Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 5 No. 1 (2026) pp: 1569-1577
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Penerapan Algoritma C4.5 dengan Teknik Pruning untuk Diagnosis Dini Penyakit Demam Berdarah

Yuda Apriansyah, Muhammad Siddik Hasibuan, Yusuf Ramadhan Nasution
Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sumatera Utara
yudaapriansyahyudaapriansyah6@gmail.com, muhammadsiddik@uinsu.ac.id, ramadhannst@uinsu.ac.id

Abstrak

Demam Berdarah Dengue (DBD) merupakan penyakit yang membutuhkan diagnosis dini secara tepat untuk mengurangi risiko komplikasi dan kematian. Penelitian ini bertujuan menerapkan algoritma Decision Tree C4.5 dengan teknik pruning untuk menghasilkan model klasifikasi diagnosis dini DBD yang akurat, sederhana, dan mampu meminimalkan risiko overfitting. Penelitian ini menggunakan dataset sekunder yang diperoleh dari Kaggle sebanyak 1.433 record. Melalui tahapan preprocessing yang meliputi pemeriksaan, pembersihan, dan seleksi data, diperoleh 1.418 data bersih yang terdiri atas 718 data Negatif DBD dan 700 data Positif DBD. Data tersebut memiliki enam atribut prediktor hematologi dan satu atribut target. Selanjutnya, dataset dibagi menjadi 80% data latih dan 20% data uji. Proses klasifikasi dilakukan menggunakan algoritma C4.5 dengan kriteria entropy, kemudian dioptimasi melalui teknik Cost Complexity Pruning menggunakan parameter ccp_alpha untuk menyederhanakan struktur pohon keputusan. Hasil penerapan pruning menunjukkan penyederhanaan struktur pohon dari delapan tingkat dan 27 daun menjadi model yang lebih ringkas dengan atribut Platelet Count (≤142.000) sebagai root node. Hasil evaluasi terhadap 284 data uji menunjukkan akurasi sebesar 97,89%. Model memperoleh Precision, Recall, dan F1-Score masing-masing sebesar 0,97, 0,99, dan 0,98 pada kelas Positif DBD, serta 0,99, 0,97, dan 0,98 pada kelas Negatif DBD. Dengan demikian, penerapan teknik pruning menghasilkan model klasifikasi yang memiliki kinerja baik dan berpotensi mendukung proses identifikasi awal DBD.

Kata kunci: Demam Berdarah Dengue, Diagnosis Dini, Machine Learning, Algoritma C4.5, Cost Complexity Pruning

Abstract

Dengue Fever (DBD) is a disease that requires early and accurate diagnosis to reduce the risk of complications and death. This study aims to apply the C4.5 Decision Tree algorithm with pruning techniques to create an early DBD diagnosis classification model that is accurate, simple, and able to minimize the risk of overfitting. The study uses a secondary dataset from Kaggle consisting of 1,433 records. Through preprocessing steps that include checking, cleaning, and selecting data, 1,418 clean records were obtained, consisting of 718 Negative DBD records and 700 Positive DBD records. These data have six hematology predictor attributes and one target attribute. Next, the dataset was split into 80% training data and 20% testing data. The classification process was carried out using the C4.5 algorithm with entropy criteria, and then optimized using techniques Cost Complexity Pruning uses the parameter ccp_alpha to simplify the structure of decision trees. The application of pruning showed a simplification of the tree structure from eight levels and 27 leaves to a more compact model with Platelet Count (≤142,000) as the root node. Evaluation on 284 test data showed an accuracy of 97.89%. The model achieved Precision, Recall, and F1-Score of 0.97, 0.99, and 0.98 for the Positive DBD class, and 0.99, 0.97, and 0.98 for the Negative DBD class. Therefore, applying the pruning technique produced a classification model with good performance and has the potential to support the early identification process of DBD.

Keywords: Dengue Fever, Early Diagnosis, Machine Learning, C4.5 Algorithm, Cost Complexity Pruning

1. Pendahuluan

Demam Berdarah *Dengue* (DBD) merupakan salah satu penyakit infeksi menular yang banyak dijumpai di wilayah dengan iklim tropis maupun subtropis dan disebabkan oleh virus *dengue* melalui vektor nyamuk *Aedes aegypti* (Dewi, 2023). Hingga saat ini, DBD masih menjadi permasalahan kesehatan masyarakat yang sangat serius karena bersifat akut dan dapat berkembang dengan cepat menuju kondisi kritis apabila tidak ditangani secara tepat (Ramadhani et al., 2023). Di Indonesia, Kementerian Kesehatan Republik Indonesia mencatat sebanyak 88.593 kasus DBD dengan 621 kasus kematian hingga minggu ke-17 tahun 2024 (Kadek et al., 2025). Data tersebut menegaskan pentingnya perhatian khusus, terutama dalam upaya diagnosis dini untuk menekan angka kesakitan dan kematian penderita.

