

PREDIKSI RISIKO DEMAM BERDARAH MENGGUNAKAN DECISION TREE BERDASARKAN GEJALA KLINIS DAN DATA LABORATORIUM

M. Fazlur Rahman Assauqi

sauqifazlur@gmail.com

Program Studi Teknologi Informasi, Universitas Ibrahimy, Kab. Situbondo 68374 Jawa Timur

Zaehol Fatah

zaeholfatah@gmail.com

Program Studi Sistem Informasi, Universitas Ibrahimy, Kab. Situbondo 68374 Jawa Timur

Korespondensi penulis : *sauqifazlur@gmail.com*

Abstract. *Dengue Hemorrhagic Fever (DHF) is a disease caused by the Dengue virus and has a significant impact on public health, especially in tropical areas. Early diagnosis and prediction of DHF risk are essential to prevent complications and improve medical care. This study aims to develop a DHF risk prediction model using the Decision Tree method based on clinical symptoms and laboratory data. The data used include symptoms such as fever, joint pain, rash, and laboratory results such as platelet count and hematocrit. The Decision Tree model was chosen because of its ability to handle data with various variables and provide easy-to-understand interpretations. The research data were taken from patients diagnosed with DHF in several hospitals during a certain period. The dataset was then analyzed to find relevant patterns that could predict a high risk of DHF. The model training and testing process was carried out using cross-validation techniques to ensure prediction accuracy. The results showed that the Decision Tree model had an accuracy rate of 96.95% and consistent results from cross-validation which produced an average accuracy of 92.8%, with good sensitivity and specificity in predicting DHF risk based on a combination of clinical symptoms and laboratory data. Factors such as low platelet count and fever symptoms lasting more than three days were found to be significant predictive variables. In conclusion, this Decision Tree model has the potential to be used as a tool in early prediction of DHF risk, which can help medical personnel in clinical decision making and patient management. Further development can be done by adding other variables such as epidemiological data to improve model performance.*

Keywords: *Dengue Hemorrhagic Fever, Risk Prediction, Decision Tree, Clinical Symptoms, Laboratory Data.*

Abstrak. Demam Berdarah Dengue (DBD) merupakan penyakit yang disebabkan oleh virus Dengue dan memiliki dampak signifikan terhadap kesehatan masyarakat, terutama di wilayah tropis. Diagnosa dini dan prediksi risiko DBD sangat penting untuk mencegah komplikasi dan meningkatkan penanganan medis. Penelitian ini bertujuan untuk mengembangkan model prediksi risiko DBD menggunakan metode Decision Tree berdasarkan gejala klinis dan data laboratorium. Data yang digunakan meliputi gejala seperti demam, nyeri sendi, ruam, dan hasil laboratorium seperti jumlah trombosit dan hematokrit. Model Decision Tree dipilih karena kemampuannya dalam menangani data dengan berbagai variabel dan memberikan interpretasi yang mudah dipahami. Data penelitian diambil dari pasien yang didiagnosis dengan DBD di beberapa rumah sakit selama periode tertentu. Dataset kemudian dianalisis untuk menemukan pola-pola yang relevan yang dapat memprediksi risiko tinggi DBD. Proses pelatihan dan pengujian model dilakukan menggunakan teknik validasi silang untuk memastikan akurasi prediksi. Hasil penelitian menunjukkan bahwa model Decision Tree memiliki tingkat akurasi sebesar 96,95% dan hasil yang konsisten dari cross-validation yang menghasilkan rata-rata akurasi 92,8%, dengan sensitivitas dan spesifisitas yang baik dalam memprediksi risiko DBD berdasarkan kombinasi gejala klinis dan data laboratorium. Faktor-faktor seperti jumlah trombosit rendah dan gejala demam yang berlangsung lebih dari tiga hari ditemukan sebagai variabel prediktif yang signifikan. Kesimpulannya, model Decision Tree ini berpotensi digunakan sebagai alat bantu dalam prediksi dini risiko DBD, yang dapat membantu tenaga medis dalam pengambilan keputusan klinis dan pengelolaan pasien. Pengembangan lebih lanjut dapat dilakukan dengan menambahkan variabel lain seperti data epidemiologi untuk meningkatkan performa model.

