

KLASIFIKASI DAMPAK DAN KONDISI PASIEN HEPATITIS MENGGUNAKAN METODE DECISION TREE

Luhur Pambudi¹, Sefrika²

^{1,2}Universitas Bina Sarana Informatika

¹cihuybusebg@gmail.com

Received: 08-03-2026

Revised: 10-04-2026

Approved: 20-04-2026

ABSTRAK

Penelitian ini bertujuan untuk mengklasifikasikan dampak dan kondisi pasien hepatitis menggunakan algoritma Decision Tree tipe C4.5 berdasarkan data klinis pasien. Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan teknik klasifikasi menggunakan data sekunder sebanyak 155 data pasien yang diperoleh dari platform Kaggle. Tahapan penelitian meliputi preprocessing data, seperti pembersihan data, penanganan missing value, serta encoding atribut kategorikal menjadi numerik. Selanjutnya, dilakukan pembagian data menggunakan beberapa skenario, yaitu 60:40, 70:30, 80:20, dan 90:10, dengan metode shuffled sampling. Proses pemodelan dilakukan menggunakan perangkat lunak RapidMiner dengan algoritma Decision Tree (C4.5), dan evaluasi model dilakukan menggunakan confusion matrix untuk mengukur nilai accuracy, precision, dan recall. Hasil penelitian menunjukkan bahwa skenario pembagian data 80:20 menghasilkan performa terbaik dengan nilai accuracy sebesar 80,65%, precision 50,00%, dan recall 66,67%. Simpulan dari penelitian ini adalah bahwa metode Decision Tree C4.5 mampu mengklasifikasikan kondisi pasien hepatitis dengan cukup baik berdasarkan atribut klinis yang digunakan, serta berpotensi menjadi sistem pendukung keputusan dalam membantu tenaga medis melakukan analisis dan pengambilan keputusan secara lebih cepat dan objektif.

Kata KUNCI: Hepatitis, Decision Tree, Klasifikasi, Machine Learning

PENDAHULUAN

Hepatitis merupakan penyakit yang menyerang organ hati (liver) dan dapat mengganggu berbagai fungsi vital tubuh, seperti metabolisme zat gizi, penyimpanan energi, produksi protein penting, serta proses detoksifikasi racun dari dalam tubuh (Kemenkes RI, 2023; World Health Organization, 2024). Peradangan hati ini dapat disebabkan oleh berbagai faktor, antara lain infeksi virus (hepatitis A, B, C, D, dan E), konsumsi alkohol berlebihan, gangguan autoimun, serta paparan zat kimia berbahaya dan obat-obatan dalam jangka panjang (Lavanchy, 2004; Shepard et al., 2006). Di antara berbagai jenis hepatitis, hepatitis B dan hepatitis C menjadi perhatian utama dalam dunia medis karena keduanya berpotensi berkembang menjadi penyakit kronis yang dapat menyebabkan komplikasi serius seperti sirosis hati dan karsinoma hepatoseluler (Dienstag, 2008; Alter, 2007). Infeksi kronis ini sering kali berlangsung tanpa gejala yang signifikan pada tahap awal, sehingga banyak pasien tidak menyadari bahwa dirinya telah terinfeksi hingga memasuki stadium lanjut (Hajarizadeh et al., 2013). Kondisi ini menyebabkan keterlambatan dalam diagnosis dan penanganan, yang pada akhirnya meningkatkan risiko kematian.

Menurut laporan World Health Organization (2024), lebih dari 350 juta orang di dunia hidup dengan hepatitis kronis, menjadikannya sebagai salah satu penyebab utama kematian akibat penyakit menular setelah pandemi COVID-19 (WHO, 2023; Polaris Observatory, 2018). Setiap tahunnya, jutaan orang meninggal akibat komplikasi yang disebabkan oleh hepatitis, terutama akibat sirosis dan kanker hati. Hal ini menunjukkan bahwa hepatitis masih menjadi masalah kesehatan global yang serius dan membutuhkan perhatian khusus dari berbagai pihak. Di Indonesia, prevalensi hepatitis juga masih tergolong tinggi. Berdasarkan data dari Kementerian Kesehatan Republik Indonesia, penyakit hepatitis termasuk dalam kategori penyakit menular yang menjadi prioritas penanganan nasional (Kemenkes RI, 2023). Tingginya angka kejadian ini

dipengaruhi oleh berbagai faktor, seperti rendahnya tingkat kesadaran masyarakat terhadap pola hidup sehat, kurangnya edukasi mengenai penularan penyakit, serta keterbatasan akses terhadap layanan kesehatan yang memadai (Sukartini et al., 2020).