Salah satu tantangan utama dalam penanganan DBD di fasilitas kesehatan adalah kesulitan dalam melakukan diagnosis dini. Pada fase awal, gejala klinis DBD seperti demam di atas 38°C sering kali menyerupai demam biasa, demam tifoid, atau infeksi virus ringan lainnya (Sofiana et al., 2024). Selain demam mendadak, penderita sering mengeluhkan sakit kepala, mual, muntah, hingga timbulnya ruam atau bintik merah pada kulit (Ahmad et al., 2023). Kemiripan gejala awal ini berpotensi memicu keterlambatan penanganan medis yang berujung pada komplikasi fatal.

Machine learning dapat dimanfaatkan dalam bidang kesehatan untuk membantu proses identifikasi dan diagnosis penyakit berdasarkan data medis. Pemanfaatan kecerdasan buatan dalam bidang kesehatan memungkinkan proses analisis dilakukan secara terkomputerisasi sehingga dapat menjadi alternatif pendukung dalam proses diagnosis. Penelitian Hasibuan dan Suhardi menunjukkan bahwa model berbasis *machine learning* dapat dikembangkan untuk membantu mendeteksi penyakit melalui data medis, dengan tahapan yang meliputi *preprocessing*, pembentukan model, pelatihan, serta evaluasi menggunakan sejumlah metrik kinerja. Pendekatan tersebut menunjukkan bahwa teknologi kecerdasan buatan memiliki potensi untuk mendukung proses pengambilan keputusan dalam bidang Kesehatan(Hasibuan & Suhardi, 2025).

Seiring perkembangan teknologi informasi, pemanfaatan teknik *data mining* dan *machine learning* di bidang kesehatan dapat membantu pengambilan keputusan medis secara lebih presisi (Kalsum et al., 2025). Algoritma *Decision Tree* C4.5 merupakan salah satu metode klasifikasi populer yang terbukti mampu menghasilkan akurasi tinggi serta membentuk aturan keputusan berbasis pohon yang mudah diinterpretasikan oleh tenaga medis (Supandi & Zaehol Fatah, 2025). Meskipun demikian, model C4.5 yang dibentuk tanpa pemangkasan (*unpruned*) cenderung menghasilkan struktur pohon yang terlalu kompleks, rentan terhadap *noise*, dan berisiko tinggi mengalami *overfitting* sehingga kinerjanya menurun saat menghadapi data uji baru (Halabaku & Bytyçi, 2024).

Untuk mengatasi masalah tersebut, teknik *pruning* (pemangkasan) sangat diperlukan untuk memotong cabang-cabang yang tidak signifikan tanpa mengorbankan informasi penting (Syahputri & Hasibuan, 2024). Penerapan *pruning* mampu menyederhanakan pohon keputusan, meningkatkan stabilitas, serta memperkuat kemampuan generalisasi model. Oleh karena itu, penelitian ini bertujuan mengombinasikan algoritma C4.5 dengan teknik *Cost Complexity Pruning* (ccp_alpha) guna menghasilkan model klasifikasi diagnosis dini DBD yang sederhana, akurat, dan bebas dari *overfitting*.

2. Metode Penelitian

2.1. Tahapan Penelitian

Penelitian ini disusun secara terstruktur melalui enam tahapan utama yaitu,Perencanaan (*Planning*), Pengumpulan Data (*Data Collection*), Analisis Kebutuhan (*Requirements Analysis*), Perancangan (*Design & Modeling*), Pengujian (*Testing & Evaluation*), serta Penerapan dan Penarikan Kesimpulan (*Implementation & Conclusion*).

Pada tahap *planning*, dilakukan studi literatur terkait tantangan diagnosis dini DBD dan identifikasi kelemahan algoritma C4.5 akibat *overfitting*. Selanjutnya, pada tahap *data collection*, *dataset* sekunder parameter hematologi diunduh dari repositori *Kaggle*. Pada tahap *requirements analysis*, dilakukan *data preprocessing* termasuk penanganan nilai hilang via imputasi median dan eliminasi duplikasi hingga menghasilkan 1.418 data bersih yang kemudian dibagi menjadi 80% data latih dan 20% data uji.