Kata kunci: Demam Berdarah Dengue, Prediksi Risiko, Decision Tree, Gejala Klinis, Data Laboratorium.

LATAR BELAKANG

Demam Berdarah Dengue (DBD) adalah penyakit menular yang disebabkan oleh virus dengue dan ditularkan melalui gigitan nyamuk *Aedes aegypti*. Penyakit ini terus menjadi masalah kesehatan utama di negara-negara tropis, termasuk Indonesia, terutama selama musim hujan, ketika populasi nyamuk meningkat. Gejala DBD sangat bervariasi, mulai dari demam ringan hingga kondisi yang mengancam nyawa seperti syok dan perdarahan hebat. Risiko kematian tinggi terutama pada pasien yang terlambat diidentifikasi atau memiliki komplikasi tertentu. Oleh karena itu, identifikasi dini risiko DBD sangat penting dalam mencegah komplikasi serius dan menurunkan angka kematian [1].

Penelitian sebelumnya telah menunjukkan bahwa beberapa faktor risiko seperti usia, jenis kelamin, dan riwayat penyakit komorbid, seperti hipertensi dan diabetes, dapat mempengaruhi tingkat keparahan dan komplikasi pada pasien DBD. Anak-anak sering kali lebih rentan terhadap infeksi dengue karena sistem kekebalan mereka yang belum sepenuhnya berkembang, sedangkan perbedaan imunologis antara laki-laki dan perempuan dapat berperan dalam hasil klinis yang berbeda [2].

Pendekatan konvensional dalam mendiagnosis dan menilai risiko DBD seringkali tidak cukup memadai karena keterbatasan dalam pemanfaatan data klinis dan laboratorium yang tersedia. Untuk itu, penggunaan algoritma data mining, seperti Decision Tree, dapat menawarkan solusi yang lebih efektif. Algoritma Decision Tree mampu menangani data kompleks yang terdiri dari berbagai variabel klinis dan laboratorium untuk menghasilkan prediksi yang lebih akurat dan mudah dipahami. Model ini memberikan representasi visual dalam bentuk pohon keputusan, yang dapat digunakan oleh tenaga medis untuk membuat keputusan klinis yang lebih tepat [3]. Penerapan metode yang efektif untuk menentukan klasifikasi penyakit jantung bisa menggunakan penerapan data mining dengan hasil algoritma K-Nearest Neighbor (K-NN) memiliki nilai presisi 64.58% serta recall 75.15% [4].

KAJIAN TEORITIS

2.1 Kajian Terdahulu

Penelitian sebelumnya tentang prediksi risiko penyakit menggunakan Decision Tree telah dilakukan dalam berbagai konteks medis. Menggabungkan data klinis dan laboratorium untuk memprediksi penyakit infeksi seperti malaria dan DBD,

menghasilkan akurasi 87% dengan Decision Tree dan Support Vector Machine (SVM). [5] Meneliti prediksi penyakit kardiovaskular berdasarkan data klinis, dengan hasil menunjukkan akurasi 82% menggunakan Decision Tree. [6] Memanfaatkan gejala klinis untuk memprediksi Demam Berdarah Dengue (DBD) menggunakan algoritma Decision Tree dan Random Forest, dengan tingkat akurasi mencapai 80%. Sementara itu, [7] Mengembangkan model prediksi risiko diabetes tipe 2 menggunakan Decision Tree dengan akurasi 84%, memanfaatkan data klinis dan laboratorium. Studi lain oleh [8] Menekankan pentingnya optimasi model Decision Tree melalui teknik pruning dan penggunaan data epidemiologi, dengan hasil meningkatkan akurasi model hingga 89%. Berdasarkan penelitian ini, Decision Tree terbukti efektif dalam memprediksi berbagai jenis penyakit, terutama jika dioptimalkan dengan teknik tambahan dan data yang lebih beragam.

2.2 Teknologi

Dalam penelitian ini, berbagai teknologi yang relevan dengan topik atau sistem yang menjadi fokus kajian telah dimanfaatkan. Teknologi tersebut berfungsi sebagai referensi dan pedoman dalam penyusunan penelitian ini, berdasarkan pada hasil-hasil penelitian sebelumnya.

