa penerapan *RFE* pada algoritma *Decision Tree* mampu meningkatkan akurasi di atas 77%. Studi [16] pada prediksi kanker payudara menggunakan algoritma *SVM* dan seleksi fitur *RFE* bahkan mencatatkan akurasi hingga 98,25%.

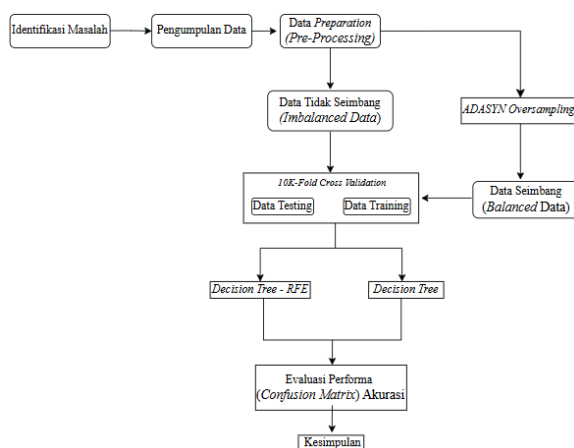
Selain seleksi fitur, tantangan lain dalam klasifikasi adalah ketidakseimbangan data. Dalam

studi ini, jumlah data kelas mayoritas (217 data) lebih banyak dibanding kelas minoritas (167 data), yang dapat memengaruhi performa model dalam mengidentifikasi kelas minoritas secara akurat [17]. Untuk mengatasi masalah ini, digunakan metode *oversampling* ADASYN yang menghasilkan data sintesis untuk kelas minoritas. Penelitian oleh [18] membuktikan bahwa ADASYN mampu meningkatkan performa klasifikasi.

[19] Membandingkan metode *oversampling*, dan hasilnya menunjukkan bahwa ADASYN memberikan akurasi lebih baik dibanding metode lainnya, mencapai 87,3%. Penelitian oleh [20] mendukung temuan ini dengan menunjukkan peningkatan akurasi dari 59% menjadi 99% setelah penerapan ADASYN.

Melalui penggabungan algoritma *Decision Tree*, seleksi fitur *RFE*, dan metode penyeimbangan ADASYN, penelitian ini bertujuan untuk menghasilkan model klasifikasi yang optimal dalam menganalisis penyakit dermatitis. Pendekatan ini diharapkan dapat membantu meningkatkan akurasi prediksi serta menjadi kontribusi bagi peningkatan kualitas layanan kesehatan, khususnya dalam penanganan dermatitis.

3. METODE PENELITIAN



Gambar 1. Alur penelitian

3.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data bencana penyakit dermatitis yang diperoleh dari UPT Puskesmas Bontang Barat pada tahun 2024. Adapun hasil dari pengumpulan data didapatkan 12 fitur untuk mengklasifikasikan data penyakit dermatitis. Terdapat 12 atribut yaitu Jenis Kelamin, Umur, Pendidikan, Sistolik, Diastolik, Nadi, RR, Temp, BB, TB, LP, dan Disease. Pada setiap atribut memiliki tipe data *string* atau *numeric* untuk menggambarkan kondisi fisik pasien, kemudian pada kolom Disease adalah target *class* dengan kategori berat dan ringan.

3.2. Data Pre-Processing

Diperlukan data lebih lanjut sebelum dimasukkan ke dalam proses pemodelan, proses akan dilakukan pada 4 langkah, yaitu data *selection*, data *cleaning*, data *transformation*, dan data *balancing*. Data ini diperoleh dari UPT Puskesmas Bontang Barat.

a. Data Selection

Pada tahap ini akan dilakukan pemilihan fitur yang paling relevan untuk klasifikasi penyakit dermatitis. Pada fitur yang tidak memberikan kontribusi signifikan pada proses klasifikasi akan di hapus.

b. Data Cleaning

Data *cleaning* bertujuan untuk menghilangkan nilai kosong, proses data *cleaning* dengan menghapus atau mengoreksi data yang salah atau tidak lengkap. Proses ini sangat penting karena memastikan keakuratan pada saat analisis dan prediksi.

c. Data Transformation

Data *transformation* adalah proses mengubah data ke dalam format atau skala yang sesuai untuk analisis dengan mengubah nilai atribut yang bersifat kategorial menjadi numerik. Untuk tahap ini dilakukan terlebih dahulu *LabelEncoder* untuk mengubah tipe data karakter menjadi numerik [21]. Proses konversi ini menggunakan *library scikit-learn* dengan memanfaatkan fungsi *LabelEncoder*.

d. Data Balancing

Data *balancing* adalah teknik penyeimbangan data yang bertujuan untuk menyeimbangkan distribusi dataset, dan mencegah kesalahan pada algoritma klasifikasi yang disebabkan ketidakseimbangan antar kelas [22]. Metode yang diterapkan dalam menangani masalah ini adalah metode ADASYN dengan berfokus pada peningkatan jumlah sampel di kelas minoritas agar model lebih efektif dalam mengenali pola dengan akurasi yang lebih baik.

