Jurnal Sistem Informasi dan Sains Teknologi
Vol.8 No.1, Februari 2026
ISSN 2684-8260

IMPLEMENTASI METODE C4.5 UNTUK PREDIKSI RISK-LEVEL KEKAMBUHAN KANKER TIROID TERDIFERENSIASI

Hesti Ifada Mustio[1)], Hasbi Firmansyah[2)], Wahyu Asriani[3)]
Universitas Pancasakti Tegal
Email: mustiohesty@gmail.com[1)], hasbifirmansyah@upstegal.ac.id[2)], asriani1409@gmail.com[3)]

ABSTRAK

Kanker tiroid terdiferensiasi (Differentiated Thyroid Cancer/DTC) umumnya berprognosis baik, namun sebagian pasien tetap mengalami kekambuhan sehingga diperlukan stratifikasi risiko yang objektif dan mudah diinterpretasi. Penelitian ini bertujuan membangun dan mengevaluasi model klasifikasi berbasis algoritma C4.5 untuk memprediksi tingkat risiko kekambuhan DTC (Low, Intermediate, High) menggunakan dataset "Differentiated Thyroid Cancer Recurrence" dari UCI Machine Learning Repository. Dataset ini merupakan kohort retrospektif 383 pasien dengan follow-up 10 tahun yang memuat 16 fitur klinikopatologis. Algoritma C4.5 dipilih karena mampu menangani atribut numerik dan kategorik, memodelkan hubungan nonlinier, serta menghasilkan pohon keputusan dan aturan if–then yang mudah ditelusuri sehingga sesuai untuk prototipe sistem pendukung keputusan klinis. Tahapan penelitian meliputi prapemrosesan dan pengkodean atribut, pembagian data menjadi data latih dan uji, pelatihan pohon keputusan C4.5 dengan pruning, serta evaluasi menggunakan akurasi, error klasifikasi, presisi tertimbang, dan recall tertimbang. Pada 116 data uji, model menghasilkan akurasi 81,03%, error klasifikasi 18,97%, presisi tertimbang 78,11%, dan recall tertimbang 78,77%. Kesalahan klasifikasi terutama terjadi antara kelas Low dan Intermediate, sementara sebagian besar pasien berisiko tinggi teridentifikasi dengan baik. Temuan ini menunjukkan bahwa C4.5 andal memetakan pola klinis dan berpotensi dimanfaatkan sebagai alat bantu awal dalam stratifikasi risiko kekambuhan DTC.

Kata Kunci : Kanker tiroid terdiferensiasi; Kekambuhan; Stratifikasi risiko; Pohon keputusan C4.5; Data mining; Sistem pendukung keputusan klinis

ABSTRACT

Differentiated thyroid cancer (DTC) generally has a favorable prognosis, yet some patients develop recurrence, making objective and interpretable risk stratification essential. This study aims to develop and evaluate a C4.5-based model to predict DTC recurrence risk levels (Low, Intermediate, High) using the "Differentiated Thyroid Cancer Recurrence" dataset from the UCI Machine Learning Repository. The dataset includes 383 patients with 10-year follow-up and 16 clinicopathological features. C4.5 was chosen for its ability to handle numerical and categorical attributes, capture nonlinear relationships, and produce traceable decision trees and if–then rules suitable for prototyping clinical decision support. The workflow comprised data preprocessing, attribute encoding, splitting the data into training and test sets, training a pruned C4.5 decision tree, and evaluating performance using accuracy, classification error, weighted precision, and weighted recall. On 116 test instances, the model achieved an accuracy of 81.03%, a classification error of 18.97%, a weighted precision of 78.11%, and a weighted recall of 78.77%. Most misclassifications occurred between Low and Intermediate classes, while high-risk patients were

largely identified correctly. These findings indicate that C4.5 can capture clinical patterns and is promising as an initial tool for DTC recurrence risk stratification.

Keyword : *Differentiated thyroid cancer; Recurrence; Risk stratification; C4.5 Decision tree; Data mining; Clinical decision support system.*

I. PENDAHULUAN

1.1 Latar Belakang

Kanker tiroid merupakan salah satu jenis keganasan endokrin yang cukup sering dijumpai. Sebagian besar kasusnya tergolong sebagai kanker tiroid *terdiferensiasi (Differentiated Thyroid Cancer, DTC)* yang secara umum memiliki prognosis baik, terutama apabila terdeteksi dan ditangani pada stadium awal. Meskipun demikian, tidak sedikit pasien yang tetap mengalami kekambuhan setelah menjalani terapi awal, baik dalam bentuk rekurensi lokal, regional, maupun metastasis jauh. Kekambuhan ini berdampak pada meningkatnya morbiditas, kebutuhan terapi lanjutan, serta beban biaya layanan kesehatan, sehingga prediksi risiko kekambuhan menjadi aspek yang sangat penting dalam tata laksana jangka panjang pasien kanker tiroid terdiferensiasi.

Dalam praktik klinis, penentuan *risk-level* biasanya bertumpu pada penilaian dokter yang mengintegrasikan faktor klinis, patologi, dan pemeriksaan penunjang. Faktor-faktor tersebut dapat mencakup usia pasien, ukuran dan karakteristik histopatologi tumor, keterlibatan kelenjar getah bening, keberadaan metastasis jauh, serta respons terhadap terapi awal. Meskipun pendekatan ini telah lama digunakan, penilaian yang terlalu bergantung pada pengalaman dan intuisi klinisi berpotensi menimbulkan variasi antar dokter maupun antar fasilitas pelayanan kesehatan. Di sisi lain, volume data rekam medis yang terus meningkat menjadikan analisis manual semakin tidak efisien dan berisiko mengabaikan pola-pola penting yang sebenarnya dapat dimanfaatkan untuk memperkirakan *risk-level*. Perkembangan data mining dan machine learning membuka peluang untuk mengolah data rekam medis secara lebih sistematis, terstruktur, dan konsisten. Berbagai algoritma klasifikasi telah dikembangkan untuk menggali pola tersembunyi dalam data dan menghasilkan model prediksi yang terukur kinerjanya. Salah satu algoritma yang banyak digunakan adalah C4.5, yang membangun model dalam bentuk pohon keputusan.

Karakteristik utama metode ini adalah kemampuannya menghasilkan aturan keputusan yang relatif mudah dipahami dan ditelusuri kembali oleh pengguna, sehingga sangat relevan untuk diaplikasikan pada domain medis yang menuntut transparansi, interpretabilitas, dan akuntabilitas dalam pengambilan keputusan. Dalam konteks prediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi, algoritma C4.5 dapat memanfaatkan berbagai atribut klinis dan penunjang yang tercatat pada rekam medis, kemudian memetakannya ke dalam kategori risiko tertentu, misalnya risiko rendah, sedang, atau tinggi.

Model yang dihasilkan diharapkan mampu membantu klinisi melakukan stratifikasi risiko secara lebih objektif dan berbasis data. Dengan adanya estimasi risiko yang lebih terukur, perencanaan tindak lanjut, jadwal kontrol, dan strategi monitoring pasien dapat disesuaikan dengan profil risiko masing-masing, sehingga pemanfaatan sumber daya layanan kesehatan menjadi lebih tepat

sasaran. Bertolak dari kebutuhan tersebut, penelitian ini berangkat dari pertanyaan mengenai bagaimana metode C4.5 dapat diimplementasikan untuk memprediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi berdasarkan data rekam medis pasien, sejauh mana kinerja model klasifikasi yang dihasilkan dalam membedakan kategori risiko kekambuhan, dan bagaimana bentuk aturan keputusan yang muncul dari pohon keputusan yang dibangun serta potensi pemanfaatannya sebagai alat bantu pengambilan keputusan klinis.