2.2. Teknik Pengumpulan Data

Data dalam penelitian ini diperoleh melalui observasi tidak langsung dengan menggunakan *dataset* sekunder dari repositori *Kaggle*. *Dataset* tersebut berisi atribut terkait kondisi pasien yang digunakan sebagai variabel prediktor serta satu atribut target yang menunjukkan hasil diagnosis DBD. Data disimpan dalam format CSV sehingga mudah diolah menggunakan Python. Pemilihan *dataset* ini disesuaikan dengan tujuan penelitian, yaitu membangun model klasifikasi menggunakan algoritma C4.5 untuk mendukung diagnosis dini DBD. *Dataset* ini terdiri dari kurang lebih 1433 baris dan 7 kolom atribut. Keterangan *dataset* dapat dilihat pada tabel 1 dibawah ini

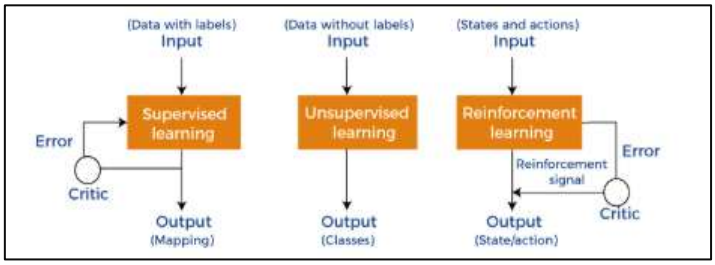
DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 1. Atribut Dataset

Nama Atribut	Keterangan	Tipe Data
Haemoglobin	Kadar Haemoglobin darah pasien	Numerik
WBC Count	Jumlah sel darah putih (Leukosit)	Numerik
Differential Count	Kondisi differensial sel darah 0/1	Biner/kategorikal
RBC Panel	Hasil pemeriksaan sel darah merah (0/1)	Biner/kategorikal
Platelet Count	Jumlah trombosit dalam darah	Numerik
PDW	Patelet distribution with	Numerik
Final Output	Status diagnosis (0 Negatif, 1 positif)	Biner

2.3. Machine Learning

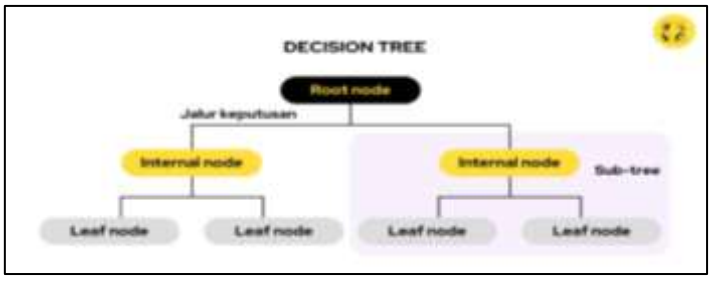
Machine Learning (ML) merupakan salah satu komponen utama dalam bidang *Artificial Intelligence* (AI) yang memanfaatkan metode statistik untuk memungkinkan komputer belajar dan mengambil keputusan secara mandiri tanpa harus diprogram secara eksplisit (Alnuaimi & Albaldawi, 2024). Secara umum, *Machine Learning* dibagi menjadi tiga kategori utama, yaitu *Supervised Learning*, *Unsupervised Learning*, dan *Reinforcement Learning* (Nurhakim et al., 2024).



Gambar 1. Kategori Machine Learning

2.4. Decision Tree

Decision Tree merupakan metode klasifikasi yang membangun model dalam bentuk struktur pohon, yang terdiri dari simpul (*node*), cabang (*branch*), dan daun (*leaf*). Simpul mewakili atribut, cabang menunjukkan hasil pengujian atribut, dan daun menyatakan kelas atau keputusan akhir (Siahaan & Muliono, 2025). Algoritma C4.5 merupakan pengembangan dari algoritma ID3 dan termasuk salah satu algoritma *Decision Tree* yang paling banyak digunakan. C4.5 memiliki keunggulan dalam menangani atribut bertipe kategorik maupun numerik serta mampu menghasilkan pohon keputusan yang mudah dipahami dan diinterpretasikan. Proses pembentukan pohon pada algoritma C4.5 dilakukan dengan menghitung nilai *information gain* untuk setiap atribut, di mana atribut dengan nilai *gain* tertinggi dipilih sebagai simpul akar. Keunggulan ini menjadikan algoritma C4.5 efektif dan efisien untuk digunakan dalam proses klasifikasi, termasuk pada kasus diagnosis penyakit (Junaidi et al., 2024).