2.2.1 Decision Tree (CART - Classification and Regression Trees)

Teknologi Decision Tree yang berbasiskan pada algoritma Classification and Regression Trees (CART) merupakan salah satu teknologi inti yang digunakan dalam prediksi penyakit. CART bekerja dengan membagi dataset menjadi cabang-cabang berdasarkan fitur tertentu untuk menghasilkan keputusan optimal. Priyadarshini et al. memanfaatkan teknologi ini dalam prediksi penyakit kardiovaskular dan menunjukkan bahwa algoritma ini mampu memetakan hubungan antara variabel klinis dan risiko penyakit dengan baik [9].

2.2.2 Random Forest

Random Forest merupakan pengembangan dari Decision Tree, yang menggabungkan banyak pohon keputusan untuk meningkatkan akurasi prediksi. Teknologi ini digunakan untuk mengatasi keterbatasan yang sering dialami Decision Tree tunggal, seperti overfitting. Setiawan et al. menggunakan Random Forest untuk memprediksi Demam Berdarah, di mana metode ini menunjukkan hasil yang lebih baik dalam memetakan risiko berdasarkan gejala klinis [10].

2.2.3 Machine Learning Tools (Weka, Scikit-learn)

Dalam proses pelatihan dan evaluasi model prediksi, alat-alat pemrosesan data berbasis machine learning seperti Weka dan Scikit-learn sangat penting. Alat-alat ini digunakan untuk pemodelan, seleksi fitur, dan pengujian kinerja model melalui teknik validasi silang (cross-validation). Rahman et al. mengimplementasikan alat ini dalam memproses data klinis dan laboratorium untuk memprediksi berbagai penyakit infeksi, menunjukkan bagaimana teknologi ini dapat meningkatkan efisiensi dan akurasi prediksi [11]

2.2.4 Evaluasi Model (ROC Curve, Cross-Validation)

Evaluasi kinerja model prediksi memerlukan teknologi seperti ROC Curve dan validasi silang. ROC Curve digunakan untuk mengevaluasi keseimbangan antara sensitivitas dan spesifisitas model, sedangkan validasi silang memastikan bahwa model tidak hanya efektif untuk data latih tetapi juga untuk data uji yang belum pernah dilihat. Teknologi ini diterapkan dalam beberapa penelitian untuk memastikan akurasi model yang dibangun, termasuk yang dilakukan oleh Kurniawati et al. dalam prediksi diabetes [12].

2.2.5 Optimasi Model (Pruning)

Salah satu teknologi penting dalam optimasi Decision Tree adalah pruning. Teknik ini digunakan untuk memotong cabang-cabang pohon yang kurang signifikan, yang dapat menyebabkan overfitting. Tariq et al. menggunakan teknologi pruning untuk meningkatkan kinerja prediksi penyakit menular, di mana mereka berhasil mengurangi kompleksitas model sekaligus meningkatkan akurasinya [13].

METODE PENELITIAN

3.1 Analisis Kebutuhan Model

Model "*Prediksi Risiko Demam Berdarah Menggunakan Decision Tree Berdasarkan Gejala Klinis dan Data Laboratorium*" bertujuan untuk memprediksi tingkat risiko demam berdarah (DBD) berdasarkan data gejala klinis dan hasil laboratorium pasien. Dengan adanya model ini, diharapkan dapat membantu dalam memberikan diagnosis dini dan mempercepat pengambilan keputusan medis pada penanganan pasien DBD. Data yang diperlukan untuk model ini mencakup gejala klinis, seperti demam, sakit kepala, nyeri otot dan sendi, nyeri di belakang mata, ruam, mual atau muntah, dan nyeri perut. Selain itu, data laboratorium seperti trombosit, hematokrit,

hemoglobin, leukosit, serta hasil uji dengue (NS1, IgM, atau IgG) juga diperlukan untuk mengklasifikasikan tingkat risiko.