3.3. Pembagian Data

Pada proses ini dilakukan pembagian data dengan memisahkan dataset menjadi dua komponen, yaitu data latih (*training*) dan data uji (*test*) [23]. Data *training* berfungsi untuk melatih model dalam mempelajari pola dan hubungan antara fitur-fitur dalam data, lalu *testing* bertugas dalam menguji kinerja model setelah selesai proses *training*. Penelitian ini akan menerapkan teknik *K-Fold Cross-Validation* dengan K=10 dalam menilai performa model *machine learning* terutama dalam mengklasifikasi data penyakit dermatitis. Pengujian K-Fold didesain untuk mengevaluasi kinerja model pada setiap nilai K yang dipilih [24].

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 tabel 1 adalah skema proses 10-Fold Cross-Validation, dimana dataset dibagi menjadi 10 bagian yang sama besar dengan setiap baris dan kolom mewakili satu *fold* ($K=10$). Setiap *fold* akan digunakan secara bergantian pada pengujian data ditandai dengan warna (abu-abu), dan data *fold* lainnya akan digunakan sebagai data pelatihan, proses ini akan terus diulang sebanyak 10 kali. Kemudian hasil evaluasi dari semua *fold* dirata-rata untuk menghasilkan estimasi performa yang lebih akurat.

3.4. Penerapan Decision Tree

Decision Tree adalah sebuah algoritma berbentuk struktur pohon yang terdiri dari *node-node* yang mempresentasikan keputusan, dan memecah data menjadi lebih kecil sesuai atribut pada data [7]. Rumus *Decision Tree* adalah sebagai berikut:

$$Gain(S, A) = Entropy(S) \sum_{i=1}^n * Entropy(S_i) \quad (1)$$

Keterangan:

S = Himpunan

A = Atribut

n = Jumlah partisi atribut A

S_i = Jumlah kasus pada partisi ke-i

S = Jumlah kasus dalam S

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (2)$$

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi dari S_i terhadap S

Secara keseluruhan, *entropy* dan *gain* adalah dua konsep fundamental dalam pembentukan *Decision Tree*. *Entropy* membantu mengukur ketidakpastian dalam data, sementara *gain* digunakan untuk memilih atribut terbaik yang dapat meminimalkan ketidakpastian tersebut, sehingga menghasilkan model yang lebih efektif dalam klasifikasi.

3.5. Penerapan Decision Tree Dengan Fitur Recursive Elimination

Recursive Elimination merupakan metode yang secara rekursif menghilangkan fitur dan membangun model berdasarkan fitur yang tersisa. Pada metode ini akan diberi peringkat fitur berdasarkan kepentingan

dan menghilangkan fitur yang tidak penting hingga jumlah fitur yang diinginkan tercapai.

3.6. Evaluasi Model

Tahap ini merupakan langkah yang penting setelah pembentukan model karena performa model akan diukur untuk mengavaluasi akurasi. Pada pengujian ini dilakukan dengan teknik *confusion matrix*, *confusion matrix* yaitu tabel yang menunjukkan klasifikasi jumlah data uji yang benar dan salah [25].

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (3)$$

- *True Positive (TP)* : Jumlah titik data dengan label 'yes' yang diidentifikasi benar. *True*
- *Negative (TN)* : Jumlah titik data dengan label 'no' yang diidentifikasi salah. *False*
- *Positive (FP)* : Jumlah titik data yang sebenarnya salah namun diberi label benar
- *False Negative (FN)* : Jumlah titik data yang sebenarnya benar namun diberi label salah

4. HASIL DAN PEMBAHASAN

4.1. Hasil Penelitian

Pada bab ini akan ditampilkan berbagai hasil analisis yang telah dilakukan pengujian secara mendalam. Analisis ini akan mencakup evaluasi dan kinerja pada model pemilihan fitur serta hasil dari semua metode yang diterapkan pada penelitian ini. Hasil yang didapatkan akan di bandingkan dengan tujuan awal penelitian yaitu menilai sejauh mana model dan metode berhasil digunakan. Pada proses analisis data terdapat beberapa tahap utama, yaitu pengumpulan data, pemrosesan data, dan analisis data menggunakan algoritma *Decision Tree* dengan metode *Recursive Elimination* dan ADASYN. Data yang digunakan pada penelitian ini berasal dari dataset penyakit dermatitis di UPT Puskesmas Bontang Barat, terdapat 36 kolom, 35 atribut, dan 1 atribut *class* sebagai label (*Disease*).