Salah satu faktor utama yang menyebabkan tingginya angka penyebaran hepatitis adalah kurangnya pemahaman masyarakat mengenai cara penularan dan gejala awal penyakit. Hepatitis, khususnya tipe C, dapat menular melalui kontak langsung dengan darah yang terinfeksi, seperti penggunaan jarum suntik secara bergantian, transfusi darah yang tidak aman, serta penggunaan alat medis yang tidak steril (Shepard et al., 2006). Selain itu, gejala awal seperti kelelahan, mual, kehilangan nafsu makan, nyeri pada perut bagian kanan atas, dan perubahan warna urin sering kali tidak disadari sebagai indikasi penyakit hepatitis (Hajarizadeh et al., 2013). Kondisi tersebut menyebabkan banyak pasien datang ke fasilitas kesehatan dalam kondisi yang sudah cukup parah. Oleh karena itu, diperlukan suatu sistem yang mampu membantu dalam mendeteksi dan mengklasifikasikan kondisi pasien secara lebih dini dan akurat. Dalam beberapa tahun terakhir, perkembangan teknologi informasi, khususnya dalam bidang kecerdasan buatan (artificial intelligence) dan machine learning, telah membuka peluang besar dalam pengembangan sistem pendukung keputusan di bidang kesehatan (Esteva et al., 2019; Topol, 2019).

Salah satu metode machine learning yang широко digunakan dalam proses klasifikasi adalah Decision Tree. Metode ini memiliki keunggulan dalam hal interpretabilitas, karena hasil model dapat divisualisasikan dalam bentuk struktur pohon yang mudah dipahami (Quinlan, 1993). Algoritma Decision Tree, khususnya C4.5, bekerja dengan membagi data ke dalam beberapa atribut berdasarkan nilai information gain tertinggi, sehingga menghasilkan aturan keputusan yang sistematis dan logis (Han et al., 2012; Kotsiantis, 2007). Dalam konteks penelitian ini, metode Decision Tree digunakan untuk mengklasifikasikan kondisi pasien hepatitis berdasarkan data klinis yang tersedia, seperti usia, jenis kelamin, hasil pemeriksaan laboratorium, serta indikator kesehatan lainnya. Dengan memanfaatkan data tersebut, sistem dapat memprediksi tingkat keparahan atau kondisi penyakit pasien secara lebih cepat dan objektif dibandingkan dengan metode konvensional.

Penelitian sebelumnya menunjukkan bahwa penggunaan algoritma Decision Tree dalam bidang kesehatan memiliki tingkat akurasi yang cukup tinggi dan mampu memberikan hasil yang dapat diinterpretasikan dengan baik oleh tenaga medis (Kotsiantis, 2007; Han et al., 2012). Oleh karena itu, penerapan metode ini diharapkan dapat menjadi solusi dalam membantu proses diagnosis dan pengambilan keputusan klinis. Selain itu, pengembangan sistem berbasis Decision Tree juga memiliki potensi untuk diimplementasikan dalam aplikasi berbasis web atau mobile, sehingga dapat diakses secara luas oleh masyarakat. Hal ini dapat meningkatkan kesadaran masyarakat terhadap pentingnya deteksi dini penyakit hepatitis serta membantu mengurangi jumlah kasus yang tidak terdiagnosis (Topol, 2019). Dengan demikian, penelitian ini bertujuan untuk mengembangkan model klasifikasi kondisi pasien hepatitis menggunakan metode Decision Tree C4.5, serta mengevaluasi kinerja model dalam memberikan prediksi yang akurat.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik klasifikasi berbasis algoritma Decision Tree tipe C4.5 untuk menganalisis dampak dan kondisi pasien hepatitis. Data yang digunakan merupakan data sekunder yang diperoleh dari

platform Kaggle, yang berisi 155 data pasien dengan berbagai atribut klinis seperti usia (age), jenis kelamin (gender), penggunaan steroid, antivirals, fatigue, malaise, anorexia, liverBig, dan liverFirm, serta atribut target berupa kondisi pasien (live atau die). Tahap awal penelitian dimulai dengan proses pengumpulan data yang kemudian diolah menggunakan perangkat lunak RapidMiner. Selanjutnya dilakukan tahap preprocessing data yang meliputi pembersihan data (data cleaning) untuk menangani nilai kosong (missing value), menghapus data duplikat, serta memastikan konsistensi data. Atribut kategorikal kemudian dikonversi ke dalam bentuk numerik menggunakan teknik label encoding, seperti nilai “yes” menjadi 1 dan “no” menjadi 0, serta “live” menjadi 1 dan “die” menjadi 0 pada variabel target.