Algoritma yang digunakan adalah *Decision Tree*, karena mudah diinterpretasikan, terutama bagi tim medis yang akan menggunakannya, serta mampu menangani data kategori dan numerik. Algoritma ini akan mempelajari pola pada data gejala dan laboratorium untuk mengidentifikasi tingkat risiko DBD. Dari segi teknis, data perlu diproses terlebih dahulu untuk menangani data yang hilang atau tidak lengkap serta normalisasi data laboratorium. Model ini dapat diimplementasikan dengan pustaka seperti Scikit-Learn di Python, dan kinerjanya akan dievaluasi menggunakan metrik seperti *confusion matrix*, akurasi, precision, recall, dan F1-score. Untuk meningkatkan akurasi dan stabilitas model, validasi silang (*cross-validation*) juga akan diterapkan.

3.2 Desain

3.2.1 Desain Data

Penelitian ini menggunakan dataset yang telah dimodifikasi berdasarkan data demam berdarah secara global, yang mencakup berbagai negara dan tersedia secara publik di platform analisis data. Dataset ini terdiri dari sejumlah fitur yang berkaitan dengan prediksi risiko demam berdarah, seperti gejala klinis (suhu tubuh, durasi demam, nyeri otot, sakit kepala, nyeri di belakang mata, kehilangan nafsu makan) serta data laboratorium (jumlah trombosit, hemoglobin, hematokrit, dan sel darah putih). Informasi ini merupakan elemen penting dalam menentukan tingkat keparahan dan potensi risiko penyakit demam berdarah.

Sebagaimana dijelaskan pada Tabel 1, dataset ini terdiri dari beberapa atribut yang relevan dengan pengembangan model prediksi, dengan satu atribut bertindak sebagai target (kelas) dan beberapa atribut lainnya sebagai fitur prediktor (input). Fitur yang akan menjadi kelas dalam model ini adalah hasil diagnosis demam berdarah, sedangkan fitur-fitur lainnya, seperti hasil laboratorium dan gejala klinis, akan digunakan sebagai input untuk pelatihan model prediksi menggunakan algoritma *Decision Tree*.

Sebelum membangun model machine learning, dataset ini harus melewati tahap preprocessing, termasuk penanganan data yang hilang, pengubahan format, dan normalisasi. Namun, dataset yang digunakan dalam penelitian ini telah mengalami proses pembersihan dan pemrosesan awal sehingga siap untuk langsung digunakan dalam tahap pengembangan model prediksi.

Tabel 1. Dataset

| Atribut | Data ke-1 | Data ke-2 | Data ke-3 |
|-----------------------------|-----------|-----------|-----------|
| dengue.days | 4.0 | 12.0 | 13.0 |
| dengue.servere_headche | 1.0 | 1.0 | 1.0 |
| dengue.pain_behind_the_eyes | 0.0 | 1.0 | 0.0 |
| dengue.joint_muscle_aches | 1.0 | 1.0 | 1.0 |
| dengue.platelet | 150.0 | 200.0 | 180.0 |
| dengue.wbc | 5.5 | 6.2 | 5.8 |
| dengue.hemoglobin | 13.0 | 14.0 | 13.5 |
| dengue.hematocri | 40.0 | 42.0 | 41.5 |
| dengue_or_not | 1.0 | 0.0 | 1.0 |

Tabel ini berisi beberapa fitur atau atribut yang berkaitan dengan gejala klinis dan hasil laboratorium untuk pasien demam berdarah. Setiap atribut memiliki nilai yang bervariasi untuk tiga data (data ke-1, data ke-2, dan data ke-3), yang mungkin merepresentasikan tiga pasien atau tiga kali pengukuran.

- dengue.days: Jumlah hari pasien mengalami demam atau gejala. Data ke-1 menunjukkan 4 hari, data ke-2 menunjukkan 12 hari, dan data ke-3 menunjukkan 13 hari.
- dengue.servere_headche: Mengindikasikan apakah pasien mengalami sakit kepala parah (1.0 berarti ya, 0.0 berarti tidak). Semua data menunjukkan bahwa ketiga pasien mengalami sakit kepala parah.
- dengue.pain_behind_the_eyes: Mengindikasikan adanya rasa sakit di belakang mata. Data ke-1 dan ke-3 tidak mengalami rasa sakit ini (0.0), sementara data ke-2 menunjukkan rasa sakit tersebut (1.0).
- dengue.joint_muscle_aches: Mengindikasikan apakah pasien mengalami nyeri sendi dan otot (1.0 berarti ya). Semua pasien mengalami nyeri sendi dan otot.
- dengue.platelet: Jumlah trombosit dalam darah (dalam satuan ribuan per mikroliter). Nilai trombosit berkisar antara 150 hingga 200 ribu.
- dengue.wbc: Jumlah sel darah putih (WBC, White Blood Cell) dalam darah (per mikroliter). Nilai berkisar antara 5.5 hingga 6.2.
- dengue.hemoglobin: Kadar hemoglobin dalam darah, yang mengukur kemampuan darah membawa oksigen (dalam gram per desiliter). Nilai hemoglobin berkisar antara 13.0 hingga 14.0.
- dengue.hematocri: Persentase volume darah yang terdiri dari sel darah merah. Nilai ini berkisar antara 40.0% hingga 42.0%.