4.2. Data Selection

Data selection adalah data yang berisi atribut yang telah diseleksi yaitu Jenis Kelamin, Sistolik, Diastolik, Nadi, RR, Temp, BB, TB, LP, dan Disease. 1 atribut *Disease* sebagai atribut kelas.

4.3. Data Cleaning

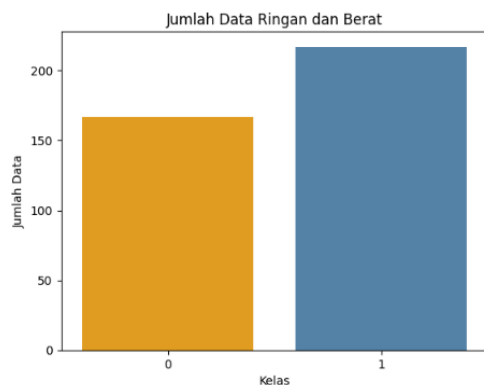
Dilakukan penghapusan kolom yang memuat nilai *N/A* (*no value is available*) data yang tidak memiliki nilai yaitu atribut LP dan menghapus atribut yang tidak berkontribusi signifikan dalam menentukan kelas serta tidak ada korelasinya dengan atribut lain yaitu atribut Pendidikan, karena dapat memengaruhi hasil akurasi. Proses ini dilakukan agar data bersih dan siap untuk analisis selanjutnya.

4.4. Data Transformation

Ada beberapa atribut yang ditransformasi yaitu jenis kelamin, dan disease. Pada proses ini akan dilakukan proses *encoding* untuk mengubah data *kategorial* menjadi *numeric* dan penggantian nilai pada kolom atribut label Disease dalam mengasumsikan kelas menjadi dua kategori. Dengan dilakukannya transformasi ini maka data akan lebih siap untuk digunakan dalam analisis serta pemodelan.

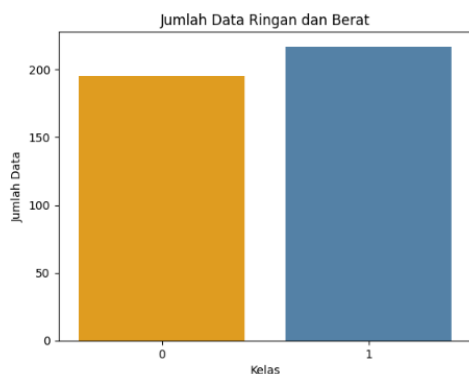
4.5. Data Balancing

Proses ini dilakukan untuk menyeimbangkan data dengan teknik ADASYN. Teknik ini berfungsi untuk menyeimbangkan jumlah sampel kelas minoritas dan kelas mayoritas pada *dataset*.



Gambar 2. Jumlah kelas sebelum *balancing*

Pada Gambar 2 terdapat perbedaan jumlah kelas dimana data kategori Ringan (0) berjumlah 167 data dan kategori Berat (1) berjumlah 217 data.



Gambar 3. Jumlah kelas sesudah *balancing*

Pada gambar 3 telah dilakukan *balancing* data sehingga kelas seimbang antara kelas mayoritas dan kelas minoritas setelah proses ADASYN jumlah kategori Ringan (0) berjumlah 195 data dan kategori Berat (1) berjumlah 217 data.

4.6. Permodelan Decision Tree

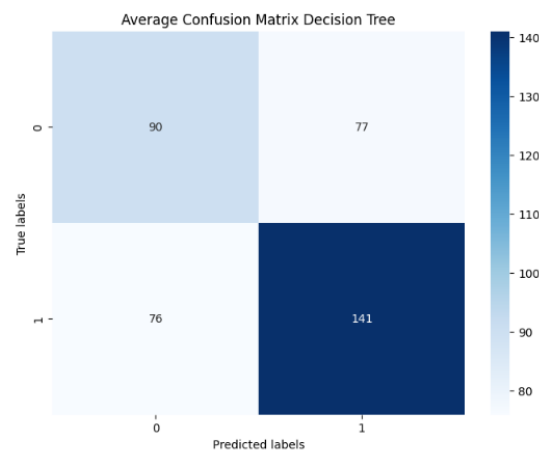
Pada tahap ini percobaan pertama pengujian dilakukan dengan menguji model menggunakan *Python*. Pada pengujian ini menerapkan algoritma *Decision Tree*. Pengujian ini bertujuan untuk menilai

kinerja dasar model *Decision Tree* dalam mengklasifikasi Hasil pengujian model *Decision Tree* pada setiap *fold* menunjukkan hasil kinerja yang cukup baik dalam memprediksi kelas pada dataset dermatitis. Setelah sepuluh *fold* dievaluasi, model ini menghasilkan akurasi yang cukup baik dengan nilai rata-rata sebesar 60.19%. Pada tingkat akurasi ini cukup baik sehingga menunjukkan kemampuan model dalam mengklasifikasi data secara akurat, namun belum maksimal.