Gambar 2. Bentuk Decision Tree

2.5. Tahapan-Tahapan C4.5

DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Menurut (Istiqomah & Sofica, 2025) tahap pertama dalam algoritma C4.5 adalah menyiapkan data latih, yaitu kumpulan data historis yang telah diklasifikasikan ke dalam kelas tertentu sesuai dengan kebutuhan analisis. Selanjutnya, ditentukan akar (*root*) pohon keputusan berdasarkan atribut yang memiliki nilai *gain* paling tinggi dibandingkan atribut lainnya. Sebelum menghitung nilai *gain*, terlebih dahulu dilakukan perhitungan *entropy* pada setiap atribut menggunakan rumus yang telah ditentukan.

1. Rumus *Entropy* :

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

2. Rumus *Information Gain*

$$Gain(S, A) = Entropy(S) - \sum_{\text{v}\in\text{Values(A)}} \frac{|\ Sv\ |}{|\ S\ |}\ Entropy(Sv)$$

3. Rumus *SplitInformation*

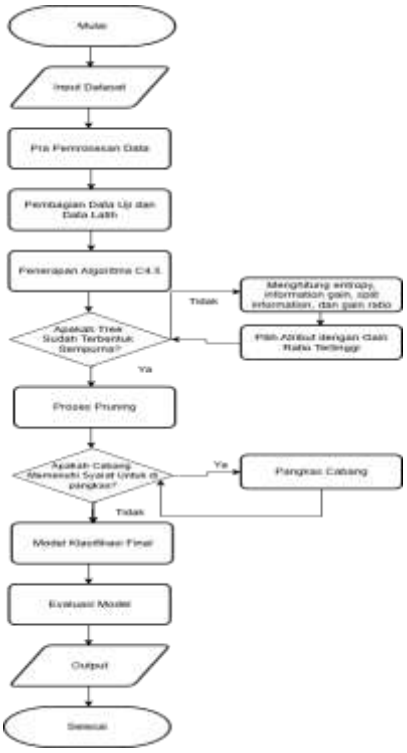
$$SplitInfo\ (S, A) = -\sum_{i=1}^{n} \frac{|\ Sv\ |}{|\ S\ |} \log 2\ \left(\frac{|\ Sv\ |}{|\ S\ |}\right)$$

4. Rumus *Gain Ratio*

$$GainRatio\ (S, A) = \frac{Gain\ (S, A)}{SplitInfo\ A(S)}$$

2.6. Metode Perancangan Model

Perancangan ini mencakup penyusunan alur logika algoritma yang divisualisasikan melalui *flowchart*, mulai dari *input* data, *preprocessing*, hingga pembentukan model. Data yang telah diolah digunakan untuk membangun model klasifikasi C4.5 berdasarkan perhitungan *entropy*, *information gain*, *split information*, dan *gain ratio* untuk menentukan atribut terbaik. Selain itu, teknik *pruning* diterapkan guna menyederhanakan struktur pohon keputusan dan mencegah *overfitting*.



Gambar 3. Perancangan Pembentukan Model C4.5

DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

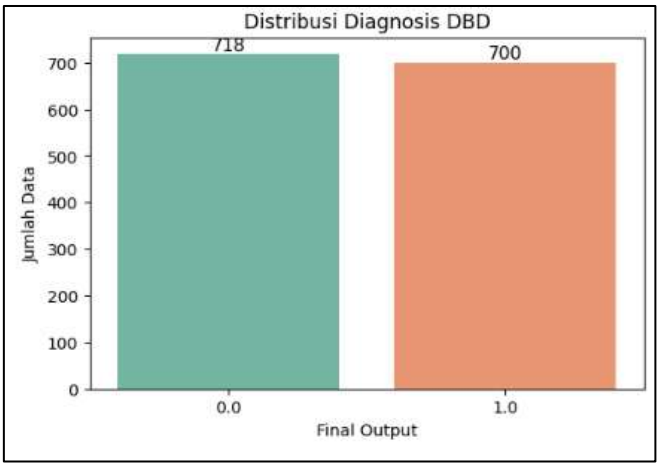
3. Hasil dan Pembahasan

Rangkaian hasil penelitian berdasarkan urutan/susunan logis untuk membentuk sebuah cerita. Isinya menunjukan fakta/data. Dapat menggunakan Tabel dan Angka tetapi tidak menguraikan secara berulang terhadap data yang sama dalam gambar, tabel dan teks. Untuk lebih memperjelas uraian, dapat mengunakan sub judul.