Tahap ini bertujuan agar data dapat diproses oleh algoritma Decision Tree secara optimal. Setelah preprocessing, dataset dibagi menjadi data latih dan data uji menggunakan beberapa skenario perbandingan, yaitu 60:40, 70:30, 80:20, dan 90:10. Pembagian data dilakukan dengan metode shuffled sampling untuk memastikan distribusi data yang acak dan seimbang. Berdasarkan hasil pengujian, skenario pembagian data 80:20 dipilih karena menghasilkan tingkat akurasi tertinggi sebesar 80,65%. Tahap berikutnya adalah proses pemodelan menggunakan algoritma Decision Tree (C4.5) yang diimplementasikan melalui RapidMiner dengan operator utama seperti Read Excel, Split Data, Decision Tree, Apply Model, dan Performance. Model dibangun menggunakan data latih dan kemudian diuji menggunakan data uji untuk mengukur performanya. Evaluasi model dilakukan dengan menggunakan metrik confusion matrix yang menghasilkan nilai accuracy, precision, dan recall. Hasil akhir dari penelitian ini divisualisasikan dalam bentuk pohon keputusan yang memudahkan interpretasi terhadap faktor-faktor yang mempengaruhi kondisi pasien hepatitis. Dengan metode ini, model yang dihasilkan diharapkan mampu memberikan prediksi yang cukup akurat serta menjadi alat bantu dalam pengambilan keputusan medis terkait kondisi pasien hepatitis.

HASIL PENELITIAN DAN PEMBAHASAN

Tabel 1.
Data Pasien Hepatitis Berdasarkan Atribut Klinis

Id	Age	Gender	Target	Steroid	Antivirals	Fatigue	Malaise	Anorexia	Liverbig	Liverfirm
1	30	Female	LIVE	Yes	No	No	No	No	Yes	No
2	50	Male	LIVE	Yes	No	Yes	No	No	Yes	No
3	78	Male	LIVE	No	No	Yes	No	No	No	No
4	31	Male	LIVE	?	Yes	No	No	No	No	No
5	34	Male	LIVE	No	No	No	No	No	No	No
6	34	Male	LIVE	No	No	No	No	No	No	No
7	51	Male	DIE	Yes	No	Yes	No	Yes	No	No
8	23	Male	LIVE	No	No	No	No	No	No	No
9	39	Male	LIVE	No	No	Yes	No	No	No	Yes
10	30	Male	LIVE	No	No	No	No	No	No	No
11	39	Male	LIVE	Yes	Yes	No	No	No	Yes	Yes
12	32	Male	LIVE	No	Yes	Yes	No	No	No	Yes
13	41	Male	LIVE	No	Yes	Yes	No	No	No	Yes
14	30	Male	LIVE	No	No	Yes	No	No	No	Yes
15	47	Male	LIVE	Yes	Yes	No	No	No	No	No
16	38	Male	LIVE	Yes	No	Yes	Yes	Yes	No	No

17	66	Male	LIVE	No	No	Yes	No	No	No	No
18	40	Male	LIVE	Yes	No	Yes	No	No	No	Yes
19	38	Male	LIVE	No	No	No	No	No	No	No
20	38	Male	LIVE	Yes	Yes	No	No	No	Yes	Yes
21	22	Female	LIVE	No	Yes	Yes	No	No	No	No

Pada penelitian ini, Data yang digunakan dalam penelitian ini merupakan data sekunder yang diperoleh dari situs Kaggle, yang berisi informasi klinis mengenai pasien penderita hepatitis. Dataset tersebut terdiri dari 155 data pasien yang masing-masing memiliki sejumlah atribut yang merepresentasikan kondisi kesehatan dan riwayat medis pasien. Berdasarkan tabel 1 penulis Atribut yang digunakan dalam penelitian ini meliputi: age (usia), gender (jenis kelamin), steroid (penggunaan steroid), antivirals (penggunaan antivirus), fatigue (kelelahan), malaise (rasa tidak enak badan), anorexia (kehilangan nafsu makan), liverBig (pembesaran hati), liverFirm (kekakuan hati). Target dari klasifikasi adalah atribut class, yang terdiri dari dua nilai yaitu "live" (pasien bertahan hidup) dan "die" (pasien meninggal dunia). Data ini digunakan untuk membangun model klasifikasi berbasis algoritma Decision Tree dalam memprediksi risiko kematian pasien hepatitis berdasarkan atribut-atribut klinis tersebut.