Tabel 2. Hasil pengujian *decision tree*

Fold	Akurasi	Precision	Recall	F1 - Score
1	64.10%	63.76%	64.10%	63.76%
2	51.28%	50.95%	51.28%	51.09%
3	56.41%	57.42%	56.41%	56.58%
4	56.41%	56.72%	56.41%	56.53%
5	63.16%	64.57%	63.16%	63.16%
6	65.79%	65.60%	65.79%	65.27%
7	57.89%	57.89%	57.89%	57.89%
8	65.79%	65.39%	65.79%	64.37%
9	60.53%	60.21%	60.53%	60.33%
10	60.53%	61.77%	60.53%	60.77%
Average	60.19%	60.43%	60.19%	59.98%

Hasil evaluasi tabel 2 pada model *Decision Tree* menunjukkan hasil yang cukup baik dengan nilai akurasi rata-rata 60.19%. Ini menunjukkan bahwa model *Decision Tree* mampu mendapatkan tingkat akurasi yang cukup baik namun perlu ditingkatkan lagi. Dalam membandingkan hasil evaluasi dilakukan perhitungan manual terhadap nilai *Confusion Matrix* dalam memastikan keakuratan dari hasil evaluasi.



Gambar 4. *Confusion matrix decision tree*

$$\text{Accuracy} = \frac{90 + 141}{90 + 141 + 77 + 76} = 0.6015 = 60.15\%$$

Setelah dilakukan perhitungan manual pada tiap parameter maka didapatkan hasil akurasi yang sama saat perhitungan menggunakan *python*, dan perhitungan manual.

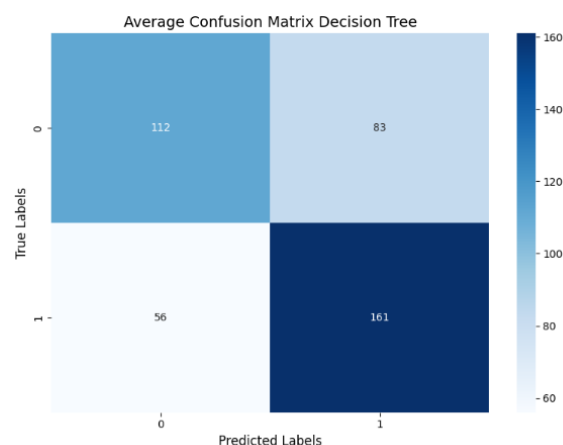
4.7. Permodelan Decision Tree dengan ADASYN

Pada tahap ini percobaan pertama pengujian dilakukan dengan menguji model menggunakan *Python*. Pada pengujian ini menerapkan algoritma *Decision Tree* dengan *ADASYN*. Pengujian ini bertujuan untuk menilai kinerja dasar model *Decision Tree* dalam mengklasifikasi data. Hasil pengujian model *Decision Tree* pada setiap *fold* menunjukkan hasil kinerja yang cukup baik dalam memprediksi kelas pada dataset dermatitis. Setelah sepuluh *fold* dievaluasi, model ini menghasilkan akurasi yang cukup baik dengan nilai rata-rata sebesar 66.26%.

Tabel 3. Hasil pengujian *Decision Tree*

<i>Fold</i>	Akurasi	<i>Precision</i>	<i>Recall</i>	<i>F1 - Score</i>
1	54.76%	54.54%	54.76%	54.37%
2	78.57%	78.59%	78.57%	78.53%
3	63.41%	66.21%	63.41%	61.33%
4	68.29%	70.45%	68.29%	67.19%
5	63.41%	63.58%	63.41%	63.15%
6	68.29%	68.35%	68.29%	67.95%
7	70.73%	70.73%	70.73%	70.73%
8	56.10%	55.68%	56.10%	53.62%
9	75.61%	80.72%	75.61%	75.02%
10	63.41%	63.31%	63.41%	63.33%
Average	66.26%	67.22%	66.26%	65.52

Hasil evaluasi tabel 3 pada model *Decision Tree* dengan *ADASYN* menunjukkan hasil yang cukup baik. Ini menunjukkan bahwa model *Decision Tree* yang telah dikembangkan mampu mendapatkan tingkat akurasi mencapai 66.26%.



Gambar 5. *Confusion matrix decision tree*

$$\text{Accuracy} = \frac{112 + 161}{112 + 161 + 83 + 56} = 0.6626 = 66.26\%$$

Setelah dilakukan perhitungan manual pada tiap parameter maka didapatkan hasil akurasi yang sama saat perhitungan menggunakan *python*, dan perhitungan manual.