Agar kajian tetap terarah dan realistis terhadap keterbatasan waktu dan sumber daya, penelitian ini tidak melakukan pengambilan data langsung dari fasilitas pelayanan kesehatan, tetapi menggunakan data sekunder berupa *Differentiated Thyroid Cancer Recurrence* Dataset dari *UCI Machine Learning Repository*. Dataset ini memuat informasi demografis, temuan klinis, hasil pemeriksaan penunjang, dan riwayat terapi awal pasien dengan kanker tiroid terdiferensiasi sesuai dengan atribut yang tersedia pada dataset.
Kategori *risk-level* kekambuhan ditentukan berdasarkan label atau penilaian yang telah ditetapkan sesuai standar atau kebijakan di fasilitas terkait. Kajian metode klasifikasi difokuskan pada algoritma C4.5, sementara penggunaan metode lain hanya diposisikan sebagai pembanding apabila diperlukan dalam ruang lingkup yang terbatas. Evaluasi terhadap model dilakukan menggunakan metrik umum dalam akurasi, *classification errors, weighted mean precision, dan weighted mean recall*, untuk memperoleh gambaran yang komprehensif mengenai kinerja model. Sejalan dengan fokus tersebut, tujuan utama penelitian ini adalah membangun model prediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi menggunakan metode C4.5 berbasis data rekam medis pasien, menilai kinerja model tersebut dalam mengklasifikasikan *risk-level*, serta menggali aturan keputusan yang terbentuk dari pohon keputusan sebagai dasar penyusunan rekomendasi yang berpotensi mendukung proses pengambilan keputusan klinis. Penelitian ini diharapkan memberikan kontribusi teoritis dan praktis. Dari sisi teoritis, hasil penelitian dapat memperkaya literatur mengenai penerapan metode data mining, khususnya algoritma C4.5, dalam bidang kesehatan, sekaligus memberikan ilustrasi mengenai pemodelan *risk-level* kekambuhan kanker tiroid terdiferensiasi yang bersifat *interpretable.*

Dari sisi praktis, model dan aturan keputusan yang dihasilkan diharapkan dapat dimanfaatkan sebagai alat bantu awal bagi klinisi dalam melakukan stratifikasi risiko kekambuhan, mendukung perencanaan tindak lanjut dan monitoring pasien secara lebih terpersonalisasi, serta menjadi salah satu rujukan dalam pengembangan sistem pendukung keputusan di lingkungan fasilitas pelayanan kesehatan.

Berdasarkan uraian tersebut, permasalahan yang dikaji dalam penelitian ini adalah bagaimana membangun model klasifikasi berbasis algoritma C4.5 untuk memprediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi dan menilai kinerja model tersebut. Tujuan penelitian ini adalah: (1) membangun model C4.5 menggunakan data klinis pasien DTC dari *UCI Machine Learning Repositor*y, (2) mengevaluasi kinerja model menggunakan metrik akurasi, *classification error, weighted mean precision, dan weighted mean recall*, serta (3) menurunkan aturan keputusan dari pohon yang terbentuk sebagai dasar rekomendasi pendukung keputusan klinis.

1.2 Analisa Permasalahan

Penentuan *risk-level* kanker tiroid terdiferensiasi (DTC) saat ini masih bergantung pada penilaian subjektif dokter, sehingga menimbulkan variasi stratifikasi risiko antar klinisi dan fasilitas

kesehatan. Pada saat yang sama, volume data rekam medis yang terus meningkat membuat analisis manual tidak efisien dan berpotensi mengabaikan pola penting untuk prediksi kekambuhan. Perkembangan data mining dan *machine learning* memberikan peluang untuk membangun model klasifikasi yang sistematis, akurat, dan konsisten, namun penerapannya untuk prediksi *risk-level* kekambuhan DTC masih terbatas. Algoritma C4.5, dengan keluaran berupa pohon keputusan yang *interpretable*, berpotensi dimanfaatkan untuk memprediksi *risk-level* sekaligus menghasilkan aturan keputusan yang dapat dijadikan alat bantu dalam pengambilan keputusan klinis.

1.3 Ruang Lingkup Penelitian

Penelitian ini terbatas pada pemanfaatan data sekunder pasien kanker tiroid terdiferensiasi yang terdapat dalam dataset *Differentiated Thyroid Cancer Recurrence* dari *UCI Machine Learning Repository*. Analisis difokuskan pada atribut demografis, klinis, dan klinikopatologis yang sudah tersedia, termasuk label *risk-level* (rendah, sedang, tinggi), tanpa pengumpulan data primer maupun penambahan variabel lain di luar dataset. Penelitian hanya mengkaji pembangunan dan evaluasi model klasifikasi menggunakan algoritma C4.5, mulai dari prapemrosesan data, pembagian data latih dan uji, pelatihan model, hingga pengukuran kinerja dengan akurasi, *classification error, weighted mean precision, dan weighted mean recall*, serta penurunan aturan keputusan sebagai prototipe alat bantu awal stratifikasi risiko kekambuhan, tanpa menguji implementasinya langsung dalam sistem klinis.

1.4 Rumusan Masalah

Secara garis besar, penelitian ini berupaya menjawab bagaimana membangun model klasifikasi berbasis algoritma C4.5 dari data rekam medis pasien kanker tiroid terdiferensiasi pada dataset *Differentiated Thyroid Cancer Recurrence* sehingga mampu memprediksi *risk-level* (rendah, sedang, tinggi) secara lebih objektif. Penelitian ini juga menanyakan seberapa baik kinerja model yang dihasilkan ketika dievaluasi menggunakan akurasi, *classification error, weighted mean precision, dan weighted mean recall*. Selain itu, dikaji pula bagaimana aturan keputusan yang diturunkan dari pohon C4.5 dapat diformulasikan dan dimanfaatkan sebagai rekomendasi awal untuk mendukung pengambilan keputusan klinis dalam stratifikasi risiko kekambuhan.

1.5 Urgensi Penelitian

Penelitian ini memiliki urgensi yang tinggi baik dari sisi klinis maupun sisi pengembangan ilmu. Secara klinis, kekambuhan DTC berdampak pada meningkatnya morbiditas, kebutuhan terapi lanjutan, dan biaya layanan kesehatan, sehingga diperlukan stratifikasi risiko yang lebih objektif untuk mengarahkan tindak lanjut dan monitoring pasien secara tepat sasaran. Ketergantungan pada penilaian subjektif klinisi tanpa dukungan model prediksi berbasis data berpotensi menimbulkan ketidakkonsistenan dalam penentuan kategori risiko, yang pada akhirnya dapat memengaruhi kualitas tata laksana jangka panjang pasien.

Dari perspektif metodologis, pemanfaatan *data mining* dan *machine learning,* khususnya algoritma C4.5 yang *interpretable*, masih perlu dikembangkan dan didokumentasikan secara sistematis dalam konteks prediksi kekambuhan DTC. Penelitian ini mendesak untuk dilakukan guna menyediakan bukti empiris mengenai kinerja dan kelayakan algoritma C4.5 sebagai bagian dari sistem pendukung keputusan klinis. Hasil penelitian diharapkan dapat menjadi dasar bagi penelitian lanjutan yang mengintegrasikan model serupa ke dalam praktik klinis sehari-hari.

1.6 Tujuan Penelitian

Secara umum, penelitian ini bertujuan untuk mengembangkan model klasifikasi yang interpretable guna memprediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi berbasis data rekam medis. Secara khusus, tujuan penelitian adalah:

1 Membangun model klasifikasi menggunakan algoritma C4.5 dengan memanfaatkan data klinis dan klinikopatologis pasien DTC dari *UCI Machine Learning Repository*.
2 Mengevaluasi kinerja model C4.5 dalam mengklasifikasikan *risk-level* kekambuhan DTC menggunakan metrik akurasi, *classification error, weighted mean precision, dan weighted mean recall.*
3 Menurunkan aturan keputusan *(decision rules)* dari pohon keputusan C4.5 yang terbentuk dan menyusunnya sebagai rekomendasi awal pendukung keputusan klinis dalam stratifikasi risiko kekambuhan DTC.