Pembahasan adalah penjelasan dasar, hubungan dan generalisasi yang ditunjukkan oleh hasil. Uraiannya menjawab pertanyaan penelitian. Jika ada hasil yang meragukan maka tampilkan secara objektif.

3.1. Pra-Pemrosesan Data

Pemeriksaan awal (*missing values*) menemukan sejumlah data kosong pada atribut *WBC Count* (30 data), *Platelet Count* (28 data), *PDW* (28 data), dan *Final Output* (15 data). Penanganan dilakukan dengan mengimputasi atribut numerik menggunakan nilai median (untuk mengantisipasi *outlier*) serta menghapus baris target yang kosong. Tahap pra-pemrosesan ini menghasilkan 1.418 data bersih yang terdistribusi secara seimbang 718 data kelas Negatif DBD (0) dan 700 data kelas Positif DBD (1). Selanjutnya, dataset dibagi dengan skema *Stratified Train-Test Split* berasio 80% data latih (1.134 sampel) dan 20% data uji (284 sampel).



Gambar 4. Distribusi Final Output

3.2. Implementasi Teknik Pruning

Setelah pohon keputusan terbentuk, dilakukan penerapan teknik *pruning* untuk menyederhanakan struktur pohon dengan menghilangkan cabang yang kurang berpengaruh terhadap proses klasifikasi. Pada penelitian ini, *pruning* diterapkan menggunakan parameter *Cost Complexity Pruning (ccp_alpha)* sehingga menghasilkan model yang lebih sederhana, mengurangi risiko *overfitting*, dan tetap mempertahankan performa klasifikasi dalam diagnosis dini Demam Berdarah *Dengue* (DBD), seperti gambar dibawah ini.

```
clf_unpruned = DecisionTreeClassifier(criterion='entropy', random_state=42)
clf_unpruned.fit(X_train, y_train)

print(f"Tree Unpruned berhasil dibentuk.")
print(f"Kedalaman Tree awal: {clf_unpruned.get_depth()}")
print(f"Jumlah total node awal: {clf_unpruned.get_n_leaves()}")
```

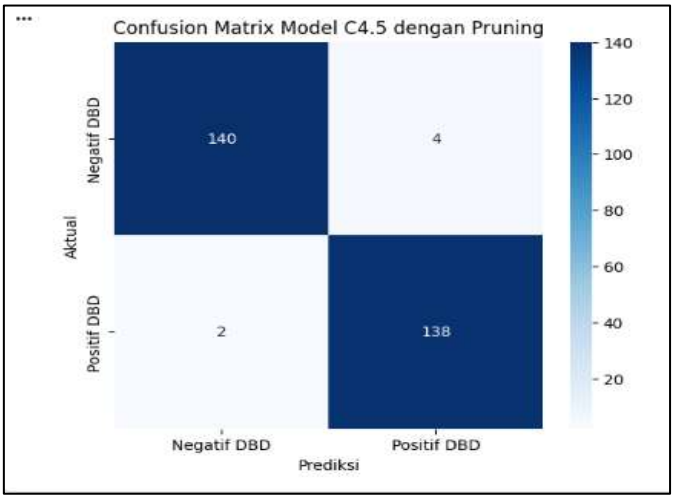
```
Tree Unpruned berhasil dibentuk.
Kedalaman Tree awal: 8
Jumlah total node awal: 27
```

Gambar 5. Implementasi Teknik Pruning

DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3.3. Confusion Matrix

Confusion Matrix ini menunjukkan hasil prediksi model C4.5 dengan *pruning*. Model berhasil mengklasifikasikan 140 data negatif DBD dan 138 data positif DBD dengan benar. Kesalahan prediksi hanya terjadi pada 6 data, yaitu 4 *False Positive* dan 2 *False Negative*. Hal ini menunjukkan bahwa model memiliki performa klasifikasi yang sangat baik.



Gambar 6. Confusion Matrix C4.5

3.4. Evaluasi Model

Evaluasi model klasifikasi C4.5 berbasis *pruning* untuk diagnosis DBD diukur menggunakan empat metrik utama: *Accuracy* untuk mengukur ketepatan prediksi keseluruhan, *Precision* untuk menilai akurasi prediksi positif DBD, *Recall* untuk mengukur kemampuan mendeteksi seluruh pasien terindikasi DBD, serta *F1-Score* untuk menilai keseimbangan keduanya. Keempat metrik ini menjadi dasar utama dalam menilai keandalan dan akurasi model sebagai sistem pendukung keputusan medis.

```
=== EVALUASI MODEL FINAL ===
Akurasi Model: 97.89%

Confusion Matrix:
[[140   4]
 [  2 138]]

Classification Report:
              precision    recall  f1-score   support

 Negatif DBD       0.99      0.97      0.98       144
 Positif DBD       0.97      0.99      0.98       140

    accuracy                           0.98       284
   macro avg       0.98      0.98      0.98       284
weighted avg       0.98      0.98      0.98       284
```

