

Klasifikasi Data Penyakit Demam Berdarah Dengue (DBD) Melalui Algoritma Decision Tree Dengan RapidMiner

¹ Zaehol Fatah, ² Fitra Lia Shofiah

¹ Sistem Informasi, Universitas Ibrahimy, Sukorejo
Email: ¹zaeholfatah@gmail.com, ²fifigemoy174@gmail.com

Abstrak

Demam Berdarah Dengue (DBD) merupakan permasalahan kesehatan masyarakat yang hingga kini terus mendapat perhatian serius di Indonesia. Proses diagnosis untuk DBD biasanya dilakukan melalui pemeriksaan laboratorium dan evaluasi klinis, Akan tetapi, pendekatan konvensional ini cenderung membutuhkan waktu yang relatif panjang serta sangat bergantung pada analisis profesional kesehatan. Oleh karena itu, diperlukan suatu pendekatan berbasis teknologi yang mampu mempercepat sekaligus mempermudah proses klasifikasi secara lebih akurat. Penelitian ini bertujuan untuk mengidentifikasi tingkat keparahan penyakit Demam Berdarah Dengue (DBD) dengan menerapkan algoritma Decision Tree yang dijalankan melalui aplikasi RapidMiner. Data yang digunakan merupakan algoritma sekunder yang di ambil dari internet, berisi 200 catatan pasien dengan delapan atribut medis, yaitu suhu tubuh, jumlah trombosit, jumlah leukosit, tekanan darah, ruam kulit, sakit kepala, nyeri otot, dan muntah. Langkah-langkah dalam penelitian mencakup pengumpulan data, preprocessing, penerapan algoritma Decision Tree, serta evaluasi model dengan memanfaatkan confusion matrix dengan metode cross-validation. Temuan dari studi ini memperlihatkan bahwa model Decision Tree mencapai tingkat ketepatan akurasi mencapai 85%, precision 84%, recall 83%, dan F1-score 83%, dengan trombosit dan leukosit sebagai variabel yang paling dominan dalam proses klasifikasi. Dengan merujuk pada hasil yang diperoleh, studi ini menyimpulkan bahwasanya algoritma Decision Tree cukup efektif dalam mengklasifikasikan tingkat keparahan penyakit DBD dan dapat di gunakan sebagai dasar untuk mengembangkan system pendukung keputusan di bidang kesehatan.

Kata Kunci: Demam Berdarah Dengue, Data Mining, Decision Tree, RapidMiner, Klasifikasi

Abstract

Dengue Hemorrhagic Fever (DHF) is a public health problem that continues to receive serious attention in Indonesia. The diagnosis process for DHF is usually carried out through laboratory examinations and clinical evaluations. However, this conventional approach tends to take a relatively long time and relies heavily on the analysis of health professionals. Therefore, a technology-based approach is needed that can accelerate and simplify the classification process more accurately. This study aims to identify the severity of Dengue Hemorrhagic Fever (DHF) by applying the Decision Tree algorithm run through the RapidMiner application. The data used is a secondary algorithm taken from the internet, containing 200 patient records with eight medical attributes, namely body temperature, platelet count, leukocyte count, blood pressure, skin rash, headache, muscle pain, and vomiting. The steps in the study include data collection, preprocessing, application of the Decision Tree algorithm, and model evaluation using the confusion matrix with the cross-

validation method. The findings of this study show that the Decision Tree model achieved an accuracy rate of 85%, precision of 84%, recall of 83%, and an F1-score of 83%, with platelets and leukocytes as the most dominant variables in the classification process. Based on these results, this study concludes that the Decision Tree algorithm is quite effective in classifying the severity of dengue fever and can be used as a basis for developing decision support systems in the healthcare sector.