1.7 Manfaat Penelitian

Adapun manfaat dari penelitian ini:

1. Menambah referensi ilmiah mengenai penerapan data mining, khususnya algoritma C4.5, dalam prediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi.
2. Menyajikan contoh model klasifikasi medis yang bersifat *interpretable* melalui pohon keputusan dan aturan *if–then.*
3. Menyediakan model prediksi yang dapat digunakan sebagai alat bantu awal bagi klinisi dalam stratifikasi risiko kekambuhan pasien DTC.
4. Mendukung perencanaan tindak lanjut dan monitoring pasien yang lebih terarah dan terpersonalisasi berdasarkan tingkat risiko.
5. Menjadi masukan awal bagi pengembangan sistem pendukung keputusan klinis berbasis data di fasilitas pelayanan kesehatan.

II. LANDASAN TEORI

2.1 Kanker Tiroid Terdiferensiasi dan Kekambuhan

Kanker tiroid merupakan keganasan endokrin yang cukup sering dijumpai, dengan mayoritas kasus tergolong sebagai kanker tiroid terdiferensiasi *(Differentiated Thyroid Cancer/DTC)*. Secara umum DTC memiliki prognosis yang baik, terutama apabila terdeteksi dan ditangani pada stadium awal. Namun demikian, sebagian pasien tetap mengalami kekambuhan dalam bentuk rekurensi lokal, regional, maupun metastasis jauh setelah terapi awal. Kondisi ini berdampak pada peningkatan morbiditas, kebutuhan terapi lanjutan, serta beban biaya layanan kesehatan, sehingga prediksi dan stratifikasi risiko kekambuhan menjadi bagian penting dalam tata laksana jangka panjang pasien DTC.

Dalam praktik klinis, *risk-level* biasanya ditentukan berdasarkan kombinasi faktor klinis dan patologi, seperti usia, ukuran dan karakteristik histopatologi tumor, keterlibatan kelenjar getah bening, adanya metastasis jauh, serta respons terhadap terapi awal. Penilaian ini sangat bergantung pada pengalaman dan intuisi klinisi, sehingga berpotensi menimbulkan variasi antar dokter maupun antar fasilitas kesehatan dan menyebabkan stratifikasi risiko yang kurang konsisten.

2.2 Data Mining dan Machine Learning dalam Bidang Medis

Data mining merupakan proses penggalian pola atau pengetahuan tersembunyi dari sekumpulan data berukuran besar dengan memanfaatkan teknik statistik, *machine learning, dan artificial intelligence*. Dalam bidang medis, data mining digunakan untuk analisis rekam medis, prediksi luaran klinis, deteksi dini penyakit, hingga pengembangan sistem pendukung keputusan klinis.

Machine learning sebagai cabang dari data mining berfokus pada pengembangan algoritma yang mampu "belajar" dari data untuk melakukan prediksi atau klasifikasi. Berbagai algoritma klasifikasi, seperti *logistic regression, support vector machine, random forest, gradient boosting*, maupun pohon keputusan, telah digunakan untuk memodelkan hubungan kompleks antar variabel klinis dan memprediksi luaran seperti kekambuhan kanker, mortalitas, atau respons terapi.

Pada konteks kekambuhan kanker tiroid, *machine learning* dimanfaatkan untuk mengolah atribut klinikopatologis dan faktor risiko lainnya sehingga menghasilkan model prediksi yang terukur kinerjanya dan dapat dibandingkan dengan stratifikasi risiko konvensional.

2.3 Pohon Keputusan dan Algoritma C4.5

Pohon keputusan *(decision tree)* adalah salah satu metode klasifikasi yang merepresentasikan proses pengambilan keputusan dalam bentuk struktur pohon, yang tersusun atas *node* akar, *node* internal, dan *node* daun. Setiap *node* internal merepresentasikan pengujian terhadap suatu atribut, cabang merepresentasikan hasil pengujian, sedangkan *node* daun merepresentasikan kelas atau keputusan akhir. Proses pembentukan pohon dimulai dari pemilihan atribut yang paling informatif, sehingga data dapat terpecah ke dalam kelompok-kelompok yang lebih homogen. Dengan representasi ini, alur penalaran model dapat diikuti secara visual dari akar hingga daun, sehingga memudahkan penelusuran bagaimana suatu keputusan dihasilkan.

Algoritma C4.5 merupakan pengembangan dari ID3 yang dirancang untuk membangun pohon keputusan berbasis konsep entropi dan *information gain*. Algoritma ini menggunakan *gain ratio* sebagai kriteria pemilihan atribut terbaik pada setiap *node* untuk mengurangi bias terhadap atribut yang memiliki banyak kategori. C4.5 mampu menangani atribut kategorik maupun numerik dengan melakukan proses pemotongan nilai numerik menjadi beberapa interval yang bermakna. Selain itu, C4.5 mendukung penanganan nilai yang hilang dan menerapkan mekanisme pemangkasan *(pruning)* untuk mengurangi *overfitting* dan meningkatkan kemampuan generalisasi model.

Keunggulan penting C4.5 adalah sifatnya yang mudah diinterpretasikan karena pohon keputusan yang dihasilkan dapat diterjemahkan menjadi sekumpulan aturan *if–then*. Aturan-aturan tersebut relatif mudah dipahami oleh pengguna non-teknis, sehingga hasil model tidak hanya berupa angka, tetapi juga logika keputusan yang eksplisit. Dalam konteks medis, karakteristik ini sangat relevan karena klinisi membutuhkan model yang transparan untuk menilai kesesuaian prediksi dengan pengetahuan klinis yang sudah ada. Dengan demikian, C4.5 menjadi salah satu algoritma yang tepat untuk dikembangkan sebagai komponen sistem pendukung keputusan klinis yang menuntut transparansi, interpretabilitas, dan akuntabilitas.

2.4 Evaluasi Kinerja Model Klasifikasi

Evaluasi kinerja model klasifikasi dilakukan menggunakan berbagai metrik untuk menilai ketepatan prediksi dan kemampuan model dalam membedakan kelas-kelas yang ada. Beberapa metrik umum yang digunakan antara lain:

1. **Akurasi (Accuracy)**
   Mengukur proporsi prediksi yang benar terhadap seluruh prediksi. Akurasi memberikan gambaran umum performa model, namun dapat kurang informatif bila distribusi kelas tidak seimbang.
2. **Classification Error**
   Merupakan komplemen dari akurasi, yaitu proporsi prediksi yang salah terhadap seluruh prediksi. Nilai ini menggambarkan ruang perbaikan model secara global.
3. **Precision**
   Precision per kelas mengukur seberapa besar proporsi prediksi positif pada kelas tertentu yang benar-benar berasal dari kelas tersebut. *Precision* yang tinggi menunjukkan bahwa model jarang memberikan "alarm palsu" *(false positive)* untuk kelas tersebut.
4. **Recall (Sensitivity)**
   Recall per kelas mengukur kemampuan model dalam mengenali seluruh sampel aktual pada kelas tersebut. Nilai *recall* yang tinggi penting terutama pada kelas risiko tinggi, karena berkaitan dengan kemampuan model mendeteksi pasien yang benar-benar membutuhkan pemantauan intensif.
5. **Weighted Mean Precision dan Weighted Mean Recall**
   Pada klasifikasi multi-kelas, *precision* dan *recall* per kelas dapat dirata-ratakan dengan bobot tertentu (misalnya bobot yang sama untuk tiap kelas) untuk menghasilkan gambaran performa rata-rata model secara lebih seimbang antar kelas.

