



E-ISSN 3032-601X & P-ISSN 3032-7105

Vol. 3, No. 1, 2026

MISTER

**Journal of Multidisciplinary Inquiry in Science,
Technology and Educational Research**

**Jurnal Penelitian Multidisiplin dalam Ilmu
Pengetahuan, Teknologi dan Pendidikan**

**UNIVERSITAS SERAMBI MEKKAH
KOTA BANDA ACEH**

mister@serambimekkah.ac.id

Journal of Multidisciplinary Inquiry in Science Technology
and Educational Research

Journal of MISTER

Vol. 3, No. 1, 2026

Pages: 1215–1226

Penerapan Algoritma C4.5 untuk Prediksi Diabetes menggunakan Rapidminer

Amirul Khaq, Hasbi Firmansyah, Wahyu Asriyani

Universitas Pancasakti Tegal

Article in Journal of MISTER

Available at : <https://jurnal.serambimekkah.ac.id/index.php/mister>

DOI : <https://doi.org/10.32672/mister.v3i1.4050>

How to Cite this Article

APA : Khaq, A., Hasbi Firmansyah, & Wahyu Asriyani. (2025). Penerapan Algoritma C4.5 untuk Prediksi Diabetes menggunakan Rapidminer. *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research*, 3(1), 1215 – 1226. <https://doi.org/10.32672/mister.v3i1.4051>

Others Visit : <https://jurnal.serambimekkah.ac.id/index.php/mister>

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is a scholarly journal dedicated to the exploration and dissemination of innovative ideas, trends and research on the various topics include, but not limited to functional areas of Science, Technology, Education, Humanities, Economy, Art, Health and Medicine, Environment and Sustainability or Law and Ethics.

MISTER: *Journal of Multidisciplinary Inquiry in Science, Technology and Educational Research* is an open-access journal, and users are permitted to read, download, copy, search, or link to the full text of articles or use them for other lawful purposes. Articles on Journal of MISTER have been previewed and authenticated by the Authors before sending for publication. The Journal, Chief Editor, and the editorial board are not entitled or liable to either justify or responsible for inaccurate and misleading data if any. It is the sole responsibility of the Author concerned.



ISSN 3032-7105

ISSN 3032-601X



Penerapan Algoritma C4.5 untuk Prediksi Diabetes menggunakan Rapidminer

Amirul Khaq¹, Hasbi Firmansyah², Wahyu Asriyani³

Program Studi Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Pancasakti Tegal, Kota Tegal, Indonesia^{1,2}

Program Studi Pendidikan Bahasa dan Sastra Indonesia, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Pancasakti Tegal, Kota Tegal, Indonesia³

Email: amirulkhaq400@gmail.com¹, hasbifirmansyah@upstegal.ac.id², asriyani1409@gmail.com³

Diterima: 15-12-2025

| Disetujui: 25-12-2025

| Diterbitkan: 27-12-2025

ABSTRACT

Diabetes is a chronic disease whose prevalence continues to increase, so accurate and easy-to-interpret prediction methods are needed. The application of machine learning is one solution in analyzing medical data, especially for diabetes prediction. The Decision Tree algorithm was chosen because it is able to produce a transparent, easy-to-understand model, and is suitable for health data analysis that requires high interpretability. This study aims to apply the Decision Tree C4.5 algorithm to predict diabetes conditions using a diabetes dataset processed with RapidMiner software. The research stages include data preparation, determining attribute roles (set roles), dividing training and test data, building a classification model, implementing the model, and evaluating performance using the Performance module. The results showed that the Glucose attribute was the most dominant factor in the formation of the decision tree. The model produced an accuracy value of 80.00%, with very good performance in the non-diabetes class, but still has limitations in detecting all diabetes cases. In conclusion, the Decision Tree C4.5 algorithm is effective as an initial interpretive model in diabetes prediction, but still requires further development to increase the model's sensitivity.

Keywords: Decision Tree C4.5, RapidMiner, Diabetes Prediction, Machine Learning, Classification.