Pemrosesan Data (Preprocessing Data)

Pada Tahapan preprocessing dilakukan untuk menyiapkan data sebelum digunakan dalam pembangunan model klasifikasi dengan algoritma Decision Tree. Proses ini meliputi beberapa langkah utama sebagai berikut:

Pemeriksaan dan Pembersihan Data (Data Cleaning)

Dataset yang digunakan terdiri dari 155 data pasien dengan berbagai atribut klinis. Pemeriksaan awal dilakukan untuk mengetahui keberadaan nilai kosong (missing values), duplikasi data, dan data yang tidak konsisten. Data yang mengandung nilai kosong ditangani dengan metode imputasi sederhana, seperti pengisian berdasarkan modus atau penghapusan baris dengan nilai yang terlalu banyak hilang.

Penyandian Atribut Kategorikal (Encoding)

Atribut kategorikal seperti: gender (male/female), steroid (yes/no), antivirals (yes/no), fatigue, malaise, anorexia, liverBig, liverFirm, dan target (live/die), disandikan ke dalam bentuk numerik menggunakan teknik Label Encoding agar dapat diproses oleh algoritma Decision Tree. Contohnya, yes diubah menjadi 1, dan no menjadi 0. Untuk target, nilai live dikodekan sebagai 1 dan die sebagai 0.

Penyesuaian Format dan Normalisasi (Jika Diperlukan)

Semua data dikonversi ke format numerik agar kompatibel dengan pustaka pemodelan seperti scikit-learn. Karena algoritma Decision Tree tidak memerlukan normalisasi, langkah ini tidak dilakukan secara eksplisit, namun pemeriksaan outlier tetap dilakukan untuk memastikan tidak ada nilai yang sangat ekstrem.

Pembagian Data (Data Splitting)

Untuk mengevaluasi performa model secara menyeluruh, dilakukan pembagian data ke dalam beberapa skema sebagai berikut:

1. Skema 1: 90% data training dan 10% data testing
2. Skema 2: 80% data training dan 20% data testing

3. Skema 3: 70% data training dan 30% data testing
4. Skema 4: 60% data training dan 40% data testing

Pembagian dilakukan menggunakan metode shuffled sampling untuk memastikan distribusi acak dan merata dari kelas target. Evaluasi model kemudian dilakukan pada setiap skema untuk membandingkan accuracy, precision, dan recall.

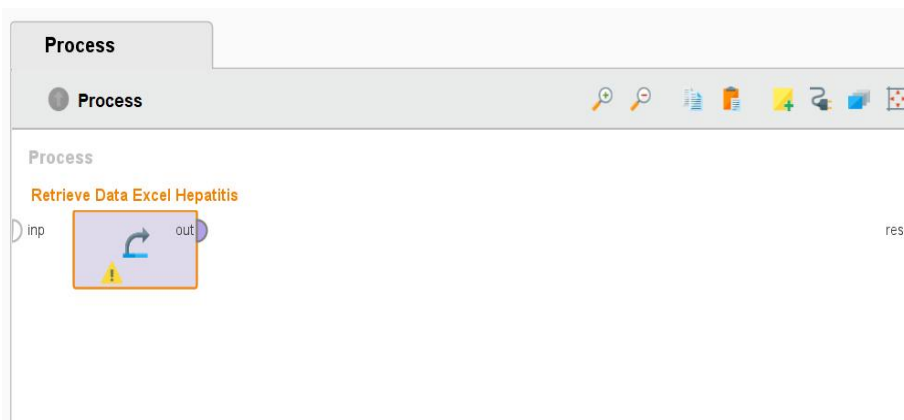
Pemodelan Algoritma Decision Tree

Pada tahap ini, metode klasifikasi diuji untuk menghitung nilai akurasi dengan menggunakan algoritma Decision Tree pada perangkat lunak Rapid Miner. Implementasi dilakukan dengan membangun proses pemodelan melalui tampilan process view dengan menyusun dan menghubungkan beberapa operator yang relevan. Operator utama yang digunakan meliputi Read Excel untuk memuat data, Split Data untuk membagi data menjadi data latih dan uji, serta Decision Tree sebagai algoritma klasifikasi utama. Evaluasi performa model dilakukan menggunakan operator Apply Model dan Performance untuk mengukur metrik seperti accuracy, precision, dan recall. Seluruh tahapan ini dirancang untuk memperoleh model klasifikasi yang optimal dalam mengklasifikasikan dampak dan kondisi pasien hepatitis secara akurat.