4.8. Permodelan Decision Tree dengan Recursive Elimination

Di tahap ini akan dilakukan percobaan kedua dengan melakukan pengujian model menggunakan *Python*. Pada pengujian ini akan melibatkan implementasi algoritma *Decision Tree* dengan menambah teknik *Recursive Elimination*

4.9. Seleksi fitur

Seleksi fitur menggunakan metode *Recursive Elimination* pada algoritma *Decision Tree* adalah pendekatan bertahap untuk memilih fitur-fitur terbaik yang dapat memberikan kontribusi secara signifikan pada akurasi model. Pada proses ini semua fitur akan ditambahkan secara berurutan dan setiap fiturnya akan dievaluasi performanya menggunakan validasi silang (*cross-validation*) kemudian dilakukan pembagian data secara proporsional (*StratifiedKFold*), dan pada penilaian performa metrik yang digunakan adalah akurasi. *Decision Tree* adalah model yang akan digunakan sebagai *estimator* utama dalam menangani banyak fitur data dan relasi *non-linier* antar variabel. Hasil dari seleksi fitur menggunakan metode *Recursive Elimination* menggunakan *Decision Tree*, fitur yang terpilih adalah Umur, BB, RR, Nadi, dan TB.

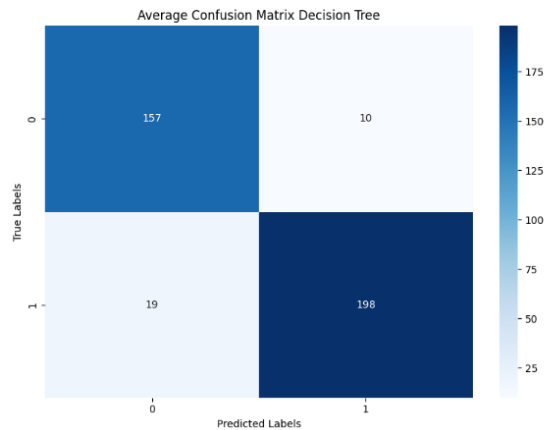
4.10. Permodelan

Pada pengujian menghasilkan kinerja yang sangat baik pada model *Decision Tree* dengan seleksi fitur *Recursive Elimination* serta kesepuluh *fold*. Akurasi pada model ini cukup tinggi dengan nilai rata-rata akurasi mencapai 92.44%. Ini menunjukkan bahwa kemampuan model cukup baik dalam mengklasifikasikan data, dengan tingkat kesalahan yang cukup rendah.

Tabel 4. Hasil pengujian model *decision tree*

<i>Fold</i>	Akurasi	<i>Precision</i>	<i>Recall</i>	<i>F1 - Score</i>
1	87.18%	87.36%	87.18%	87.21%
2	87.18%	87.86%	87.18%	86.97%
3	92.31%	92.47%	92.31%	92.33%
4	100%	100%	100%	100%
5	92.11%	92.18%	92.11%	92.08%
6	92.11%	92.27%	92.11%	92.12%
7	97.37%	97.51%	97.37%	97.37%
8	94.74%	95.32%	94.74%	94.77%
9	92.11%	93.35%	92.11%	92.15%
10	89.47%	91.58%	89.47%	89.53%
Average	92.46%	92.99%	92.46%	92.45%

Hasil evaluasi tabel 4 dari model *Decision Tree* dengan *Recursive Elimination* mendapatkan hasil yang sangat baik, dimana hasil akurasi mengalami peningkatan setelah seleksi fitur diterapkan. Dalam membandingkan hasil evaluasi maka akan dilakukan perhitungan manual nilai *confusion matrix* untuk melihat keakuratan hasil pengujian.



Gambar 6. Confusion matrix decision tree

$$\text{Accuracy} = \frac{157 + 198}{157 + 198 + 10 + 19} = 0.9244 = 92.44\%$$

Tabel 5. Perbandingan hasil akurasi pengujian *decision tree*, *adasyn*, dan *rfe*

Fold	Decision Tree	Decision Tree sesudah ADASYN	Decision Tree Sesudah ADASYN dan Recursive Elimination	Perubahan Decision Tree ke ADASYN	Perubahan Decision Tree ADASYN ke RFE
1	64.10%	82.05%	87.18%	+17.95%	+5.13%
2	51.28%	74.36%	87.18%	+23.08%	+12.82%
3	56.41%	82.05%	92.31%	+25.64%	+10.26%
4	56.41%	79.49%	100%	+23.08%	+20.51%
5	63.16%	76.32%	92.11%	+13.16%	+15.79%
6	65.79%	84.21%	92.11%	+18.42%	+7.9%
7	57.89%	84.21%	97.37%	+26.32%	+13.16%
8	65.79%	81.58%	94.74%	+15.79%	+13.16%
9	60.53%	78.95%	92.11%	+18.42%	+13.16%
10	60.53%	81.58%	89.47%	+21.05%	+7.89%
Average	60.19%	66.26%	92.44%	+20.39%	+11.98%