ABSTRAK

Diabetes merupakan penyakit kronis yang prevalensinya terus meningkat sehingga diperlukan metode prediksi yang akurat dan mudah diinterpretasikan. Penerapan machine learning menjadi salah satu solusi dalam menganalisis data medis, khususnya untuk prediksi diabetes. Algoritma Decision Tree dipilih karena mampu menghasilkan model yang transparan, mudah dipahami, serta sesuai untuk analisis data kesehatan yang menuntut interpretabilitas tinggi. Penelitian ini bertujuan menerapkan algoritma Decision Tree C4.5 untuk memprediksi kondisi diabetes menggunakan dataset diabetes yang diolah dengan perangkat lunak RapidMiner. Tahapan penelitian meliputi persiapan data, penentuan peran atribut (set role), pembagian data latih dan data uji, pembangunan model klasifikasi, penerapan model, serta evaluasi kinerja menggunakan modul Performance. Hasil penelitian menunjukkan bahwa atribut Glucose menjadi faktor paling dominan dalam pembentukan pohon keputusan. Model menghasilkan nilai akurasi sebesar 80,00%, dengan performa sangat baik pada kelas non-diabetes, namun masih memiliki keterbatasan dalam mendeteksi seluruh kasus diabetes. Kesimpulannya, algoritma Decision Tree C4.5 efektif digunakan sebagai model awal yang

interpretatif dalam prediksi diabetes, namun masih memerlukan pengembangan lebih lanjut untuk meningkatkan sensitivitas model.

Kata Kunci: Decision Tree C4.5, RapidMiner, Prediksi Diabetes, Machine Learning, Klasifikasi.

PENDAHULUAN

Diabetes merupakan salah satu penyakit kronis yang menjadi perhatian serius di bidang kesehatan global karena prevalensinya yang terus meningkat dari tahun ke tahun. Penyakit ini dapat menimbulkan berbagai komplikasi serius apabila tidak terdeteksi dan ditangani sejak dini. Oleh karena itu, diperlukan suatu pendekatan yang mampu membantu proses deteksi dan prediksi diabetes secara lebih akurat berdasarkan data medis yang tersedia. Seiring dengan berkembangnya teknologi informasi, pemanfaatan *machine learning* dalam bidang kesehatan semakin banyak digunakan untuk menganalisis data medis dan mendukung pengambilan keputusan. Teknik *machine learning* memungkinkan pengolahan data dalam jumlah besar untuk menemukan pola tersembunyi yang sulit diidentifikasi secara manual. Dalam konteks prediksi diabetes, penggunaan algoritma klasifikasi menjadi salah satu pendekatan yang efektif untuk menentukan kondisi pasien berdasarkan atribut-atribut kesehatan tertentu. Salah satu algoritma klasifikasi yang banyak digunakan adalah Decision Tree, khususnya algoritma C4.5. Algoritma ini memiliki keunggulan dalam menghasilkan model prediksi yang bersifat transparan dan mudah diinterpretasikan dalam bentuk struktur pohon keputusan (Afifuddin & Hakim, 2023). Karakteristik tersebut sangat penting dalam analisis data medis, karena hasil prediksi tidak hanya dituntut memiliki tingkat akurasi yang baik, tetapi juga harus dapat dijelaskan secara logis agar faktor-faktor yang memengaruhi keputusan dapat dipahami oleh pengguna (Devi et al., 2025). Selain itu, Decision Tree mampu menangani data numerik maupun kategorikal serta tidak memerlukan tahapan prapemrosesan data yang kompleks. (Hana, 2020)

Berdasarkan permasalahan tersebut, penelitian ini menerapkan algoritma Decision Tree C4.5 untuk memprediksi kondisi diabetes dengan memanfaatkan dataset diabetes yang diolah menggunakan perangkat lunak RapidMiner. Penelitian ini bertujuan untuk mengevaluasi kinerja model klasifikasi yang dibangun serta menganalisis kemampuan algoritma Decision Tree dalam mengidentifikasi pola yang berkaitan dengan kondisi diabetes (Hermansyah et al., 2025). Diharapkan hasil penelitian ini dapat memberikan kontribusi dalam pengembangan model prediksi diabetes yang bersifat interpretatif dan dapat digunakan sebagai dasar untuk penelitian selanjutnya di bidang analisis data kesehatan.