Gambar 7. Hasil Evaluasi Model

Gambar tersebut menampilkan hasil evaluasi final dari sebuah model klasifikasi yang memprediksi status "Negatif DBD" dan "Positif DBD". Model ini mencatatkan tingkat akurasi yang sangat tinggi, yaitu sebesar 97,89%. Berdasarkan *Confusion Matrix*, dari total 284 data uji, model berhasil mengklasifikasikan 140 sampel True Negative dan 138 sampel True Positive, dengan kesalahan prediksi yang sangat minim, yakni hanya 4 False Positive dan 2 False Negative. Laporan klasifikasi (*Classification Report*) menunjukkan performa yang seimbang dan konsisten di kedua kelas, dengan nilai *precision*, *recall*, dan *f1-score* rata-rata berada pada kisaran 0,97 hingga

DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

0,99, yang menandakan bahwa model memiliki kemampuan sangat baik dalam memprediksi maupun mendeteksi kasus DBD dengan akurat. Berdasarkan Gambar 7 yang menyajikan hasil evaluasi model, berikut adalah rincian perhitungan manual untuk masing-masing metrik evaluasi:

$$Accuracy = \frac{138 + 140}{284} x\ 100\% = \frac{278}{284}\ x\ 100\% = 97{,}89\%$$

$$Precision_{Positif} = \frac{TP}{TP + FP} = \frac{138}{138 + 4} = \frac{138}{142} = 0{,}97\%$$

$$Precision_{Negatif} = \frac{TN}{TN + FN} = \frac{140}{140 + 2} = \frac{140}{142} = 0{,}99\%$$

$$Recall_{Positif} = \frac{TP}{TP + FN} = \frac{138}{138 + 2} = \frac{138}{140} = 0{,}99\%$$

$$Recall_{Negatif} = \frac{TN}{TN + FP} = \frac{140}{140 + 4} = \frac{140}{144} = 0{,}97\%$$

$$F1 - Score_{Positif} = 2x\ \frac{0{,}9718\ x\ 0{,}9857}{0{,}9718 + 0{,}9857} = 2x\frac{0{,}9579}{1{,}9575} = 0{,}98\%$$