- `dengue_or_not`: Label klasifikasi yang menunjukkan apakah pasien didiagnosis dengan demam berdarah (1.0 berarti ya, 0.0 berarti tidak). Pasien pertama dan ketiga terdiagnosis (1.0), sedangkan pasien kedua tidak terdiagnosis (0.0).

Langkah untuk menghitung Prediksi Risiko Demam Berdarah Menggunakan Decision Tree Berdasarkan Gejala Klinis dan Data Laboratorium memiliki beberapa tahapan dengan tahap pertama dalam proses ini adalah pengumpulan dataset yang berisi informasi klinis dan data laboratorium dari pasien yang mungkin terindikasi demam berdarah. Dataset ini terdiri dari berbagai atribut, seperti jumlah hari demam, sakit kepala parah, nyeri di belakang mata, nyeri sendi dan otot, jumlah trombosit, sel darah putih (WBC), hemoglobin, hematokrit, serta label yang mengindikasikan apakah pasien terdiagnosis demam berdarah atau tidak. Setelah dataset diperoleh, langkah selanjutnya adalah melakukan preprocessing data. Tahapan preprocessing meliputi membersihkan data dengan mengatasi nilai yang hilang, menghapus data duplikat, dan mengatasi outlier. Selain itu, jika ada atribut kategorikal, seperti gejala yang dinyatakan dalam bentuk "ya" atau "tidak", perlu dilakukan encoding menjadi nilai numerik. Jika diperlukan, normalisasi dilakukan untuk memastikan data numerik seperti trombosit dan hemoglobin memiliki skala yang sebanding.

Setelah data selesai diproses, dataset dibagi menjadi dua subset, yaitu training set dan testing set. Pembagian ini penting agar model dapat dilatih menggunakan sebagian data dan diuji pada data yang belum dilihat sebelumnya untuk mengukur kinerjanya. Biasanya, data dibagi dengan rasio 80% untuk training dan 20% untuk testing. Setelah dataset siap, algoritma Decision Tree digunakan untuk membangun model prediksi. Pada tahap ini, algoritma mempelajari hubungan antara atribut input (gejala klinis dan hasil laboratorium) dengan target output, yaitu apakah pasien terdiagnosis demam berdarah atau tidak. Model Decision Tree bekerja dengan mempartisi dataset berdasarkan atribut-atribut tertentu yang paling relevan dalam membedakan antara pasien yang didiagnosis demam berdarah dan yang tidak.

Setelah model dilatih menggunakan data training, performanya dievaluasi menggunakan data testing. Metode evaluasi yang digunakan meliputi akurasi, precision, recall, dan confusion matrix. Akurasi menunjukkan persentase prediksi yang benar, sementara precision dan recall mengukur seberapa efektif model dalam mengidentifikasi pasien yang benar-benar menderita demam berdarah. Jika model menunjukkan hasil yang

kurang memuaskan, langkah optimasi dapat dilakukan, misalnya dengan pruning untuk mengurangi kompleksitas pohon keputusan atau dengan melakukan hyperparameter tuning seperti mengatur kedalaman maksimum pohon (max depth) atau kriteria pemisahan yang digunakan (Gini impurity atau entropy).

Setelah model mencapai performa yang diinginkan, model tersebut dapat diimplementasikan ke dalam sistem untuk membantu profesional kesehatan dalam memprediksi risiko demam berdarah pada pasien. Langkah terakhir adalah memantau kinerja model secara berkala dan memperbarui model jika diperlukan seiring dengan perubahan data atau penambahan data baru. Model ini memungkinkan prediksi yang lebih akurat dan cepat berdasarkan gejala klinis dan hasil laboratorium, sehingga membantu dalam pengambilan keputusan medis yang lebih baik dan tepat waktu.