Read Excel

Tahap awal dalam proses ini adalah penyusunan model, sebagaimana ditunjukkan pada Gambar 1 di bawah ini.

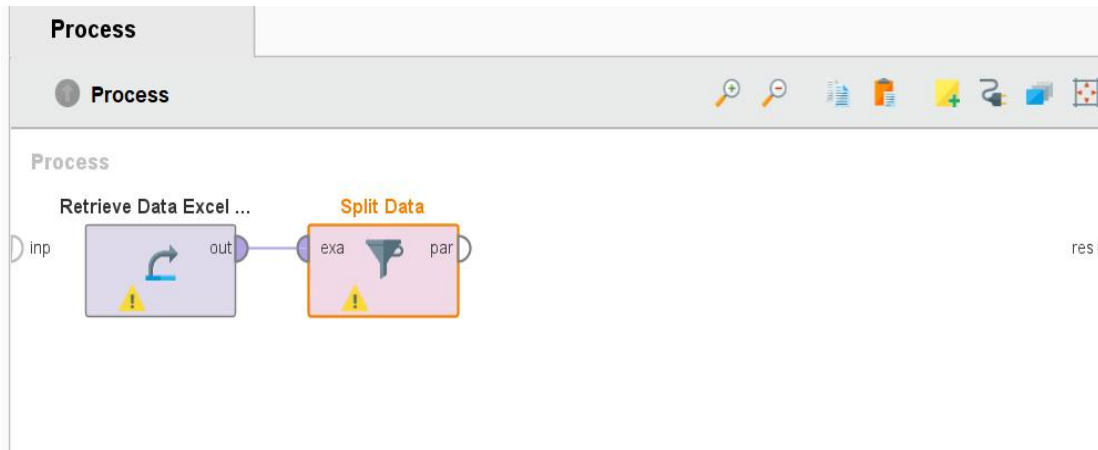
Gambar . 1 Read Exel



Seperti Gambar 1 file Excel yang berisi dataset dimuat ke sistem melalui operator Read Excel, yang berfungsi untuk membaca data dari file tersebut.

Split Data

Setelah data dibaca menggunakan operator Read Excel, proses dilanjutkan dengan menghubungkannya ke operator Split Data, sebagaimana diperlihatkan pada gambar di bawah ini.



Gambar III. 2 Split Data

Dalam tahap ini, penulis melakukan pengujian terhadap beberapa rasio pembagian data, yaitu 60:40, 70:30, 80:20, dan 90:10, guna memperoleh hasil akurasi yang maksimal. Pembagian data dilakukan dengan menggunakan metode shuffled sampling pada parameter Split Data. Rincian hasil akurasi dari masing-masing proporsi ditampilkan pada tabel berikut :

Tabel III.2 Proporsi Pembagian Data

Proporsi	Hasil Akurasi
60:40	64.52%
70:30	65.96%
80:20	80.65%
90:10	73.33%

Berdasarkan tabel di atas, penulis memilih proporsi pembagian data 80:20 karena memberikan nilai akurasi tertinggi.