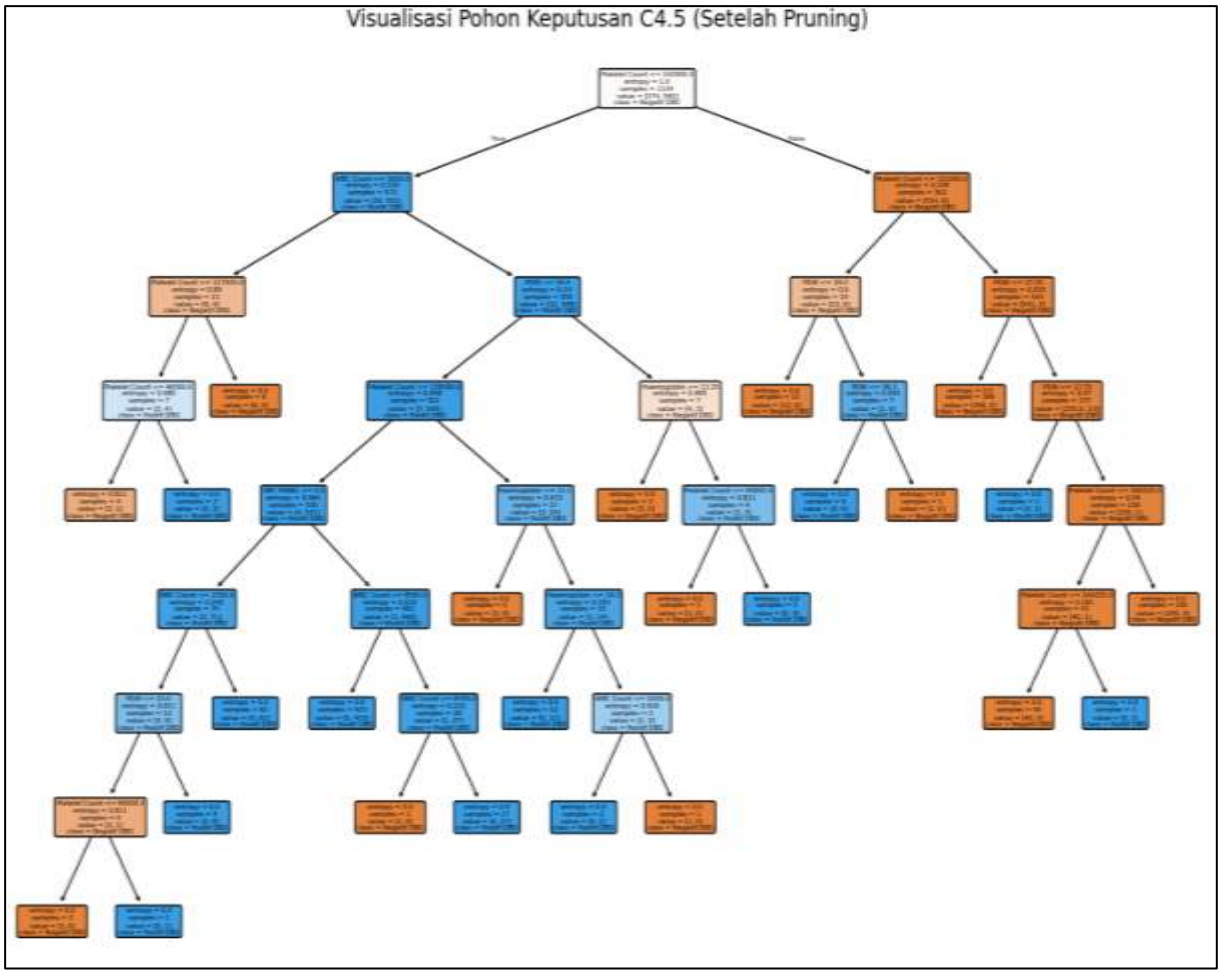
$$F1 - Score_{Negatif} = 2x\ \frac{0{,}9859\ x\ 0{,}9722}{0{,}9859 + 0{,}9722} = 2x\frac{0{,}9585}{1{,}9581} = 0{,}98\%$$

Evaluasi model klasifikasi C4.5 berbasis *pruning* untuk diagnosis DBD menggunakan empat metrik utama: *Accuracy* untuk mengukur ketepatan prediksi keseluruhan, *Precision* untuk menilai akurasi prediksi terindikasi DBD, *Recall* untuk mengukur kemampuan mendeteksi seluruh kasus positif, serta *F1-Score* untuk menilai keseimbangan keduanya. Keempat metrik ini menjadi dasar dalam mengevaluasi tingkat akurasi dan keandalan model sebagai sistem pendukung keputusan medis.

3.5. Visualisasi Pohon Keputusan

Adapun hasil visualisasi pohon keputusan final yang telah dioptimasi dengan teknik *pruning* disajikan pada Gambar 4.10 berikut:

DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 8. Visualisasi Pohon Keputusan C4.5

4. Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, penerapan algoritma *Decision Tree* C4.5 dengan teknik *Cost Complexity Pruning* (*ccp_alpha*) mampu menghasilkan model klasifikasi yang efektif untuk mendukung diagnosis dini penyakit Demam Berdarah *Dengue* (DBD). Tahap *preprocessing* menghasilkan 1.418 data bersih yang terdiri dari 718 data Negatif DBD dan 700 data Positif DBD, kemudian dibagi menjadi 80% data latih dan 20% data uji. Penerapan teknik *pruning* berhasil menyederhanakan struktur pohon keputusan sehingga dapat mengurangi risiko *overfitting* dan mempertahankan performa klasifikasi. Hasil pengujian terhadap 284 data uji menunjukkan bahwa model memperoleh akurasi sebesar 97,89%, dengan 140 data Negatif DBD dan 138 data Positif DBD berhasil diklasifikasikan dengan benar. Model juga menghasilkan 4 *False Positive* dan hanya 2 *False Negative.* Nilai *Precision*, *Recall*, dan *F1-Score* berada pada kisaran 0,97 hingga 0,99, dengan *Recall* kelas *Positif* DBD sebesar 0,99. Hasil tersebut menunjukkan bahwa model C4.5 berbasis *pruning* memiliki kemampuan yang sangat baik dalam mengidentifikasi kasus DBD serta dapat digunakan sebagai salah satu pendukung keputusan dalam proses diagnosis dini.