2.5 Sistem Pendukung Keputusan Klinis Berbasis Model Prediksi

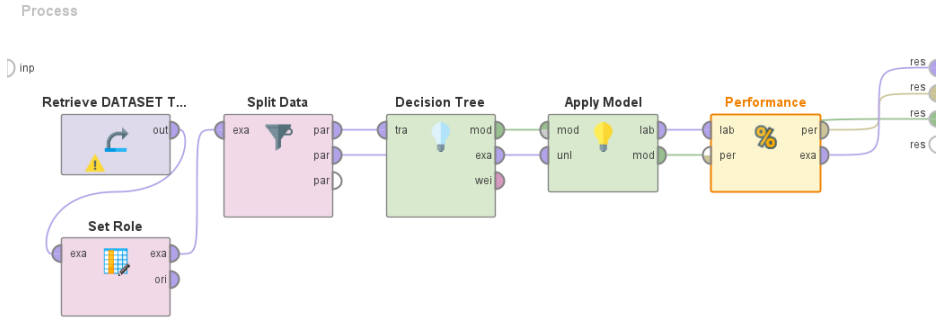
Sistem pendukung keputusan klinis *(Clinical Decision Support System/CDSS)* adalah sistem informasi yang dirancang untuk membantu klinisi dalam pengambilan keputusan dengan memberikan informasi, rekomendasi, atau peringatan berbasis data dan pengetahuan. Dalam konteks DTC, model prediksi *risk-level* kekambuhan yang dibangun dengan algoritma C4.5 dapat diintegrasikan sebagai komponen CDSS untuk membantu stratifikasi risiko secara lebih objektif dan konsisten. Integrasi ini dapat berupa tampilan kategori risiko kekambuhan berdasarkan data klinis pasien yang dimasukkan, misalnya usia, karakteristik tumor, status kelenjar getah bening, dan respons terapi awal.

Aturan keputusan yang dihasilkan dari pohon C4.5 dapat diterjemahkan ke dalam bentuk logika klinis yang mudah dipahami, misalnya kombinasi atribut tertentu yang mengarah pada kategori risiko rendah, sedang, atau tinggi. Penyajian aturan dalam format *if–then* menjadikan model lebih transparan dan dapat ditelusuri kembali oleh klinisi saat mengevaluasi hasil prediksi. Dengan demikian, landasan teori ini menegaskan bahwa pemanfaatan data mining dan algoritma C4.5 berpotensi mendukung klinisi dalam menyusun rencana tindak lanjut dan monitoring pasien DTC yang lebih terarah dan terpersonalisasi.

III. METODE PENELITIAN

3.1 Desain Penelitian dan Workflow Pemodelan

Penelitian ini menggunakan pendekatan kuantitatif berbasis data mining untuk membangun model klasifikasi yang memprediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi *(Differentiated Thyroid Cancer Recurrence/DTC)*. Model yang dibangun adalah pohon keputusan menggunakan algoritma C4.5, sedangkan evaluasi kinerja model dilakukan dengan metrik akurasi, *classification error, weighted mean precision, dan weighted mean recall*. Seluruh proses pemodelan diimplementasikan menggunakan RapidMiner agar tahapan prapemrosesan, pembagian data, pelatihan, pengujian, dan evaluasi dapat dilakukan secara konsisten serta dapat direplikasi.



Gambar III.1 Desain Penelitian dan Workflow Pemodelan

Workflow pemodelan yang digunakan pada penelitian ini ditunjukkan pada Gambar 1. Secara ringkas, proses dimulai dari pengambilan dataset DTC dari *UCI Machine Learning Repository*, dilanjutkan prapemrosesan dan penetapan label kelas *(Risk)* sebagai variabel target, pembagian data menjadi data latih dan data uji dengan rasio 70:30, pembangunan model C4.5 pada data latih, penerapan model pada data uji, serta evaluasi kinerja dan penarikan kesimpulan.

3.2 Data dan Sumber Data

Dataset yang digunakan dalam penelitian ini berasal dari *UCI Machine Learning Repository* dengan nama *Differentiated Thyroid Cancer Recurrence*. Dataset tersebut memuat 383 baris data pasien kanker tiroid terdiferensiasi yang dikumpulkan selama kurang lebih 15 tahun, dengan setiap pasien diikuti minimal selama 10 tahun untuk memantau perkembangan kondisi klinis dan status risiko. Data disajikan dalam bentuk tabular, sehingga setiap baris merepresentasikan satu pasien, sedangkan setiap kolom merepresentasikan atribut yang menggambarkan karakteristik demografis, riwayat klinis, temuan pemeriksaan, serta informasi klinikopatologis yang relevan terhadap penentuan tingkat risiko. Ringkasan karakteristik dataset yang digunakan pada penelitian ini disajikan pada Tabel 3.1.data mentah secara langsung. Penyajian kamus variabel ini juga membantu memperjelas atribut mana yang menjadi masukan model dan atribut mana yang menjadi keluaran yang diprediksi. Ringkasan pada Tabel 3.1 menunjukkan karakteristik dataset yang digunakan sebagai dasar tahap prapemrosesan dan pemodelan pada subbab berikutnya. Secara keseluruhan, dataset memiliki 16 fitur, di mana 13 di antaranya merupakan fitur klinikopatologis yang digunakan sebagai variabel prediktor dalam proses klasifikasi. Atribut target pada penelitian ini adalah *Risk*, yaitu label kelas yang memetakan pasien ke dalam kategori *risk-level* rendah

(Low), sedang *(Intermediate)*, dan tinggi *(High)*. Penetapan Risk sebagai variabel target dilakukan karena atribut ini secara langsung merepresentasikan keluaran klasifikasi yang ingin diprediksi oleh model, sementara atribut lainnya berfungsi menjelaskan pola pembeda antar kelas risiko.

Tabel III.1 Ringkasan Dataset

Parameter	**Nilai**
Sumber	UCI Machine
Nama Dataset	Differentiated Thyroid Cancer Reccurence
Jumlah Instance	383
Jumlah Atribut	16 Fitur + Label Risk
Tipe Data	Tabular (Campuran numerik & Kategorikal)
Task	Klasifikasi
Missing Value	Tidak Ada
Tools Pemodelan	Rapid Miner

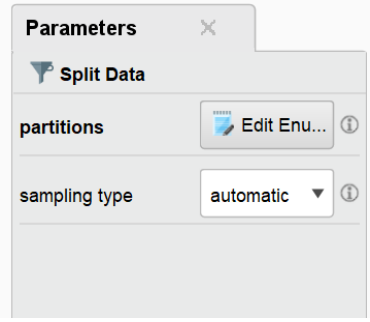
Untuk menjaga keterbacaan dan konsistensi analisis, atribut-atribut pada dataset dikelompokkan berdasarkan fungsinya, yaitu: (1) atribut demografis (misalnya usia dan jenis kelamin), (2) atribut riwayat dan kebiasaan (misalnya status merokok dan riwayat terkait), (3) atribut pemeriksaan/temuan klinis, serta (4) atribut klinikopatologis dan staging (misalnya kategori T, N, M, dan stage bila digunakan sebagai prediktor). Rincian nama atribut, tipe data, serta peran atribut (prediktor atau label) disajikan pada Tabel Tabel Kamus Variabel agar pembaca dapat memahami struktur dataset secara sistematis tanpa perlu mengacu pada data mentah.

3.3 Prapemrosesan Data

Tahap prapemrosesan dilakukan untuk memastikan atribut pada dataset memiliki peran yang sesuai dengan kebutuhan pemodelan klasifikasi di RapidMiner. Pada penelitian ini, prapemrosesan difokuskan pada penetapan variabel target menggunakan operator *Set Role*, yaitu menetapkan atribut *Risk* sebagai label (variabel target) dan atribut lainnya sebagai regular *attribute* (prediktor). Penetapan ini penting agar proses pembelajaran model C4.5 dan evaluasi performa dilakukan terhadap target yang tepat. Output tahap prapemrosesan adalah dataset yang telah memiliki struktur peran atribut yang benar dan siap digunakan pada tahap pembagian data *(Split Data)*, pelatihan model *(Decision Tree)*, penerapan model *(Apply Model)*, dan evaluasi *(Performance)*.