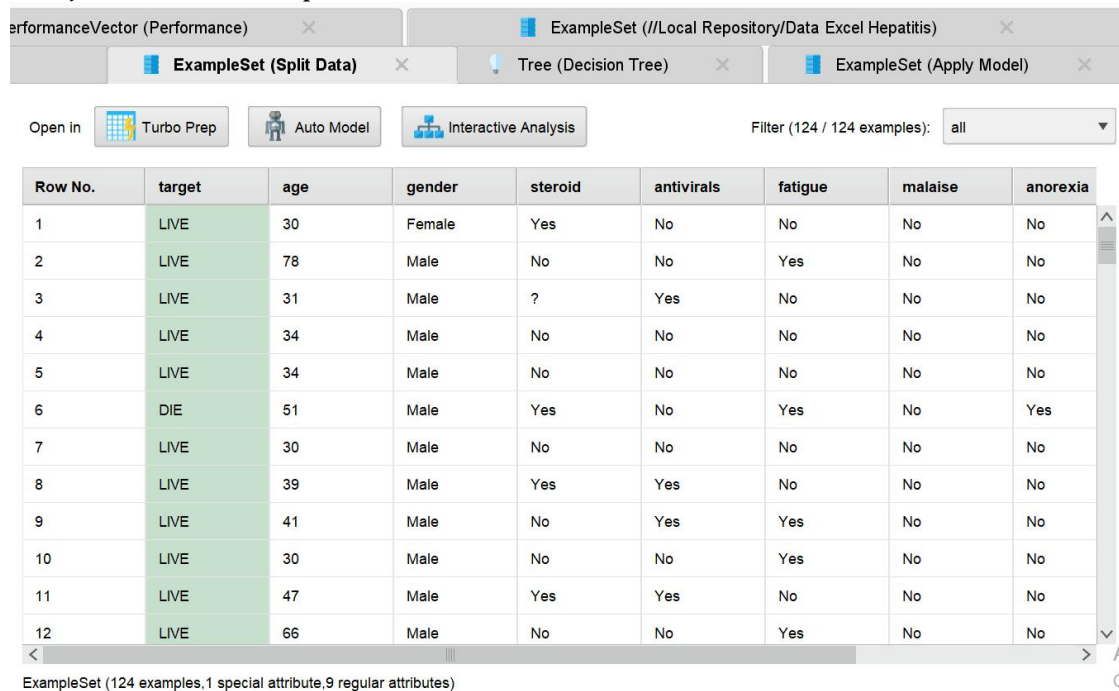
Hasil dari proses pembagian data menjadi data latih dan data uji menggunakan Rapid Miner ditunjukkan pada Gambar III.3 untuk data latih, Gambar III.4 untuk data uji, serta Tabel III.2 yang menampilkan atribut dari data uji tersebut.

Row No.	target	age	gender	steroid	antivirals	fatigue	malaise
1	LIVE	30	Female	Yes	No	No	No
2	LIVE	50	Male	Yes	No	Yes	No
3	LIVE	78	Male	No	No	Yes	No
4	LIVE	31	Male	?	Yes	No	No
5	LIVE	34	Male	No	No	No	No
6	LIVE	34	Male	No	No	No	No
7	DIE	51	Male	Yes	No	Yes	No
8	LIVE	23	Male	No	No	No	No
9	LIVE	39	Male	No	No	Yes	No
10	LIVE	30	Male	No	No	No	No
11	LIVE	39	Male	Yes	Yes	No	No

ExampleSet (155 examples, 1 special attribute, 9 regular attributes)

Gambar . 3 Tampilan Data Training

Melalui proses pada Rapid Miner, sebanyak 155 data digunakan sebagai data traning hasil pengolahan split data, sebagaimana ditunjukkan dalam Gambar III.3, yang menjadi dasar dalam pembuatan model.



Row No.	target	age	gender	steroid	antivirals	fatigue	malaise	anorexia
1	LIVE	30	Female	Yes	No	No	No	No
2	LIVE	78	Male	No	No	Yes	No	No
3	LIVE	31	Male	?	Yes	No	No	No
4	LIVE	34	Male	No	No	No	No	No
5	LIVE	34	Male	No	No	No	No	No
6	DIE	51	Male	Yes	No	Yes	No	Yes
7	LIVE	30	Male	No	No	No	No	No
8	LIVE	39	Male	Yes	Yes	No	No	No
9	LIVE	41	Male	No	Yes	Yes	No	No
10	LIVE	30	Male	No	No	Yes	No	No
11	LIVE	47	Male	Yes	Yes	No	No	No
12	LIVE	66	Male	No	No	Yes	No	No

ExampleSet (124 examples, 1 special attribute, 9 regular attributes)