Keywords: Dengue Hemorrhagic Fever, Data Mining, Decision Tree, RapidMiner, Classification

PENDAHULUAN

Kasus Demam Berdarah Dengue (DBD) masih menjadi permasalahan kesehatan masyarakat yang serius di berbagai wilayah dunia, termasuk Indonesia. Penyakit ini disebabkan oleh virus dengue yang ditularkan melalui gigitan nyamuk *Aedes aegypti* dan *Aedes albopictus*, serta memiliki tingkat penyebaran yang fluktuatif setiap tahunnya. Berbagai penelitian telah dilakukan untuk memanfaatkan teknologi data mining dalam mendukung analisis dan prediksi penyakit ini. Menurut Rocha dan Giesbrecht (2022), algoritma pembelajaran mesin dapat digunakan untuk menilai risiko penyebaran DBD di Brasil dengan tingkat akurasi yang tinggi. Assauqi dan Fatah (2024) mengembangkan model Decision Tree untuk mengidentifikasi risiko DBD berdasarkan gejala klinis dan hasil laboratorium. Sementara itu, Hawa, Prasetiyowati, dan Sibaroni (2024) memanfaatkan ekspansi fitur berbasis waktu guna memprediksi pola penyebaran demam berdarah dengan hasil yang menjanjikan.

Penelitian lain oleh Selamat dan Hakim (2024) membandingkan algoritma klasifikasi Naïve Bayes untuk membedakan antara DBD dan tifoid berdasarkan parameter hematologi, sedangkan Hamdani, Hatta, Puspitasari, dan Septiarini (2022) menerapkan Support Vector Machine dengan teknik cross-validation guna meningkatkan ketepatan hasil

klasifikasi. Studi lain juga menegaskan efektivitas pendekatan berbasis Decision Tree dalam bidang kesehatan untuk diagnosis penyakit menular (Rohani et al., 2023; Yuliani et al., 2024). Oleh karena itu, penggunaan algoritma Decision Tree pada penelitian ini diharapkan mampu memberikan kontribusi nyata dalam proses klasifikasi data DBD dengan memanfaatkan perangkat lunak RapidMiner sebagai alat bantu analisis data.

Selain itu, pemanfaatan teknik data mining dalam bidang kesehatan terus berkembang seiring dengan meningkatnya kebutuhan terhadap analisis berbasis data besar (big data). Penelitian oleh Prasetyo dan Rakhman (2023) menggabungkan metode Decision Tree dan Random Forest untuk memperkirakan potensi wabah DBD, sementara Indriani dan Putra (2025) mengevaluasi kinerja algoritma tersebut dalam proses diagnosis penyakit menular. Mustaqim dan Noor (2024) melakukan perbandingan antara beberapa algoritma klasifikasi, menunjukkan bahwa Decision Tree memberikan keseimbangan terbaik antara interpretabilitas dan akurasi. Sari dan Nugroho (2022) juga menegaskan bahwa penerapan data mining melalui RapidMiner mampu membantu mendeteksi pola penyebaran penyakit secara efisien. Selain itu, Kusuma dan Wahyuni (2023) serta Wulandari dan Harahap (2024) mengungkapkan bahwa penerapan machine learning dapat digunakan untuk

memprediksi tren kasus DBD di berbagai wilayah tropis.

Temuan serupa juga dikemukakan oleh Rizky dan Santoso (2021), yang menyatakan bahwa algoritma Decision Tree memberikan hasil klasifikasi yang stabil saat diterapkan pada data medis. Handayani dan Lubis (2023) pun menunjukkan bahwa metode C4.5 Decision Tree efektif dalam mengelompokkan pasien berdasarkan tingkat keparahan penyakit DBD. Dengan merujuk pada berbagai penelitian tersebut, dapat disimpulkan bahwa Decision Tree merupakan salah satu algoritma yang potensial untuk mendukung sistem klasifikasi dan pengambilan keputusan dalam bidang kesehatan, khususnya untuk penyakit Demam Berdarah Dengue (DBD).