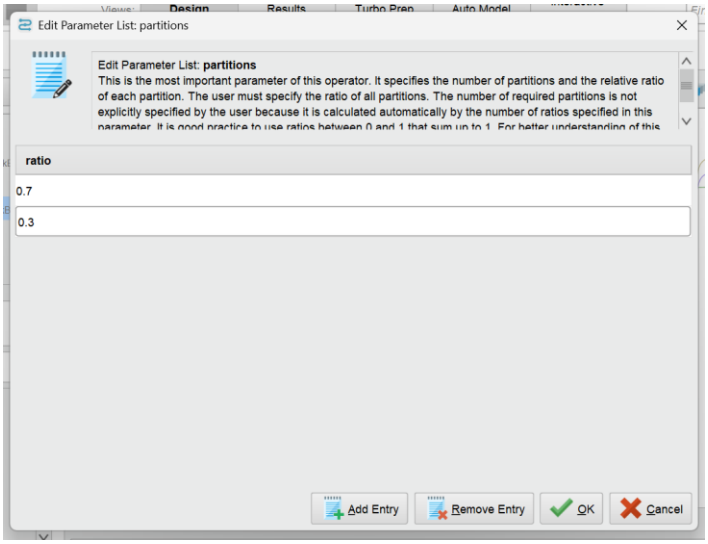
3.4 Pembangunan Model C4.5

Pembangunan model dilakukan setelah data siap diproses dan variabel target telah ditetapkan sebagai label. Pada penelitian ini, dataset dibagi menjadi data pelatihan dan data pengujian menggunakan operator *Split Data* pada RapidMiner. Pengaturan *sampling type* menggunakan nilai *automatic* (Gambar 2) sehingga proses pemilahan *record* dilakukan otomatis oleh RapidMiner sesuai konfigurasi operator.



Gambar III.2 Parameter Splitt Data

Rasio partisi ditetapkan 70:30, yaitu 70% untuk data pelatihan dan 30% untuk data pengujian (Gambar 4). Skema ini bertujuan menyediakan data latih yang memadai untuk membentuk struktur pohon keputusan sekaligus menyediakan data uji yang cukup untuk mengukur kemampuan generalisasi model pada data yang tidak terlibat dalam proses pembelajaran.



Gambar III.3 Rasio Partisi

Setelah pembagian data, model C4.5 dibangun pada data pelatihan dengan langkah utama sebagai berikut:

1 Menghitung nilai entropi awal berdasarkan distribusi label kelas *Risk* pada data pelatihan.
2 Untuk setiap atribut kandidat, menghitung *information gain* dan *gain ratio* sebagai dasar pemilihan atribut pemisah.
3 Memilih atribut dengan *gain ratio* tertinggi sebagai *splitting attribute*, lalu membentuk cabang-cabang pohon sesuai nilai atribut.
4 Mengulangi proses pemilihan atribut secara rekursif sampai node daun homogen (berisi satu kelas *Risk*) atau tidak ada atribut yang bermakna untuk pemisahan.

Implementasi dilakukan menggunakan operator C4.5 *Decision Tree* pada RapidMiner dengan kriteria pemilihan atribut *gain ratio*. Untuk mengurangi risiko *overfitting*, diterapkan *pruning* (*apply pruning* = Ya) dengan *confidence* = 0,1 serta *prepruning* (*apply prepruning* = Ya) dengan

minimal *gain* = 0,01 dan minimal *leaf size* = 2. Kedalaman pohon dibatasi dengan *maximal depth* = 10. Rangkuman parameter model disajikan pada Tabel 2.

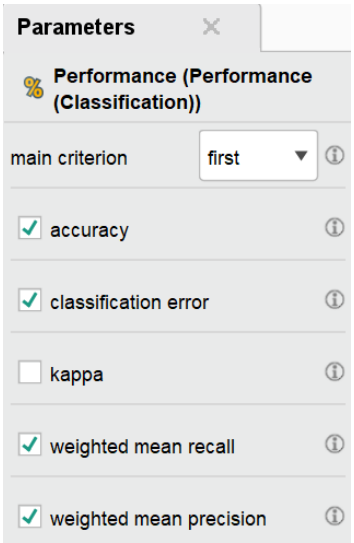
Tabel III.2 Pengaturan Parameter Model

Criterion	Max Depth	Apply Pruning	Confidence	Apply Prepruning	Minimal Gain	Minimal Leaf Size
Gain Ratio	10	Ya	0,1	Ya	0,01	2

Output tahap ini adalah model C4.5 terlatih yang selanjutnya diterapkan pada data pengujian menggunakan operator *Apply Model* untuk menghasilkan label prediksi *(pred. Low, pred. Intermediate, pred. High)* beserta nilai *confidence* untuk setiap kelas.

3.5 Metode Evaluasi Matrik

Evaluasi kinerja bertujuan menilai seberapa baik model C4.5 memprediksi *risk-level (Low, Intermediate, High)* pada data uji yang tidak terlibat saat pelatihan. Evaluasi dilakukan menggunakan operator *Performance (Classification)* pada RapidMiner dengan metrik utama: *accuracy, classification error, weighted mean precision, dan weighted mean recall* (Gambar 4). Seluruh metrik dihitung berdasarkan confusion matrix hasil prediksi sehingga interpretasi dapat dilakukan baik secara global maupun per kelas.



Gambar III.4 Parameter Performance Classification

3.5.1 Akurasi

Akurasi mengukur proporsi prediksi yang benar terhadap seluruh prediksi. Untuk kasus multi-kelas, akurasi dapat dipahami sebagai jumlah prediksi benar pada diagonal *confusion matrix* dibagi total data uji. Secara umum, akurasi dirumuskan sebagai:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Keterangan :
TP (True Positives) adalah jumlah sampel yang sebenarnya positif dan diprediksi model sebagai positif dengan benar.
TN (True Negatives) adalah jumlah sampel yang sebenarnya negatif dan diprediksi model sebagai negatif dengan benar.
FP (False Positives) adalah jumlah sampel yang sebenarnya negatif tetapi diprediksi model secara keliru sebagai positif.
FN (False Negatives) adalah jumlah sampel yang sebenarnya positif tetapi diprediksi model secara keliru sebagai negatif.

3.5.2 Classification Error

Classification error merupakan komplemen dari akurasi, yaitu proporsi prediksi yang salah terhadap seluruh prediksi. Metrik ini menunjukkan besaran kesalahan klasifikasi model dan menjadi indikator ruang perbaikan. Dirumuskan sebagai:

$$CE = \frac{E}{N} \times 100\% \tag{2}$$

Keterangan
CE *(Classification Error)* : Tingkat kesalahan klasifikasi model (proporsi prediksi yang salah).
E : Jumlah prediksi salah
N : Jumlah data uji
Nilai klasifikasi error menunjukkan seberapa besar proporsi prediksi yang masih keliru dan menjadi indikator ruang perbaikan bagi model.