Berdasarkan pada tabel 5 perbandingan hasil akurasi pengujian diatas menunjukkan bahwa model yang menggunakan *Recursive Elimination* mendapatkan akurasi yang lebih tinggi dalam mengklasifikasi. Pada 10 *folds* dengan *cross validation*, model *Decision Tree* dengan seleksi fitur *Recursive Elimination* menunjukkan peningkatan akurasi dibandingkan tanpa menggunakan seleksi fitur. Hasil akurasi yang meningkat adalah 26.2% ini menunjukkan bahwa seleksi fitur *Recursive Elimination* efektif dalam meningkatkan kinerja model *Decision Tree*.

4.12. Pembahasan

Berdasarkan hasil analisis, penelitian ini mengevaluasi kinerja model *Decision Tree* dalam mengklasifikasikan penyakit dermatitis dengan menerapkan teknik seleksi fitur *RFE* dan *ADASYN*. Pada tahap *preprocessing* meliputi data *cleaning*, data *transformation*, dan data *balancing* menggunakan *ADASYN* untuk menyeimbangkan kelas minoritas dan mayoritas. Validasi menggunakan *confusion matrix* untuk menunjukkan hasil apakah konsisten antara perhitungan manual dan implementasi menggunakan *python*.

Setelah dilakukan perhitungan manual pada tiap parameter maka didapatkan hasil akurasi yang sama saat perhitungan menggunakan *python*, hasil ini sangat baik dan sesuai dengan perhitungan manual, ini menunjukkan bahwa kemampuan model cukup baik dalam menerapkan data latih ke data uji serta mampu mengklasifikasikan data dengan akurat.

4.11. Perbandingan Hasil

Pada perbandingan ini dilakukan antara dua model *Decision Tree*, yaitu menerapkan *ADASYN* pada model lalu dilanjutkan dengan menerapkan seleksi fitur *Recursive Elimination*, kemudian akan dilihat perbedaan hasil akurasi dari setiap model untuk dapat di evaluasi efektif atau tidaknya seleksi fitur dalam meningkatkan kinerja model.

4.13. Hasil Seleksi Fitur RFE

Hasil Hasil seleksi fitur *RFE* yang diterapkan pada dataset dermatitis UPT Puskesmas Bontang Barat berhasil mengidentifikasi fitur-fitur yang paling berpengaruh terhadap hasil klasifikasi yaitu fitur Umur, BB, RR, Nadi, dan TB. Fitur-fitur yang terpilih ini mempunyai pengaruh yang signifikan dalam membedakan kelas pada dataset dermatitis. Dalam penelitian yang menggunakan fitur *Recursive Elimination* pada klasifikasi kanker payudara berhasil mengidentifikasi lima belas fitur terbaik yaitu *concave_points3*, *concavity3*, *symmetry3*, *compactness3*, *smoothness3*, *texture2*, *radius3*, *radius1*, *concavity1*, *concave_points1*, *symmetry1*, *compactness1*, *smoothness1*, *perimeter2*, *dantexture3* [26].

4.14. Hasil Peningkatan Akurasi dari Algoritma Decision Tree, dan ADASYN

Penerapan *Decision Tree* yang telah dikombinasikan dengan teknik oversampling *ADASYN* mampu menghasilkan akurasi 66.26%. *ADASYN* digunakan untuk mengatasi kelas yang tidak seimbang pada dataset dermatitis. Kemudian setelah diterapkan seleksi fitur *RFE* hasil akurasi meningkat signifikan

mencapai 92.44%, peningkatan yang bertambah sebesar 26.18% ini menunjukkan bahwa *RFE* sangat efektif meningkatkan efisiensi dan efektivitas model *Decision Tree* pada klasifikasi dermatitis. Berikut adalah penelitian yang membahas klasifikasi penyakit dengan menggunakan model yang sama. Pada penelitian [15] terkait identifikasi diabetes yang menggabungkan *Decision Tree* dengan *RFE* berhasil mencapai akurasi 75.97%, lalu penelitian [26] yang menerapkan *RFE* pada klasifikasi kanker payudara mencapai akurasi 98%. Selanjutnya penelitian [20] yang menerapkan *ADASYN* berhasil mencapai akurasi 99%. Ini menunjukkan bahwa seleksi fitur sangat penting dalam membantu model mengidentifikasi fitur-fitur yang relevan, sehingga dapat meningkatkan akurasi klasifikasi.