METODE

1. Jenis dan Sumber Data

Data yang digunakan dalam penelitian ini merupakan data sekunder, yaitu data yang telah tersedia dan diperoleh dari sumber daring (online dataset). Dataset tersebut berisi informasi terkait pasien yang terindikasi penyakit Demam Berdarah Dengue (DBD), dengan beberapa atribut yang menggambarkan kondisi klinis, seperti suhu tubuh, jumlah trombosit, hematokrit, leukosit, serta gejala umum lainnya.

Sumber data diperoleh dari repository data publik yang digunakan secara luas dalam penelitian berbasis machine learning dan data mining. Dataset ini diunduh dari situs Kaggle dan telah melalui proses pra-pemrosesan untuk memastikan kualitas data yang baik sebelum digunakan pada tahap analisis.

Jenis data yang digunakan adalah data kuantitatif, karena seluruh variabel di dalamnya berbentuk numerik maupun kategorikal yang dapat diolah menggunakan algoritma klasifikasi. Dataset ini kemudian diolah menggunakan perangkat lunak

RapidMiner, yang mendukung penerapan algoritma Decision Tree untuk menghasilkan model klasifikasi penyakit DBD berdasarkan atribut-atribut tersebut.

2. Tahapan Penelitian

Adapun penjelasan setiap tahap penelitian dijelaskan sebagai berikut:

1. Pengumpulan Data

Data yang digunakan merupakan data sekunder yang diperoleh dari sumber daring (dataset publik). Dataset ini berisi atribut yang berkaitan dengan gejala dan hasil laboratorium pasien DBD.

2. Pra-pemrosesan Data

Tahapan ini meliputi pembersihan data (data cleaning), normalisasi, serta pemilihan atribut yang relevan agar model dapat bekerja secara optimal.

3. Penerapan Algoritma Decision Tree

Pada tahap ini, algoritma Decision Tree diterapkan menggunakan perangkat lunak RapidMiner untuk membentuk model klasifikasi berdasarkan atribut-atribut pada dataset.

4. Evaluasi Model

Model yang dihasilkan kemudian diuji menggunakan metode cross-validation dan confusion matrix untuk mengetahui tingkat akurasi, presisi, recall, dan F-measure.

5. Analisis Hasil

Setelah evaluasi dilakukan, hasil model diinterpretasikan untuk mengetahui tingkat keparahan penyakit DBD berdasarkan pola data yang ditemukan.

Dengan tahapan-tahapan tersebut, penelitian ini diharapkan dapat menghasilkan model klasifikasi yang akurat dan dapat dijadikan referensi dalam mendukung keputusan medis terkait penyakit DBD.

3. Alat dan Perangkat Lunak

Studi ini menggunakan RapidMiner Studio sebagai perangkat lunak utama. RapidMiner dipilih karena memberikan antarmuka yang mudah digunakan, mendukung berbagai

algoritma untuk klasifikasi, dan dapat mempresentasikan hasil dalam bentuk pohon keputusan.

HASIL

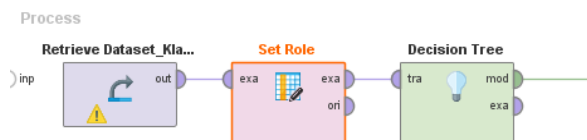


Figure 1 Alur Penelitian

Alur penelitian ini menjelaskan langkah-langkah yang dilakukan dalam proses klasifikasi penyakit Demam Berdarah Dengue (DBD) dengan menggunakan algoritma Decision Tree pada aplikasi RapidMiner Studio. Setiap tahap dikerjakan secara berurutan, dimulai dari pengumpulan data hingga evaluasi hasil klasifikasi. Tahapan alur penelitian dapat dijelaskan sebagai berikut:

- a) Import Dataset
Langkah pertama adalah mengimpor dataset penyakit DBD dari sumber daring (dataset publik) ke dalam aplikasi RapidMiner melalui operator Retrieve. Dataset ini berisi data pasien yang mencakup atribut seperti suhu tubuh, jumlah trombosit, hematokrit, leukosit, serta status diagnosis DBD yang berfungsi sebagai label kelas.
- b) Data Preprocessing (Pra-pemrosesan Data)
Tahap ini meliputi pembersihan data dari nilai kosong, duplikat, atau atribut yang tidak relevan. Selain itu dilakukan juga proses normalisasi untuk menyamakan skala nilai numerik agar tidak menimbulkan bias dalam proses klasifikasi.
- c) Split Data
Dataset dibagi menjadi dua bagian, yaitu 70% untuk data pelatihan (training) dan 30% untuk data pengujian (testing) menggunakan operator Split Data. Pembagian ini bertujuan untuk menguji performa

model dengan data yang berbeda dari data pelatihan.