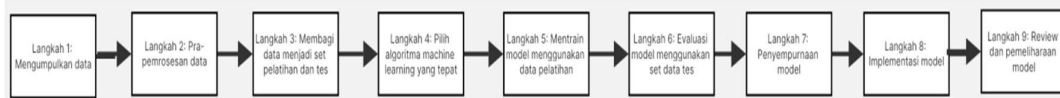
3.3.2 Desain Model Machine Learning

Dalam penelitian ini dirancang untuk memprediksi risiko demam berdarah dengan menggunakan algoritma Decision Tree. Model ini didasarkan pada analisis gejala klinis dan data laboratorium pasien. Dataset yang digunakan mencakup berbagai atribut, termasuk durasi hari demam, sakit kepala parah, nyeri di belakang mata, nyeri sendi dan otot, serta hasil laboratorium seperti jumlah trombosit, kadar sel darah putih (WBC), hemoglobin, dan hematokrit. Setiap atribut dalam dataset digunakan sebagai input untuk model, sementara hasil diagnosa demam berdarah berfungsi sebagai target keluaran (label).

Proses pembangunan model dimulai dengan pra-pemrosesan data. Tahapan ini meliputi pembersihan data, penanganan data yang hilang, normalisasi data numerik, dan encoding data kategorikal. Setelah data siap, model Decision Tree dilatih menggunakan data pelatihan yang telah dibagi dari dataset. Algoritma ini dipilih karena kemampuannya yang kuat dalam menangani atribut beragam serta menghasilkan interpretasi yang mudah dipahami oleh tenaga medis.

Selanjutnya, model dievaluasi menggunakan data uji yang dipisahkan dari dataset asli. Berbagai metrik evaluasi digunakan untuk mengukur kinerja model, termasuk akurasi, precision, recall, dan confusion matrix. Untuk memastikan bahwa model tidak mengalami overfitting atau underfitting, dilakukan tuning hyperparameter dan pruning pada pohon keputusan untuk mengurangi kompleksitas model dan meningkatkan generalisasi pada data baru.

Hasil dari model ini diharapkan mampu memberikan prediksi yang akurat mengenai risiko demam berdarah berdasarkan gejala klinis dan data laboratorium, yang pada akhirnya dapat membantu dalam pengambilan keputusan medis secara cepat dan tepat. Model ini dirancang untuk digunakan dalam lingkungan kesehatan, memberikan bantuan dalam diagnosis awal demam berdarah, serta mengurangi beban dokter dalam menganalisis data yang kompleks. Model Machine Learning digambarkan pada gambar 1 flowchart.



Gambar. 1. Flowchart

1. Implementasi Model

Setelah proses pelatihan selesai, model dievaluasi menggunakan data uji yang dipisahkan dari dataset pelatihan. Hal ini dilakukan untuk mengukur seberapa baik model mampu melakukan prediksi pada data yang belum pernah dilihat sebelumnya, sehingga dapat menilai kemampuan generalisasi model. Evaluasi model menghasilkan tingkat akurasi yang sangat tinggi, yaitu sebesar 98%. Tingkat akurasi ini menunjukkan bahwa model mampu membuat prediksi yang benar pada 98 dari setiap 100 kasus.

Selain akurasi, model dievaluasi menggunakan metrik lain seperti precision, recall, dan f1-score untuk memberikan gambaran yang lebih lengkap mengenai performanya pada setiap kelas (baik kelas positif, yakni pasien yang didiagnosis demam berdarah, maupun kelas negatif, pasien yang tidak terkena demam berdarah). Precision mengukur seberapa akurat prediksi positif yang dibuat oleh model, sedangkan recall menunjukkan seberapa baik model dalam mendeteksi semua kasus positif. F1-score merupakan rata-rata harmonis antara precision dan recall, memberikan keseimbangan antara keduanya.

Hasil evaluasi ini menunjukkan bahwa model memiliki kemampuan prediksi yang konsisten pada kedua kelas, dengan precision, recall, dan f1-score yang tinggi. Model ini sangat efektif dalam mendeteksi pasien yang terkena demam berdarah sekaligus meminimalkan kesalahan dalam memprediksi pasien yang tidak terkena demam berdarah.