Referensi

1. Ahmad, Z. F., Salsabila Mongilong, N., Kadir, L., Indah Nurdin, St. S., & Rahmawaty Moo, D. (2023). Perbandingan Manifestasi Klinis Penderita Demam Berdarah. *Indonesian Journal of Pharmaceutical Education*, *3*(1). https://doi.org/10.37311/ijpe.v3i1.19231
2. Alnuaimi, A. F. A. H., & Albaldawi, T. H. K. (2024). An overview of machine learning classification techniques. *BIO Web of Conferences*, *97*. https://doi.org/10.1051/bioconf/20249700133
3. Dewi, A. M. N. P. T. N. G. L. (2023). *Prediksi Daerah Rawan Penyakit Demam Berdarah Menggunakan Algoritma C4.5*. *Volume 2*(2).
4. Halabaku, E., & Bytyçi, E. (2024). Overfitting in Machine Learning: A Comparative Analysis of Decision Trees and Random Forests. *Intelligent Automation and Soft Computing*, *39*(6), 987–1006. https://doi.org/10.32604/iasc.2024.059429

DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

5. Hasibuan & Suhardi. (2025). *Higher Education and Health Breakthroughs: Residual Network for Pneumonia Detection from Chest X-ray Images*. *15*(5).
6. Istiqomah, K., & Sofica, V. (2025). Penerapan Data Mining Menggunakan Algoritma Decision Tree Untuk Menganalisis Penggunaan Media Sosial Dengan Konsentrasi Belajar Mahasiswa. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(4), 53–67. https://doi.org/10.31004/riggs.v4i4.3228
7. Junaidi, S., Beno, I. S., Farkhan, M., Supartha, I. K. D. G., Pasaribu, A. A., Kmurawak, R. M., Supiyanto, Sroyer, A. M., Reba, F., Fitriyanto, R., Syafiqoh, U., & Rizal, A. A. (2024). *Buku Ajar Machine Learning* (Efitra, Ed.). SonpediaPublishingIndonesia.
8. Kadek, I., Prayoga, Y., Made, I., Sunarya, G., & Suputra, P. H. (2025). *Klasifikasi Penyakit Demam Berdarah Menggunakan Algoritma Stacking Ensemble Learning*. *Vol. 8*(2). https://doi.org/10.31598
9. Kalsum, D. N., Alkadri, S. P. A., & Istikoma, I. (2025). Implementation of Data Mining to Predict Dengue Prone Areas Using C4.5 Algorithm (Case Study: Sanggau Regency). *Smart Comp: Jurnalnya Orang Pintar Komputer*, *14*(1). https://doi.org/10.30591/smartcomp.v14i1.7540
10. Nurhakim, B., Dikananda, A. R., Hamonangan, R., & Prasetyo, R. A. (2024). *Data Cerdas Di Dunia Pendidikan Penerapan Machine Learning Untuk Kepuasan Mahasiswa* (A. Sofatunisa, Ed.). MegaPressNusantara.
11. Ramadhani, F., Satria, A., & Sari, I. P. (2023). Implementasi Metode Fuzzy K-Nearest Neighbor dalam Klasifikasi Penyakit Demam Berdarah. *Hello World Jurnal Ilmu Komputer*, *2*(2), 58–62. https://doi.org/10.56211/helloworld.v2i2.253
12. Siahaan, R. F., & Muliono, R. (2025). JITE (Journal of Informatics and Telecommunication Engineering) Coffee Quality Classification Based on Customer Reviews Using C4.5 Algorithm. *JITE*, *8*. https://doi.org/10.31289/jite.v8i3Spc.14427
13. Sofiana, E., Herin, L., & Prasetyaningru, P. T. (2024). Sistem Pakar Diagnosa Penyakit Demam Berdarah Dengue Dan Tifoid Menggunakan Metode Certanty Factor Berbasis Web. In *Journal of Computer and Information Systems Ampera* (Vol. 5, Number 3). https://journal-computing.org/index.php/journal-cisa/index
14. Supandi, & Zaehol Fatah. (2025). Penerapan Data Mining Untuk Prediksi Diagnosis Demam Berdarah Dengan Algoritma Decision Tree C4.5. *Journal Of Computer Science And Technology (JOCSTEC)*, *3*(2), 74–82. https://doi.org/10.59435/jocstec.v3i2.437
15. Syahputri, C. N., & Hasibuan, M. S. (2024). Optimasi Klasifikasi Decision Tree Dengan Teknik Pruning Untuk Mengurangi Overfitting. *JSiI (Jurnal Sistem Informasi)*, *11*(2), 87–96. https://doi.org/10.30656/jsii.v11i2.9161

DOI: https://doi.org/10.69693/ijmst.v5i1.14735
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)