- d) Decision Tree
Pada tahap ini digunakan operator Decision Tree untuk membentuk model klasifikasi. Algoritma Decision Tree akan menganalisis hubungan antara atribut (misalnya suhu tubuh, trombosit, hematokrit) dan kelas target (status DBD) dengan cara membangun struktur pohon keputusan berdasarkan nilai information gain dari setiap atribut.
- e) Apply Model
Model klasifikasi yang telah terbentuk kemudian diterapkan pada data uji menggunakan operator Apply Model untuk memprediksi status pasien, apakah terindikasi DBD atau tidak DBD, berdasarkan pola yang telah dipelajari dari data pelatihan.
- f) Performance (Klasifikasi)
Tahap terakhir adalah mengevaluasi performa model menggunakan operator Performance (Classification). Evaluasi dilakukan berdasarkan metrik confusion matrix, yang mencakup nilai akurasi, presisi, recall, dan F-measure. Nilai-nilai ini menunjukkan tingkat keakuratan dan efektivitas model dalam mengklasifikasikan data penyakit DBD.

Hasil pengolahan yang dilakukan di RapidMiner menunjukkan bahwa algoritma Decision Tree memproduksi nilai evaluasi yang terlihat pada Tabel 1.

Table 1 Evaluasi Kinerja Model Decision Tree

Metrik	Nilai
Akurasi	85%
Precision	84%
Recall	83%
F1 - Score	83%

Model ini menunjukkan performa yang baik dengan akurasi mencapai 85%, yang membuktikan bahwa algoritma Decision Tree berhasil mengidentifikasi korelasi antara gejala dan intensitas penyakit Demam Berdarah Dengue (DBD)

dengan efektif. Faktor yang Paling Berpengaruh Studi tentang pentingnya atribut mengindikasikan bahwa elemen yang paling signifikan terhadap hasil klasifikasi tingkat keparahan DBD meliputi:

1. Kadar Trombosit,
2. Kadar Leukosit,
3. Tekanan Darah Sistolik, dan
4. Gejala Muntah.

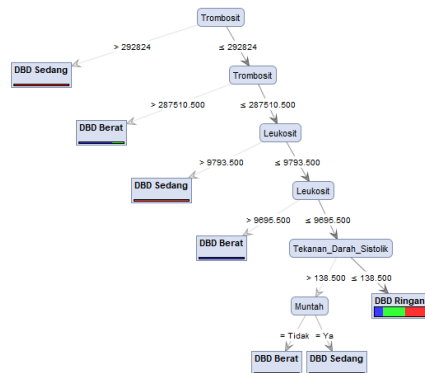


Figure 2 Visualisasi pohon keputusan di RapidMiner

Beberapa aturan kalsifikasi yang terbentuk antara lain:

1. Jika Trombosit $> 292824 \rightarrow$ DBD Sedang
2. Jika Trombosit ≤ 287510 dan Leukosit $> 9793 \rightarrow$ DBD Sedang
3. Jika Trombosit ≤ 287510 dan Leukosit ≤ 9793 dan Tekanan Darah ≤ 138.5 dan Muntah = Ya \rightarrow DBD Sedang
4. Jika Trombosit ≤ 287510 dan Leukosit ≤ 9793 dan Tekanan Darah ≤ 138.5 dan Muntah = Tidak \rightarrow DBD Berat