3.5.3 Weighted Mean Precision

Precision dihitung untuk setiap kelas *Risk* sebagai perbandingan jumlah data yang benar diprediksi pada kelas tersebut dengan seluruh data yang diprediksi sebagai kelas tersebut. Untuk kelas ke-i:

$$\text{Precision}_i = \frac{TP_i}{TP_i + FP_i} \tag{3}$$

Keterangan :
TP *True Positives* untuk kelas ke-i
FP *False Positives* untuk kelas ke-i

Dengan TP_i menyatakan jumlah *true positive* pada kelas ke-iii, dan FP_i menyatakan jumlah *false positive* pada kelas ke-iii. Nilai *precision* per kelas kemudian dirata-ratakan dengan bobot tertentu, sehingga menghasilkan *weighted mean precision* :

$$\text{Weighted Mean Precision} = \frac{\sum_{i=1}^{k} w_i \cdot \text{Precision}_i}{\sum_{i=1}^{k} w_i} \tag{4}$$

Keterangan :
Dengan k adalah jumlah kelas pada label *Risk* dan w_i adalah bobot untuk kelas ke-iii. Pada pengaturan RapidMiner yang digunakan, bobot dapat diatur sama (misalnya W_i = 1 untuk semua kelas), sehingga *weighted mean precision* ekuivalen dengan rata-rata *precision* per kelas.

3.5.4 Weighted Mean Recall

Recall (sensitivity) per kelas mengukur kemampuan model mengenali seluruh sampel aktual pada kelas tersebut. Untuk kelas ke-i:

$$Recall_i = \frac{TP_i}{TP_i + FN_i} \quad (5)$$

Keterangan :

TP_i (*True positives* kelas i) : jumlah data yang sebenarnya termasuk kelas i dan di prediksi sebagai kelas i.

FN_i (*False negative* kelas i) : jumlah data yang sebenarnya termasuk kelas i tetapi diprediksi sebagai kelas selain i.

i : indeks kelas misalnya *(low, intermediate, high).*

Nilai *recall* per kelas kemudian dirata-ratakan dengan bobot tertentu sehingga diperoleh *weighted mean recall*:

$$\text{Weighted Mean Recall} = \sum \text{i} = 1\text{kwi} \sum \text{i} = 1\text{kwi} \cdot \text{Recall}i \quad (6)$$

Keterangan :

k : Jumlah kelas atau kategori pada dataset.

w_i : Bobot untuk kelas ke-iii, biasanya proporsional terhadap jumlah data aktual pada kelas tersebut.

Ketika bobot w_i ditetapkan sama untuk setiap kelas, nilai *weighted mean recall* mencerminkan rata-rata kemampuan model dalam mengenali pasien pada masing-masing kelas *Risk* tanpa memihak pada kelas

Keempat metrik tersebut digunakan secara komplementer. Akurasi dan *classification error* memberikan gambaran performa global, sedangkan *weighted mean precision* dan *weighted mean recall* memberikan ringkasan performa per kelas secara lebih seimbang pada kasus multi-kelas.

IV. ANALISIS DAN HASIL PERANCANGAN

Penerapan algoritma C4.5 pada dataset *Differentiated Thyroid Cancer Recurrence* menghasilkan model klasifikasi tiga kelas risiko kekambuhan, yaitu *Low*, *Intermediate*, dan *High*. Evaluasi dilakukan menggunakan data uji yang tidak dilibatkan dalam proses pelatihan, dengan empat parameter utama, yakni *accuracy*, *classification error*, *weighted mean precision*, dan *weighted mean recall*. Keempat parameter tersebut ditampilkan pada empat keluaran RapidMiner yang masing-masing diawali dengan nilai ukuran kinerja dan diikuti oleh tabel *confusion matrix* seperti pada Tabel IV.1 sampai Tabel IV.4

Secara umum, Pada seluruh tabel memiliki isi yang sama. Baris menunjukkan kelas hasil prediksi model *(pred. Low, pred. Intermediate, pred. High)*, sedangkan kolom menunjukkan kelas sebenarnya *(true Low, true Intermediate, true High)*. Dari tabel terlihat bahwa terdapat 75 pasien dengan risiko rendah (64 + 11 + 0), 31 pasien dengan risiko intermediate (7 + 22 + 2), dan 10 pasien dengan risiko tinggi (2 + 0 + 8), sehingga total data uji berjumlah 116 pasien. Prediksi yang

tepat terjadi pada diagonal utama tabel (64 pasien *Low*, 22 pasien *Intermediate*, dan 8 pasien *High*), sedangkan nilai di luar diagonal menggambarkan kesalahan klasifikasi.

4.1 Hasil Akurasi Model

Tabel IV.1 menampilkan nilai *accuracy* sebesar 81,03% dengan *confusion matrix* yang sama. Akurasi dihitung sebagai perbandingan antara jumlah prediksi yang benar dan total prediksi. Dari tabel terlihat bahwa jumlah prediksi benar adalah 94 kasus (64 *Low* + 22 *Intermediate* + 8 *High*) dari total 116 kasus.

Tabel IV.1 Hasil Akurasi Model

Akurasi : 81.03%	**True Low**	**True Intermediate**	**True High**	**Class Precision**
Pred Low	64	7	2	87.67%
Pred Intermediate	11	22	0	66.67%
Pred High	0	2	8	80.00%
Class Recall	85.33%	70.79%	80.00%	

Nilai ini menunjukkan bahwa sekitar delapan dari sepuluh pasien pada data uji diklasifikasikan dengan *risk-level* yang sesuai. Untuk klasifikasi multi-kelas dengan tiga kategori risiko yang saling berdekatan, akurasi di atas 80% dapat dikategorikan cukup baik. Kombinasi akurasi 81,03%, *classification error* 18,97%, *weighted mean precision* 78,11%, dan *weighted mean recall* 78,77% menunjukkan bahwa model C4.5 mampu memetakan pola data klinis pasien dengan cukup andal, meskipun masih ada ruang perbaikan terutama dalam membedakan kelas *Low* dan *Intermediate*.

Secara ringkas, akurasi 81,03% yang diperoleh model C4.5 dalam penelitian ini masih berada di bawah akurasi di atas 90% yang dilaporkan beberapa penelitian lain pada prediksi kekambuhan DTC, khususnya studi yang memanfaatkan algoritma ensemble atau teknik pembelajaran mesin yang lebih kompleks dengan optimasi parameter dan strategi penyeimbangan kelas yang lebih agresif.

4.2 Hasil Klasifikasi Error

Berdasarkan hasil evaluasi pada data uji, diperoleh *classification error* sebesar 18,97%. Rincian performa klasifikasi antar kelas ditunjukkan melalui *confusion matrix*, yang menggambarkan jumlah data pada setiap kelas aktual *(True Low, True Intermediate, True High)* beserta hasil prediksi model *(Pred Low, Pred Intermediate, Pred High)*. Selain itu, ditampilkan pula nilai precision untuk setiap kelas prediksi serta *recall* untuk setiap kelas aktual sebagai indikator ketepatan dan kelengkapan prediksi pada kasus multi-kelas. Secara umum, model menunjukkan kinerja yang cukup baik pada kelas *Low* dan *High*, namun masih terdapat pertukaran prediksi pada kelas yang berdekatan.

Kesalahan terbesar terjadi antara kelas *Low* dan *Intermediate*, yaitu 11 pasien risiko rendah yang diprediksi sebagai *intermediate* serta 7 pasien risiko *intermediate* yang diprediksi sebagai rendah. Selain itu, terdapat 2 pasien risiko *intermediate* yang diprediksi sebagai risiko tinggi dan 2 pasien

risiko tinggi yang diprediksi sebagai risiko rendah. Pola ini menunjukkan bahwa batas pemisah antara kelas yang berdekatan *(Low–Intermediate dan Intermediate–High)* masih belum sepenuhnya tegas, sehingga sebagian pasien berpindah kelas akibat kemiripan karakteristik klinis.