Gambar 4 Data Testing

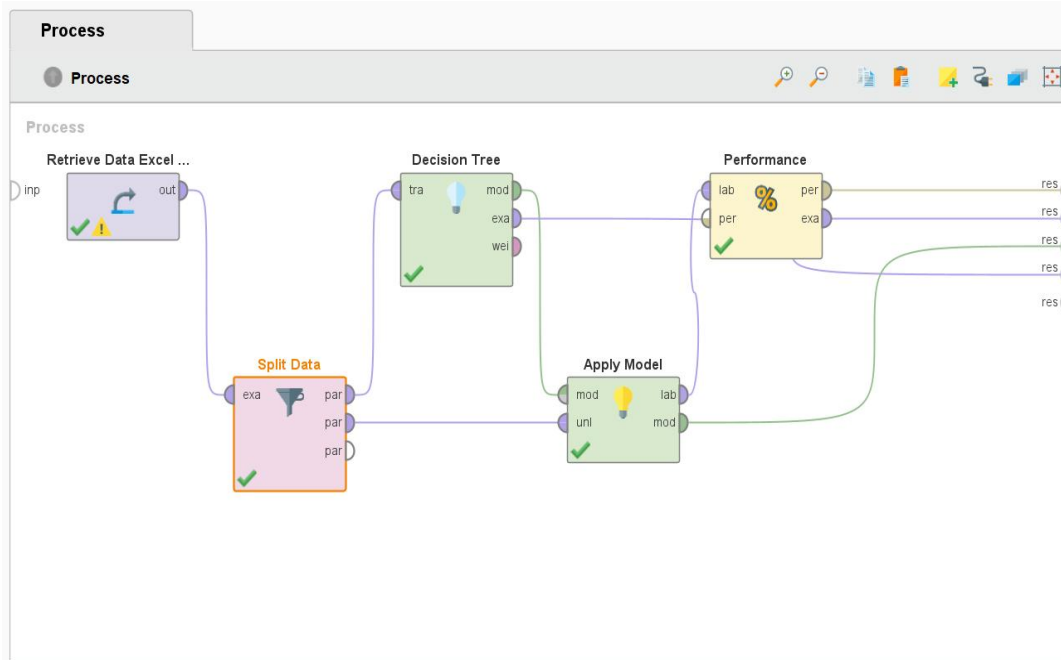
Mengacu pada Gambar III.4, proses pemisahan data yang dilakukan di Rapid Miner menghasilkan sebanyak 124 data sebagai data pengujian. Data ini dimanfaatkan untuk mengevaluasi performa model yang sebelumnya telah dibentuk menggunakan data pelatihan. Data uji tersebut mencakup sejumlah atribut penting, antara lain: usia (age), jenis kelamin (gender), status kondisi akhir pasien (target), penggunaan steroid dan antivirus, serta gejala klinis seperti kelelahan (fatigue), rasa tidak nyaman (malaise), hilangnya nafsu makan (anorexia), pembesaran hati (liverBig), dan kekakuan hati (liverFirm). Data ini berfungsi sebagai acuan untuk menilai sejauh mana model klasifikasi Decision Tree mampu memprediksi dampak dan kondisi pasien hepatitis secara akurat.

Tabel.3
 Atribut Data Testing

Atribut	Tipe Nilai	Deskripsi
Age	Numerik	Usia pasien saat data dikumpulkan.
Gender	Kategorikal (Male/Female)	Jenis kelamin pasien (Laki-laki/Perempuan).
Target	Kategorikal(Live/Die)	Status akhir pasien, apakah hidup atau meninggal.
Steroid	Kategorikal (Yes/No)	Penggunaan steroid sebagai bagian dari pengobatan.
Antivirals	Kategorikal (Yes/No)	Penggunaan antivirus dalam penanganan hepatitis.
Fatigue	Kategorikal (Yes/No)	Gejala kelelahan yang dialami pasien.
Malaise	Kategorikal (Yes/No)	Kondisi umum tidak nyaman atau lemas.
Anorexia	Kategorikal (Yes/No)	Kehilangan selera makan sebagai gejala.
LiverBig	Kategorikal (Yes/No)	Apakah hati pasien membesar.
LiverFirm	Kategorikal (Yes/No)	Apakah hati pasien terasa kaku saat diperiksa.

Berdasarkan Tabel .3, atribut-atribut yang digunakan dalam data testing mencakup informasi klinis penting dari pasien hepatitis.

Untuk melakukan pengujian terhadap data, setelah proses pembagian dataset menggunakan operator Split Data, langkah selanjutnya adalah menerapkan operator Decision tree, Apply Model, dan Performance, sebagaimana ditampilkan pada gambar berikut:



Gambar 5 Proses Decion Tree

Gambar 5 memperlihatkan alur proses dari masing-masing operator, di mana data latih dihubungkan dengan operator Decision Tree, sementara data uji diarahkan ke operator Apply Model dan Performance untuk keperluan pengujian.