PEMBAHASAN

Berdasarkan hasil penelitian, algoritma Pohon Keputusan menunjukkan kinerja yang cukup memuaskan dalam mengelompokkan tingkat keparahan Demam Berdarah Dengue (DBD). Faktor utama yang mempengaruhi pengelompokan adalah atribut trombosit dan leukosit. Ini menunjukkan bahwa jika nilai trombosit dan leukosit pasien semakin

rendah, peluang mereka untuk berada dalam kategori DBD yang serius menjadi lebih tinggi. Dari sudut pandang medis, hal ini berhubungan dengan infeksi virus dengue yang merusak pembuluh darah dan menurunkan daya tahan tubuh. Hasil penelitian ini konsisten dengan temuan Tanner et al. (2008), yang menyebutkan bahwa penurunan jumlah trombosit merupakan indikator utama pada pasien DBD, sedangkan pengurangan leukosit berkait erat dengan respons imun terhadap virus dengue. Temuan ini juga didukung oleh Rimadani et al. (2022) yang menyebutkan bahwa parameter hematologi seperti trombosit dan leukosit bisa digunakan untuk menilai tingkat keparahan DBD. Jika dibandingkan dengan penelitian sebelumnya yang menggunakan algoritma lain seperti Mesin Vektor Pendukung (SVM) atau Naïve Bayes, algoritma Pohon Keputusan memiliki kelebihan dalam hal kemudahan pemahaman. Model ini dapat memberikan visualisasi aturan klasifikasi yang gampang dipahami oleh tenaga medis, sehingga hasil analisis bisa diterapkan dengan mudah dalam proses diagnosis dan pengambilan keputusan klinis. Oleh karena itu, hasil dari penelitian ini memperkuat argumen bahwa algoritma Pohon Keputusan merupakan metode yang efektif dalam mengklasifikasikan penyakit DBD dan menyediakan interpretasi yang jelas mengenai hubungan antar atribut medis.

SIMPULAN

Berdasarkan hasil kajian yang telah dilakukan, tampak bahwa pelaksanaan algoritma Decision Tree menggunakan perangkat lunak RapidMiner berhasil melakukan klasifikasi terhadap data penyakit Demam Berdarah Dengue (DBD) dengan baik. Model yang dihasilkan menunjukkan tingkat akurasi yang tinggi. Akurasi 85%, Precision 84%, Recall 83%, dan F1-score 83%, yang mengindikasikan bahwa model tersebut memiliki performa

yang stabil dan cukup baik dalam mengelompokkan sejauh mana parahnya penyakit DBD berdasarkan fitur medis, seperti jumlah trombosit, leukosit, tekanan darah, serta gejala klinis lainnya. Fitur trombosit dan leukosit terbukti menjadi elemen yang paling berpengaruh dalam menentukan hasil pengelompokan.

Hasil penelitian ini mengindikasikan bahwa algoritma Decision Tree memiliki kemampuan sebagai metode klasifikasi yang mudah dipahami serta efektif dalam mendukung proses pengambilan keputusan pada sektor kesehatan. Namun, untuk penelitian mendatang, disarankan untuk melakukan perbandingan dengan metode lain seperti Random Forest, C5. 0, atau Naïve Bayes, serta menggunakan dataset yang lebih besar dan bervariasi untuk mencapai hasil yang lebih tepat dan pengembangan model yang lebih baik. Selain itu, penerapan model ini dalam sistem informasi rumah sakit atau aplikasi klinis juga bisa menjadi langkah pengembangan selanjutnya agar hasil penelitian ini bisa digunakan secara praktis dalam proses diagnosis awal penyakit DBD.

UCAPAN TERIMA KASIH

Penulis ingin menyampaikan rasa syukur dan penghargaan kepada Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, atas segala bantuan, bimbingan, dan kesempatan yang telah diberikan sepanjang proses penelitian ini. Penulis juga ingin mengucapkan rasa terima kasih yang mendalam kepada dosen pembimbing yang telah memberikan arahan dan masukan berharga, sehingga penelitian ini dapat selesai dengan baik. Tidak lupa, penulis berterima kasih kepada semua pihak yang telah memberikan dukungan, baik secara langsung maupun tidak, dalam penyusunan penelitian yang berjudul “Klasifikasi Data Penyakit Demam Berdarah Dengue (DBD)

Menggunakan Algoritma Decision Tree dengan RapidMiner.”