5. KESIMPULAN DAN SARAN

Dari pengujian yang telah dilakukan seleksi fitur *RFE* pada dataset dermatitis berhasil mendapatkan 5 fitur terpilih yang paling signifikan terhadap klasifikasi dermatitis yaitu Umur, BB, RR, Nadi, dan TB. Hasil pengujian dengan menerapkan model *Decision Tree*, *oversampling ADASYN* dan seleksi fitur *RFE* terbukti mampu meningkatkan akurasi dari 66.26% menjadi 92.44%. Pada penelitian selanjutnya penerapan metode *Decision Tree* dengan seleksi fitur *RFE* disarankan tetap menggunakan metode yang sama namun dengan dataset yang berbeda dari berbagai wilayah lain agar mendapatkan hasil yang beragam, dan disarankan untuk mengeksplorasi metode *balancing* seperti *SMOTE* atau *Hybrid Sampling* agar dapat menambah wawasan dalam menangani *class imbalance* pada dataset.

DAFTAR PUSTAKA

- [1] BPS, "Sepuluh Besar Penyakit di Kota Kendari." Accessed: Apr. 12, 2025. [Online]. Available: <https://kendarikota.bps.go.id/id/statistics-table/1/MTE4MSMx/sepuluh-besar-penyakit-di-kota-kendari-2021.html>
- [2] 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.
- [3] M. Ula, A. Zulfikri, A. F. Ulva, and R. A. Rizal, "Penerapan Machine Learning Clustering K-Means dan Linear Regression Dalam Penentuan Tingkat Resiko Tuberkulosis Paru," *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 336–348, 2023, doi: 10.33022/ijcs.v12i1.3162.
- [4] M. Saputra, J. P. Sidabuke, R. P. Sinulingga, and R. B. Tamba, "Analisis Metode Algoritma K-Nearest Neighbor (KNN) dan Naive Bayes Untuk Klasifikasi Diabetes Mellitus," *J. TEKINKOM*, vol. 6, no. 2, pp. 723–729, 2023, doi: 10.37600/tekinkom.v6i2.942.
- [5] Dewi Nasien et al., "Perbandingan Implementasi Machine Learning Menggunakan Metode KNN, Naive Bayes, dan Logistik Regression Untuk Mengklasifikasi Penyakit Diabetes," *JEKIN - J. Tek. Inform.*, vol. 4, no. 1, pp. 10–17, 2024, doi: 10.58794/jekin.v4i1.640.
- [6] L. Nafi and Z. Fatah, "Implementasi Algoritma Decision Tree Untuk Pendeteksian Penyakit Jantung," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 3, no. 2, pp. 160–165, 2024, doi: <https://doi.org/10.70609/jusifor.v3i2.5729>.
- [7] R. N. Ramadhon, A. Ogi, A. P. Agung, R. Putra, S. S. Febrihartina, and U. Firdaus, "Implementasi Algoritma Decision Tree untuk Klasifikasi Pelanggan Aktif atau Tidak Aktif pada Data Bank," *Karimah Tauhid*, vol. 3, no. 2, pp. 1860–1874, 2024, doi: 10.30997/karimahtauhid.v3i2.11952.
- [8] M. Awad and S. Fraihat, "Recursive Feature Elimination with Cross-Validation with Decision Tree: Feature Selection Method for Machine Learning-Based Intrusion Detection Systems," *J. Sens. Actuator Networks*, vol. 12, no. 5, 2023, doi: 10.3390/jsan12050067.
- [9] W. X. Yang, F. F. Wang, Y. Y. Pan, J. Q. Xie, M. H. Lu, and C. G. You, "Comparison of Ischemic Stroke Diagnosis Models Based on Machine Learning," *Front. Neurol.*, vol. 13, 2022, doi: 10.3389/fneur.2022.1014346.
- [10] A. S. Biyantoro and B. Prasetyo, "Penerapan Decision Tree untuk Klasifikasi Status Kesehatan dengan Perbandingan KNN dan Naive Bayes," *Indones. J. Inform. Res. Softw. Eng.*, vol. 4, no. 1, pp. 47–55, 2024, doi: journal.irpi.or.id/index.php/ijirse.
- [11] C. N. Syahputri and M. S. Hasibuan, "Optimasi Klasifikasi Decision Tree Dengan Teknik Pruning Untuk Mengurangi Overfitting," *JSII J. Sist. Inf.*, vol. 11, no. 2, pp. 87–96, 2024, doi: 10.30656/jsii.v11i2.9161.
- [12] T. A. Y. Siswa and R. P. Wibowo, "Komparasi Metode Seleksi Fitur Dalam Prediksi Keterlambatan Pembayaran Biaya Kuliah," *Teknika*, vol. 12, no. 1, pp. 73–82, 2023, doi: 10.34148/teknika.v12i1.601.
- [13] A. M. Priyatno and T. Widiyaningtyas, "A Systematic Literature Review: Recursive Feature Elimination Algorithms," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 9, no. 2, pp. 196–207, 2024, doi: 10.33480/jitk.v9i2.5015.
- [14] J. Teng, C. Zhang, H. Gong, and C. Liu, "Machine Learning-based Urban Noise Appropriateness Evaluation Method and Driving Factor Analysis," *PLoS One*, vol. 19, no. 12, pp. 1–18, 2024, doi: 10.1371/journal.pone.0311571.
- [15] N. A. Maori and N. Azizah, "Penerapan Recursive Feature Elimination (RFE) pada Tree-Based Classifier untuk Identifikasi Risiko Diabetes," *J. Inform. dan Rekayasa Perangkat Lunak*, vol. 6, no. 2, pp. 465–471, 2024, [Online]. Available: <https://publikasiilmiah.unwahas.ac.id/JINRPL/ar>