Hasil dan Pembahasan

Hasil akhir dari proses ini ditunjukkan dalam bentuk Performance Vector, yang mencakup Confusion Matrix. Melalui matriks ini, sejumlah metrik evaluasi seperti accuracy, presisi, recall, dan pohon keputusan dapat dianalisis.

KESIMPULAN

Bahwa model yang dibangun mampu mengklasifikasikan kondisi pasien dengan cukup baik berdasarkan atribut klinis yang digunakan, seperti usia, jenis kelamin, penggunaan steroid dan antivirals, serta gejala seperti fatigue, malaise, anorexia, liverBig, dan liverFirm. Dari beberapa skenario pembagian data, proporsi 80:20 menghasilkan performa terbaik dengan nilai akurasi sebesar 80,65%, precision 50,00%, dan recall 66,67%, sehingga dipilih sebagai model optimal. Hasil ini menunjukkan bahwa algoritma Decision Tree efektif dalam mengidentifikasi faktor-faktor yang mempengaruhi kemungkinan pasien bertahan hidup (live) atau meninggal (die), serta menghasilkan model yang mudah diinterpretasikan melalui visualisasi pohon keputusan. Demikian, model ini berpotensi digunakan sebagai sistem pendukung keputusan untuk membantu tenaga medis dalam melakukan analisis awal dan pengambilan keputusan terkait kondisi pasien hepatitis secara lebih cepat dan objektif.

DAFTAR PUSTAKA

- Alter, M. J. (2007). Epidemiology of hepatitis C virus infection. *World Journal of Gastroenterology*, 13(17), 2436–2441. <https://doi.org/10.3748/wjg.v13.i17.2436>
- Dienstag, J. L. (2008). Hepatitis B virus infection. *New England Journal of Medicine*, 359(14), 1486–1500. <https://doi.org/10.1056/NEJMra0801644>

- Esteva, A., Robicquet, A., Ramsundar, B., et al. (2019). A guide to deep learning in healthcare. *Nature Medicine*, 25(1), 24–29. <https://doi.org/10.1038/s41591-018-0316-z>
- Hajarizadeh, B., Grebely, J., & Dore, G. J. (2013). Epidemiology and natural history of hepatitis C virus infection. *Nature Reviews Gastroenterology & Hepatology*, 10(9), 553–562. <https://doi.org/10.1038/nrgastro.2013.107>
- Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-61819-5>
- Kementerian Kesehatan Republik Indonesia. (2023). *Profil Kesehatan Indonesia 2023*. Kementerian Kesehatan RI. <https://www.kemkes.go.id>
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, 31(3), 249–268. <https://www.informatica.si/index.php/informatica/article/view/148>
- Lavanchy, D. (2004). Global burden of hepatitis C. *Journal of Viral Hepatitis*, 11(2), 97–107. <https://doi.org/10.1046/j.1365-2893.2003.00457.x>
- Polaris Observatory Collaborators. (2018). Global prevalence, treatment, and prevention of hepatitis B virus infection. *The Lancet Gastroenterology & Hepatology*, 3(6), 383–403. [https://doi.org/10.1016/S2468-1253\(18\)30056-6](https://doi.org/10.1016/S2468-1253(18)30056-6)
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers. <https://doi.org/10.1016/C2009-0-29260-4>
- Shepard, C. W., Finelli, L., & Alter, M. J. (2006). Global epidemiology of hepatitis C virus infection. *The Lancet Infectious Diseases*, 6(9), 558–567. [https://doi.org/10.1016/S1473-3099\(06\)70581-5](https://doi.org/10.1016/S1473-3099(06)70581-5)
- Sukartini, T., Arifin, H., & Kusumaningrum, T. (2020). Analisis penyakit menular di Indonesia. *Jurnal Keperawatan Indonesia*, 23(2), 85–92. <https://doi.org/10.7454/jki.v23i2.1100>
- Topol, E. (2019). *Deep medicine: How artificial intelligence can make healthcare human again*. Basic Books. <https://www.basicbooks.com>
- World Health Organization. (2023). *Global hepatitis report 2023*. WHO Press. <https://www.who.int>
- World Health Organization. (2024). *Hepatitis fact sheets and global updates*. WHO Press. <https://www.who.int>