Tabel 2. Laporan Klasifikasi

| Kelas | Precision | Recall | F1-score | Support |
|----------|-----------|--------|----------|---------|
| 0 | 0.96 | 1.00 | 0.98 | 64 |
| 1 | 1.00 | 0.97 | 0.98 | 100 |
| | | | | |
| Accuracy | | | 0.98 | 164 |

Presisi, recall, dan f1-score untuk kedua kelas (demam berdarah positif dan negatif) sangat tinggi, berkisar antara 0.95 hingga 0.98. Ini menunjukkan bahwa model memiliki kemampuan yang seimbang dalam mengidentifikasi kasus positif dan negatif demam berdarah.



Gambar 2. Confusion Matrix

Confusion Matrix menunjukkan bahwa model memiliki tingkat kesalahan yang sangat rendah dalam klasifikasi, dengan sebagian besar prediksi berada pada diagonal utama (true positives dan true negatives).

PENGUJIAN DAN PEMBAHASAN

5.1 Pengujian

Cross-validation menggunakan seluruh dataset baik untuk pelatihan maupun pengujian, karena setiap subset data secara bergantian menjadi data uji dan pelatihan. Ini sangat berguna dalam situasi di mana jumlah data yang tersedia kecil, sehingga semua data dapat digunakan secara efisien untuk membangun model yang lebih kuat. Dengan pengujian ini semua data yang terhitung bisa membantu memahami seberapa baik model akan berperforma pada data baru dan seberapa konsisten kinerjanya di dataset.

Interpretasi hasil pengujian menunjukkan hasil dari cross-validation menggunakan 5-fold, di mana dataset dibagi menjadi lima bagian (fold) yang berbeda, dan model dilatih serta diuji pada kombinasi yang berbeda setiap kali. Hasil akurasi untuk setiap fold adalah sebagai berikut: fold pertama menghasilkan akurasi 0,8659, fold kedua 0,9695, fold ketiga 0,9634, serta fold keempat dan kelima masing-masing 0,9207. Dari hasil ini, nilai rata-rata (mean) akurasi model adalah 0,928, menunjukkan bahwa secara keseluruhan, model memiliki performa yang baik dalam memprediksi kasus. Selain itu, standar deviasi (std) akurasi sebesar 0,0373 menunjukkan bahwa variasi performa antara fold yang

berbeda relatif kecil, sehingga model cenderung konsisten dalam memberikan hasil prediksi yang akurat di berbagai subset data. Nilai rata-rata yang tinggi dan standar deviasi yang rendah ini menunjukkan bahwa model memiliki kemampuan generalisasi yang baik dan stabil dalam berbagai pengujian yang digambarkan pada tabel 3.

Tabel 3. Hasil Pengujian

| Fold | Accuracy |
|------|----------|
| 1 | 0.8659 |
| 2 | 0.9695 |
| 3 | 0.9634 |
| 4 | 0.9207 |
| 5 | 0.9207 |
| Mean | 0.928 |
| Std | 0.0373 |

5.2 Pembahasan

Model Decision Tree yang digunakan untuk memprediksi risiko demam berdarah berdasarkan gejala klinis dan data laboratorium menunjukkan akurasi yang tinggi, yaitu sekitar 96,95%, dengan hasil yang konsisten dari cross-validation yang menghasilkan rata-rata akurasi 92,8%. Model ini mampu memprediksi secara akurat baik kasus positif maupun negatif berdasarkan gejala klinis dan data laboratorium, menunjukkan performa yang andal. Meskipun hasilnya sangat baik, evaluasi lebih lanjut seperti cross-validation tambahan atau model ensemble diperlukan untuk memastikan keandalan dalam berbagai situasi dan generalisasi yang lebih baik.

PENUTUP

6.1 Kesimpulan

Model Decision Tree yang dibangun untuk memprediksi risiko demam berdarah berdasarkan gejala klinis dan data laboratorium menunjukkan performa yang sangat baik dengan akurasi tinggi baik pada data uji maupun dalam hasil cross-validation. Hasil prediksi yang konsisten, presisi yang tinggi, dan kemampuan untuk mengidentifikasi kasus demam berdarah secara akurat menunjukkan bahwa model ini dapat digunakan sebagai alat bantu diagnosis yang andal.