METODOLOGI PENELITIAN

Sumber dan Dasar Pemilihan Metode

Beberapa penelitian menunjukkan bahwa Decision Tree C4.5 relatif stabil dan memberikan akurasi tinggi pada dataset medis, termasuk prediksi diabetes. Menerapkan C4.5 untuk prediksi diabetes dan menemukan kombinasi *information gain ratio* memberikan hasil yang kompetitif dibandingkan metode lain seperti K-Nearest Neighbors dan Naïve Bayes (Afifuddin & Hakim, 2023). Penelitian lain menggunakan C4.5 yang dikombinasikan dengan teknik penyeimbangan data seperti SMOTE untuk meningkatkan performa kelas minoritas untuk kasus positif diabetes (Damayanti & Purwinarko, 2024). Data ini diperoleh melalui platform RapidMiner pada proses awal menggunakan operator Retrieve yang mengimpor dataset dari repositori internal RapidMiner. Seluruh atribut selain *Outcome* ditetapkan sebagai variabel prediktor (Nuryamin & Risyda, 2025).

Tahapan Prosedur Penelitian

1. Retrieve (Pengambilan Dataset)

Tahap *Retrieve* merupakan proses awal dalam penelitian, yaitu mengambil dataset diabetes yang telah disimpan di dalam repositori RapidMiner. Dataset ini berisi atribut-atribut medis yang digunakan sebagai variabel prediktor serta satu atribut target yang merepresentasikan kondisi diabetes. Pada tahap ini, data masih berada dalam kondisi mentah dan belum mengalami proses pemodelan. Fungsi utama dari *Retrieve* adalah memastikan bahwa dataset yang digunakan dapat diakses dengan benar dan siap untuk diproses pada tahap selanjutnya.

2. Split Data (Pembagian Data)

Tahap *Split Data* bertujuan untuk membagi dataset menjadi dua bagian, yaitu data latih (training data) dan data uji (testing data). Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur kinerja model terhadap data yang belum pernah dilihat sebelumnya. Pembagian ini penting untuk menghindari *overfitting* dan untuk memastikan bahwa performa model yang dihasilkan bersifat objektif dan dapat digeneralisasi.

3. Set Role (Penentuan Peran Atribut)

Pada tahap *Set Role*, dilakukan penentuan peran masing-masing atribut dalam dataset. Atribut *Outcome* ditetapkan sebagai label atau variabel target, sementara atribut lainnya berperan sebagai atribut prediktor. Tahap ini sangat krusial karena model klasifikasi tidak dapat dibangun tanpa penentuan label yang jelas. Dengan adanya *Set Role*, RapidMiner dapat membedakan antara variabel input dan output dalam proses pemodelan.

4. Decision Tree (Pembangunan Model Klasifikasi)

Tahap *Decision Tree* merupakan proses inti dalam penelitian, yaitu pembangunan model klasifikasi menggunakan algoritma C4.5. Pada tahap ini, model mempelajari pola hubungan antara atribut prediktor dan label berdasarkan data latih. Algoritma C4.5 membentuk struktur pohon keputusan yang terdiri dari node dan cabang, dimana setiap percabangan merepresentasikan aturan keputusan. Model ini bersifat transparan dan mudah diinterpretasikan, sehingga cocok untuk analisis data medis seperti prediksi diabetes.