Tabel IV.2 Hasil Klasifikasi Error

Klasifikasi Error : 18.97%	**True Low**	**True Intermediate**	**True High**	**Class Precision**
Pred Low	64	7	2	87.67%
Pred Intermediate	11	22	0	66.67%
Pred High	0	2	8	80.00%
Class Recall	85.33%	70.79%	80.00%	

Kesalahan terbesar terjadi antara kelas *Low* dan *Intermediate*, yaitu 11 pasien risiko rendah yang diprediksi sebagai *intermediate* serta 7 pasien risiko *intermediate* yang diprediksi sebagai rendah. Selain itu, terdapat 2 pasien risiko *intermediate* yang diprediksi sebagai risiko tinggi dan 2 pasien risiko tinggi yang diprediksi sebagai risiko rendah. Pola ini menunjukkan bahwa batas pemisah antara kelas yang berdekatan *(Low–Intermediate* dan *Intermediate–High)* masih belum sepenuhnya tegas, sehingga sebagian pasien berpindah kelas akibat kemiripan karakteristik klinis**.**

4.3 Hasil Weighted mean precision

Tabel IV.3 menampilkan nilai *weighted mean precision* sebesar 78,11% dengan bobot sama untuk masing-masing kelas. Pada kolom paling kanan terlihat nilai *class precision* untuk tiap kelas prediksi, yaitu 87,67% untuk pred. Low, 66,67% untuk *pred. Intermediate*, dan 80,00% untuk *pred. High*. Nilai-nilai ini menggambarkan tingkat ketepatan model ketika mengeluarkan suatu prediksi. Semakin tinggi nilai *precision* pada suatu kelas, semakin kecil kemungkinan model menghasilkan prediksi positif yang salah untuk kelas tersebut. Hal ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam membedakan antara kelas-kelas yang ada, meskipun terdapat perbedaan performa antar kelas. Perbedaan ini mengindikasikan bahwa model relatif lebih percaya diri dan akurat ketika memprediksi kelas *Low* dan *High* dibandingkan kelas *Intermediate*, yang umumnya memang lebih sulit dibedakan karena karakteristik klinisnya cenderung berada di antara dua ekstrem risiko tersebut. Dalam konteks penerapan klinis, pola ini perlu diperhatikan karena dapat memengaruhi interpretasi hasil model dan strategi tindak lanjut terhadap pasien dengan kategori risiko menengah. Temuan ini juga memberi ruang untuk pengembangan lebih lanjut, misalnya dengan penambahan fitur klinis yang lebih spesifik atau pengkombinasian dengan metode klasifikasi lain guna meningkatkan *precision* pada kelas *Intermediate* tanpa mengorbankan kinerja pada kelas lainnya.

Tabel IV.3 Hasil Weighted Mean Precision

Weighted Mean Precision: 78.11%, 1,1,1	**True Low**	**True Intermediate**	**True High**	**Class Precision**
Pred Low	64	7	2	87.67%
Pred Intermediate	11	22	0	66.67%
Pred High	0	2	8	80.00%
Class Recall	85.33%	70.79%	80.00%	

Sebagai contoh, *precision* 87,67% untuk prediksi *Low* berarti dari seluruh pasien yang diprediksi *Low*, sekitar 87,67% benar-benar berasal dari kelas *Low*, sementara sisanya merupakan pasien *Intermediate* atau *High* yang salah diklasifikasikan. *Precision* 66,67% untuk prediksi *Intermediate* menunjukkan bahwa sepertiga prediksi kelas ini masih salah (sebagian berasal dari kelas Low), sedangkan *precision* 80,00% untuk prediksi *High* menunjukkan bahwa delapan dari sepuluh pasien yang diprediksi *High* memang benar berisiko tinggi.

Weighted mean precision 78,11% merupakan rata-rata precision ketiga kelas dengan bobot yang sama. Nilai ini menunjukkan bahwa secara rata-rata, sekitar 78% pasien yang diprediksi berada pada suatu kelas risiko memang benar termasuk kelas tersebut. Dari sudut pandang klinis, angka ini menandakan bahwa model relatif jarang memberikan "alarm palsu", meskipun pada kelas *Intermediate* masih terdapat ketidaktepatan prediksi yang lebih tinggi dibanding dua kelas lainnya.

4.4 Hasil Weighted mean recall

Tabel IV.4 menampilkan nilai *weighted mean recall* sebesar 78,77%, yang diperoleh dari rata-rata *class recall* untuk tiap kelas, yaitu 85,33% untuk kelas *Low*, 70,97% untuk *Intermediate*, dan 80,00% untuk *High* (ditampilkan pada baris paling bawah tabel).
Recall menggambarkan kemampuan model dalam mengenali kasus yang benar-benar berasal dari suatu kelas. *Recall* 85,33% untuk kelas *Low* berarti 64 dari 75 pasien risiko rendah berhasil teridentifikasi dengan benar sebagai *Low*, sedangkan 11 sisanya "terlewat" dan justru diprediksi *Intermediate*. *Recall* 70,97% untuk kelas *Intermediate* menunjukkan bahwa sekitar 29% pasien *intermediate* diprediksi ke kelas lain (*Low* atau *High*), menandakan bahwa kelas ini merupakan kelas yang paling sulit dikenali. *Recall* 80,00% untuk kelas *High* berarti delapan dari sepuluh pasien yang benar-benar berisiko tinggi dapat diidentifikasi dengan tepat oleh model.

Tabel IV.4 Hasil Weighted Mean Recall

Weighted Mean Recall: 78.77% 1,1,1	**True Low**	**True Intermediate**	**True High**	**Class Precision**
Pred Low	64	7	2	87.67%
Pred Intermediate	11	22	0	66.67%
Pred High	0	2	8	80.00%
Class Recall	85.33%	70.97%	80.00%	

Nilai *weighted mean recall* 78,77% menunjukkan bahwa, secara rata-rata berbobot, model mampu mengenali sekitar tiga perempat pasien pada setiap kelas risiko. Dalam konteks klinis, hal ini penting khususnya untuk kelas *High*, karena kemampuan mengenali sebagian besar pasien berisiko tinggi akan membantu mengarahkan pemantauan dan intervensi yang lebih intensif kepada kelompok yang benar-benar membutuhkan.

4.5 Klasifikasi Risk-Level Kekambuhan Kanker Tiroid Terdiferensiasi Menggunakan Algoritma C4.5

Karena tujuan model adalah melakukan stratifikasi *risk-level (Low, Intermediate, High)*, interpretasi kinerja tidak hanya didasarkan pada akurasi global, tetapi juga pada kemampuan model membedakan tiap tingkat risiko. Pola kesalahan yang dominan pada kelas yang berdekatan (*Low–Intermediate* dan sebagian *Intermediate–High*) mengindikasikan adanya kemiripan karakteristik klinikopatologis di sekitar batas kelas, sehingga perpindahan kelas masih dapat terjadi pada pasien dengan profil yang *borderline*. Dari sisi penerapan, kesalahan yang mengarah pada penurunan kelas risiko (misalnya *High* diprediksi *Low*) perlu mendapat perhatian khusus karena berpotensi menyebabkan underestimation risiko, sedangkan kesalahan peningkatan kelas risiko (misalnya *Low* menjadi *Intermediate*) cenderung menghasilkan rekomendasi tindak lanjut yang lebih konservatif.

V. KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penerapan algoritma C4.5 pada dataset *Differentiated Thyroid Cancer Recurrence* dari *UCI Machine Learning Repository* mampu menghasilkan model klasifikasi yang cukup baik untuk memprediksi *risk-level* kekambuhan kanker tiroid terdiferensiasi. Model dibangun dengan memanfaatkan atribut klinis dan klinikopatologis yang tersedia pada dataset dan dilatih untuk membedakan tiga kategori risiko, yaitu risiko rendah *(Low)*, risiko menengah *(Intermediate)*, dan risiko tinggi *(High)*. Evaluasi menggunakan data uji yang tidak dilibatkan dalam proses pelatihan menunjukkan bahwa model C4.5 memiliki akurasi sebesar 81,03%, *classification error* 18,97%, *weighted mean precision*

78,11%, dan *weighted mean recall* 78,77%. Kombinasi nilai-nilai tersebut mengindikasikan bahwa secara keseluruhan model mampu memetakan pola klinis dengan tingkat ketepatan prediksi yang cukup tinggi untuk suatu problem klasifikasi multi-kelas.