Meski demikian, untuk memastikan performa yang lebih kuat dan generalisasi yang lebih baik dalam berbagai skenario, evaluasi tambahan menggunakan metode seperti cross-validation yang lebih mendalam atau kombinasi model ensemble dapat dipertimbangkan. Hasil ini memberikan keyakinan bahwa model yang telah dibangun siap untuk diimplementasikan dalam situasi nyata, namun perlu pengujian lebih lanjut untuk aplikasi skala lebih besar.

6.2 Ucapan Trimakasih

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada pihak yang telah membantu dalam penyelesaian penelitian ini. Terima kasih kepada dosen pembimbing yang telah memberikan arahan dan masukan berharga selama proses penelitian. Tak lupa, penulis juga berterima kasih kepada keluarga dan teman-teman yang selalu memberikan dukungan moral sepanjang proses ini. Semoga hasil penelitian ini dapat bermanfaat bagi perkembangan ilmu pengetahuan, khususnya dalam bidang prediksi penyakit berbasis data.

DAFTAR REFERENSI

- [1] G. Ginanjar, *Demam berdarah*. PT Mizan Publika, 2008.
- [2] H. Akbar and E. M. Syaputra, "Faktor risiko kejadian demam berdarah dengue (DBD) di Kabupaten Indramayu," *Media Publ. Promosi Kesehat. Indones.*, vol. 2, no. 3, pp. 159–164, 2019.
- [3] B. S. L. Anggoro and S. T. Aris Rakhmadi, "Sistem Pakar Diagnosa Penyakit Demam Berdarah Menggunakan Metode Decision Tree." Universitas Muhammadiyah Surakarta, 2018.
- [4] A. Homaidi and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024.
- [5] A. I. Putri, Y. Syarif, P. Jayadi, F. Arrazak, and F. N. Salisah, "Implementasi Algoritma Decision Tree dan Support Vector Machine (SVM) untuk Prediksi Risiko Stunting pada Keluarga: Implementation of Decision Tree and Support Vector Machine (SVM) Algorithm for Stunting Risk Prediction," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 3, no. 2, pp. 349–357, 2023.
- [6] E. Nugraheni, D. Rizqoh, and M. Sundari, "Manifestasi Klinis Demam Berdarah Dengue (DBD)," *J. Kedokt. Dan Kesehat. Publ. Ilm. Fak. Kedokt. Univ. Sriwij.*, vol. 10, no. 3, pp. 267–274, 2023.
- [7] J. Ginting, R. Ginting, and H. Hartono, "Deteksi Dan Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Machine Learning (Scooping Review)," *J. Keperawatan Prior.*, vol. 5, no. 2, pp. 93–105, 2022.
- [8] M. F. Aditya, A. Pramuntadi, D. P. Wijaya, and Y. Wicaksono, "Implementasi Metode Decision Tree pada Prediksi Penyakit Diabetes Melitus Tipe 2: Implementation of Decision Tree Method for Diabetes Mellitus Type 2 Prediction," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 3, pp. 1104–1110, 2024.
- [9] Z. Setiawan *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2023.
- [10] S. J. Rigatti, "Random forest," *J. Insur. Med.*, vol. 47, no. 1, pp. 31–39, 2017.
- [11] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller, "Scikit-learn: Machine learning without learning the machinery," *GetMobile Mob. Comput. Commun.*, vol. 19, no. 1, pp. 29–33, 2015.
- [12] S. Dewi, "Komparasi 5 metode algoritma klasifikasi data mining pada prediksi keberhasilan pemasaran produk layanan perbankan," *J. Techno Nusa Mandiri*, vol. 13, no. 1, pp. 60–65, 2016.
- [13] I. Harjanto, M. Amiruddin, and B. H. Kunaryo, "OPTIMASI MODEL DEEP LEARNING UNTUK RASPBERRY MENGGUNAKAN PRUNING DAN KUANTISASI," *J. Teknol. Inf. DAN Komun.*, vol. 15, no. 2, pp. 360–370, 2024.