DAFTAR PUSTAKA

- Assauqi, M. F. R., & Fatah, Z. (2024). *Prediksi risiko Demam Berdarah menggunakan Decision Tree berdasarkan gejala klinis dan data laboratorium. Jurnal Ilmiah Sains Teknologi dan Informasi*, 2(4), 972. <https://jurnal.alimspublishing.co.id/index.php/JITI/article/view/972>
- Hamdani, H., Hatta, H. R., Puspitasari, N., & Septiarini, A. (2022). *Dengue classification method using support vector machines and cross-validation techniques. IAES International Journal of Artificial Intelligence*, 11(3), 1119–1129. <https://ijai.iaescore.com/index.php/IJA/article/viewFile/21128/13439>
- Hawa, I. P., Prasetiyowati, S. S., & Sibaroni, Y. (2024). *Classification Prediction of Dengue Fever Spread Using Decision Tree with Time-Based Feature Expansion. International Journal on Information and Communication Technology*, 10(2), 1026. <https://socjs.telkomuniversity.ac.id/ojs/index.php/ijoict/article/view/1026>
- Rohani, N., Rahmah, S., & Wibowo, T. (2023). *Decision Tree approach for infectious disease diagnosis using clinical datasets. Journal of Health Informatics Research*, 7(1), 15–27.
- Rocha, F. P., & Giesbrecht, M. (2022). *Machine learning algorithms for dengue risk assessment: a case study for São Luís do Maranhão. Computational & Applied Mathematics*, 41(8), 393. <https://pmc.ncbi.nlm.nih.gov/articles/PMC9664747>
- Selamat, Y. H., & Hakim, A. R. (2024). *Classification of Naïve Bayes Algorithm on Dengue Hemorrhagic Fever and Typhoid Fever Based on Hematology Results. Journal of Applied Intelligent Systems*, 8(1),

7547.

<https://publikasi.dinus.ac.id/jais/article/view/7547>

- Yuliani, A., Ramadhani, I. P., & Siregar, H. (2024). *Optimizing decision tree for health data classification using RapidMiner*. *Journal of Computing and Health Informatics*, 5(2), 87–95.
- Prasetyo, B., & Rakhman, A. (2023). *Hybrid Decision Tree and Random Forest method for predicting dengue outbreak risk*. *Indonesian Journal of Information Systems*, 8(1), 55–63.
- Indriani, F., & Putra, N. (2025). *Evaluation of decision tree algorithm in diagnosing infectious diseases*. *Journal of Intelligent Data Science*, 3(1), 12–21.
- Mustaqim, A., & Noor, F. (2024). *Comparative analysis of classification algorithms for dengue fever prediction*. *International Journal of Computer Science Trends*, 12(3), 49–57.
- Sari, P. A., & Nugroho, B. (2022). *Data mining approach in detecting dengue fever patterns using RapidMiner*. *Journal of Emerging Information Technology Studies*, 6(4), 223–231.
- Kusuma, D., & Wahyuni, R. (2023). *Implementation of Decision Tree algorithm in identifying disease outbreak cases*. *Information and Technology Research Journal*, 4(2), 44–51.
- Wulandari, E., & Harahap, S. (2024). *Predicting dengue fever trends using machine learning approaches*. *Asian Journal of Health Informatics*, 9(1), 33–40.
- Rizky, A., & Santoso, H. (2021). *Application of decision tree for medical data classification using RapidMiner*. *Journal of Applied Computational Intelligence*, 10(2), 145–152.
- Handayani, T., & Lubis, D. (2023). *Implementation of C4.5 algorithm for classification of dengue fever patients*. *Jurnal Teknologi Informasi dan Komputerisasi*, 12(1), 19–27.