Jika ditinjau lebih dalam, performa model tidak hanya baik secara *agregat*, tetapi juga relatif seimbang antara kemampuan mengenali kasus pada tiap kelas *(recall)* dan kemampuan mempertahankan ketepatan prediksi *(precision)*. Sebagian besar pasien dengan risiko rendah dan tinggi dapat diklasifikasikan dengan tepat, sementara kelas risiko menengah *(Intermediate)* menjadi kelas yang paling menantang dan paling sering terlibat dalam kesalahan klasifikasi, yang dapat dimaklumi karena kedekatan karakteristik klinis dengan kelas risiko rendah. Meskipun demikian, kemampuan model dalam mendeteksi sebagian besar pasien berisiko tinggi dan minimnya prediksi *"false high"* menunjukkan bahwa algoritma C4.5 cukup menjanjikan sebagai alat bantu awal untuk mengidentifikasi kelompok pasien yang membutuhkan pemantauan lebih intensif.

Temuan-temuan ini menegaskan bahwa algoritma C4.5 mampu memanfaatkan atribut yang tersedia untuk membangun pohon keputusan yang representatif terhadap proses kekambuhan kanker tiroid terdiferensiasi. Keunggulan utama tidak hanya terletak pada nilai akurasi yang kompetitif, tetapi juga pada interpretabilitas struktur pohon keputusan yang dihasilkan. Aturan-aturan *if–then* yang diturunkan dari cabang dan daun pohon dapat diterjemahkan menjadi logika klinis yang mudah dipahami, sehingga memudahkan klinisi menilai kesesuaian keputusan model dengan penalaran medis yang sudah ada. Dengan demikian, model yang dikembangkan memiliki potensi untuk diintegrasikan sebagai komponen dalam sistem pendukung keputusan klinis yang membantu proses stratifikasi risiko kekambuhan secara lebih objektif dan konsisten, tanpa menggantikan peran utama penilaian klinis dokter.

5.2 Saran

Meskipun hasil yang diperoleh menunjukkan kinerja model yang cukup baik, penelitian ini memiliki sejumlah keterbatasan yang perlu diperhatikan dan menjadi dasar bagi saran pengembangan selanjutnya. Sumber data yang digunakan sepenuhnya berasal dari satu dataset sekunder UCI dengan jumlah sampel 383 pasien dan karakteristik tertentu, sehingga generalisasi hasil ke populasi pasien di rumah sakit atau wilayah lain perlu dilakukan secara hati-hati. Selain itu, *classification error* yang masih mendekati 19% menunjukkan bahwa sekitar dua dari sepuluh pasien belum dapat diprediksi *risk-level* kekambuhannya secara tepat, terutama pada kelas risiko yang saling berdekatan. Kondisi ini mengisyaratkan perlunya upaya lebih lanjut untuk memperjelas batas pemisah antar kelas risiko.

Untuk penelitian di masa mendatang, disarankan agar digunakan dataset dengan ukuran dan keragaman yang lebih besar, termasuk data dari populasi lokal atau *multi-center*, sehingga variasi klinis di lapangan dapat terwakili dengan lebih baik. Penambahan atribut klinikopatologis yang lebih spesifik terhadap kekambuhan, seperti rincian terapi *radioiodine*, dinamika kadar *thyroglobulin*, maupun temuan imaging lanjutan, berpotensi meningkatkan kemampuan model dalam membedakan pasien berisiko tinggi dari mereka yang risikonya relatif rendah. Penerapan teknik prapemrosesan lanjutan seperti *feature selection*, penyeimbangan kelas melalui *oversampling* atau *undersampling*, serta pengaturan bobot kesalahan untuk kelas risiko tinggi *(cost-sensitive learning)* juga layak dipertimbangkan untuk menurunkan tingkat kesalahan

klasifikasi yang secara klinis merugikan.

Dari sisi metode, model C4.5 yang telah dikembangkan dapat dijadikan titik awal untuk studi komparatif dengan algoritma lain, baik algoritma tunggal seperti *logistic regression, support vector machine, random forest, dan XGBoost*, maupun pendekatan *ensemble* dan *stacking*. Perbandingan tersebut penting untuk mengetahui apakah terdapat model lain yang mampu memberikan peningkatan kinerja tanpa mengorbankan interpretabilitas yang dibutuhkan di domain klinis. Selain itu, sebelum diintegrasikan ke dalam sistem pendukung keputusan klinis, diperlukan validasi eksternal menggunakan data dari fasilitas kesehatan lain serta uji coba terbatas di lingkungan praktik nyata untuk menilai aspek kegunaan, penerimaan oleh klinisi, dan dampak potensialnya terhadap alur pengambilan keputusan dan tata laksana pasien.

DAFTAR PUSTAKA

[1]
[2] L. Tang, et al., "A machine learning-based model for predicting recurrence in intermediate- and high-risk differentiated thyroid cancer: insights from a retrospective single-center study of 2388 patients," Frontiers in Endocrinology, vol. 16, 2025, doi: 10.3389/fendo.2025.1552479.
[3] L. Giovanella, L. Milan, W. Roll, M. Weber, S. Schenke, M. Kreissl, et al., "Thyroglobulin measurement is the most powerful outcome predictor in differentiated thyroid cancer: a decision tree analysis in a European multicenter series," Clinical Chemistry and Laboratory Medicine (CCLM), 2024, doi: 10.1515/cclm-2024-0405.
[4] L. He, J. Xiang, and H. Zhang, "Rethinking the prognosis model of differentiated thyroid carcinoma," Frontiers in Endocrinology, vol. 15, 2024, doi: 10.3389/fendo.2024.1419125.
[5] H. Donmez, et al., "Predicting thyroid cancer recurrence using supervised CatBoost: A SHAP-based explainable AI approach," Medicine, vol. 104, 2025, doi: 10.1097/MD.0000000000042667.
[6] F. F. Atay, F. Yağın, C. Çolak, E. Elkiran, N. Mansuri, F. Ahmad, and L. P. Ardigò, "A hybrid machine learning model combining association rule mining and classification algorithms to predict differentiated thyroid cancer recurrence," Frontiers in Medicine, vol. 11, 2024, doi: 10.3389/fmed.2024.1461372.
[7] Y. Li, J.-Y. Tian, K. Jiang, Z. Wang, S. Gao, K. Wei, A. Yang, and Q. Li, "Risk factors and predictive model for recurrence in papillary thyroid carcinoma: a single-center retrospective cohort study based on 955 cases," Frontiers in Endocrinology, vol. 14, 2023, doi: 10.3389/fendo.2023.1268282.
[8] P. Ruchong, H. Tang, and X. Wang, "A five-gene prognostic nomogram predicting disease-free survival of differentiated thyroid cancer," Disease Markers, 2021, doi: 10.1155/2021/5510780.
[9] Y. M. Park and B.-J. Lee, "Machine learning-based prediction model for papillary thyroid carcinoma recurrence," Scientific Reports, vol. 10, 2020, doi: 10.21203/rs.3.rs-113105/v1.Clinical Chemistry and Laboratory Medicine (CCLM), 2024, doi: 10.1515/cclm-2024-0405.
[10] C. Zhang and Y. Ma, "An improved C4.5 decision tree algorithm for medical data classification," IEEE Access, vol. 9, pp. 137955–137965, 2021. doi: 10.1109/ACCESS.2021.3119134
[11] M. M. Rahman, M. S. Islam, and M. Ahmed, "Decision support systems in healthcare: Trends, techniques, and challenges," Artificial Intelligence in Medicine, vol. 129, p. 102320, 2022. doi: 10.1016/j.artmed.2022.102320
[12] J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques, 4th ed. Cambridge, MA: Morgan Kaufmann, 2022. doi: 10.1016/C2018-0-02071-0