- ticle/view/12642
- [16] T. Elizabeth, U. Multi, D. Palembang, P. S. Informatika, P. Studi, and S. Informasi, "Prediksi Jenis Kanker Payudara Menggunakan Metode Support Vector Machine Berbasis Recursive Feature Elimination," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, no. x, pp. 1–9, 2024, doi: 10.35957/jatisi.v11i3.10079.
- [17] E. Saputro and D. Rosiyadi, "Penerapan Metode Random Over-Under Sampling Pada Algoritma Klasifikasi Penentuan Penyakit Diabetes," *Bianglala Inform.*, vol. 10, no. 1, pp. 42–47, 2022, doi: 10.31294/bi.v10i1.11739.
- [18] C. Sonjaya, A. F. Nur Masruriyah, D. Sulistya Kusumaningrum, and A. Rizky Pratama, "The Performance Comparison of Classification Algorithm in Order to Detecting Heart Disease," *Intern. (Information Syst. Journal)*, vol. 5, no. 2, pp. 166–175, 2022, doi: 10.32627/internal.v5i2.595.
- [19] A. Nurhopipah and C. Magnolia, "Perbandingan Metode Resampling Pada Imbalanced Dataset Untuk Klasifikasi Komentar Program Mbkm," *J. Publ. Ilmu Komput. dan Multimed.*, vol. 2, no. 1, pp. 9–22, 2023, doi: 10.55606/jupikom.v2i1.862.
- [20] M. Tiara, T. Br, N. S. Fathonah, and M. N. Fauzan, "Pemanfaatan Algoritma ADASYN Dan Support Vector Machine Dalam Meningkatkan Akurasi Prediksi Kanker Paru-Paru," vol. 8, no. 5, pp. 8773–8778, 2024, doi: <https://doi.org/10.36040/jati.v8i5.10752>.
- [21] I. R. Pratama, M. Maimunah, and E. R. Arumi, "Sistem Klasifikasi Penjualan Produk Alat Listrik Terlaris Untuk Optimasi Pengadaan Stok Menggunakan Naïve Bayes," *J. Media Inform. Budidarma*, vol. 6, no. 4, p. 2135, 2022, doi: 10.30865/mib.v6i4.4418.
- [22] A. M. A. Rahim, Ingrid Yanuar Risca Pratiwi, and Muhammad Ainul Fikri, "Klasifikasi Penyakit Jantung Menggunakan Metode Synthetic Minority Over-Sampling Technique Dan Random Forest Classifier," *Indones. J. Comput. Sci.*, vol. 12, no. 5, pp. 2995–3011, 2023, doi: 10.33022/ijcs.v12i5.3413.
- [23] 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.
- [24] D. Anitasari *et al.*, "Klasifikasi Kecelakaan Lalu Lintas di Kota Samarinda Menggunakan Algoritme K-Nearest Neighbor," *J. Mnemon. J. Ilmu Komput. dan Tek. Inform.*, vol. 8, no. 1, pp. 33–39, 2025, doi: <https://doi.org/10.36040/mnemonic.v8i1.12747>.
- [25] D. Normawati and S. A. Prayogi, "Implementasi Naïve Bayes Classifier Dan Confusion Matrix Pada Analisis Sentimen Berbasis Teks Pada Twitter," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 697–711, 2021, [Online]. Available: <https://ejurnal.tunasbangsa.ac.id/index.php/jsakti/article/view/369>
- [26] H. Sundari, M. A. Amrustian, A. Dwi, and P. Wicaksono, "Penerapan Recursive Feature Elimination pada Support Vector Machine untuk Klasifikasi Kanker Payudara," vol. 8798, pp. 60–65, 2024