5. Apply Model (Penerapan Model)

Tahap *Apply Model* digunakan untuk menerapkan model Decision Tree yang telah dibangun pada data uji. Proses ini menghasilkan keluaran berupa kelas hasil prediksi serta nilai *confidence* untuk masing-masing kelas. Dengan demikian, tahap ini berfungsi untuk menunjukkan bagaimana model melakukan klasifikasi terhadap data baru yang tidak digunakan dalam proses pelatihan.

6. Performance (Evaluasi Kinerja Model)

Tahap *Performance* merupakan tahap akhir dalam alur penelitian, yang bertujuan untuk mengevaluasi kinerja model klasifikasi. Evaluasi dilakukan menggunakan metrik performa seperti akurasi, precision, dan recall, yang dihitung berdasarkan hasil prediksi pada data uji. Hasil evaluasi ini digunakan untuk menilai sejauh mana model mampu memprediksi kondisi diabetes secara tepat dan menjadi dasar dalam pembahasan serta penarikan kesimpulan penelitian.



Gambar 1. Alur Proses RapidMiner (Process Diagram)

HASIL DAN PEMBAHASAN

Dataset Penelitian

Dataset yang digunakan dalam penelitian ini merupakan dataset diabetes yang diperoleh dari sumber dataset terbuka yang umum digunakan dalam penelitian prediksi penyakit. Dataset ini terdiri dari sejumlah data pasien dengan beberapa atribut medis, seperti kadar glukosa, tekanan darah, indeks massa tubuh, usia, dan riwayat kehamilan. Atribut *Outcome* digunakan sebagai variabel target yang menunjukkan kondisi diabetes atau non-diabetes, sedangkan atribut lainnya berperan sebagai variabel prediktor. Dataset ini diolah menggunakan perangkat lunak RapidMiner dan digunakan sebagai dasar dalam proses pembangunan dan evaluasi model klasifikasi Decision Tree C4.5.

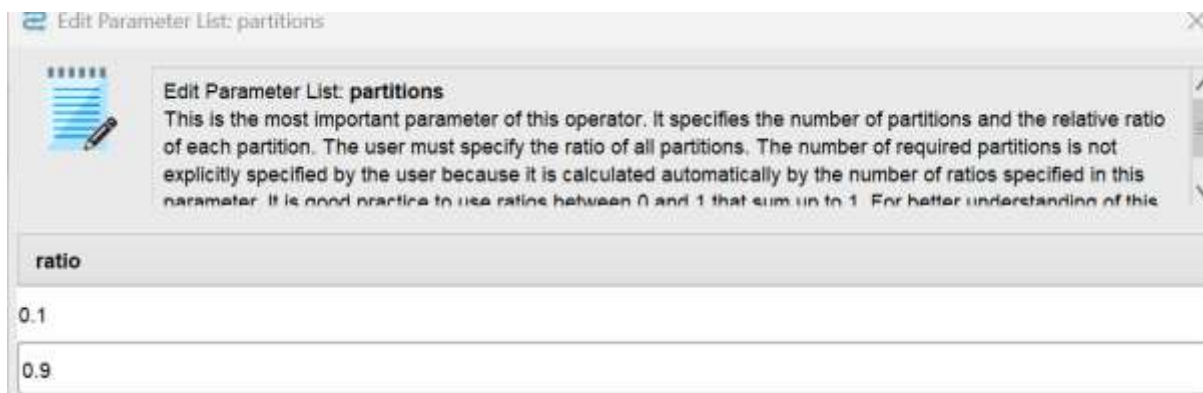
Tabel 1. Dataset Diabetes

No	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI		Age	outcome
1.	2	138	62	35	0	33.6		47	1
2.	0	84	82	31	125	38.2		23	0
3.	0	145	0	0	0	44.2		31	1
4.	0	135	68	42	250	42.3		24	1

5.	1	139	62	41	480	40.7		21	0
.....									
899	1	181	64	30	180	34.1		38	1
900	0	135	94	46	145	40.6		26	0
901	1	95	82	25	180	35		43	1

Pembagian Data

Pada penelitian ini, pembagian data dilakukan menggunakan operator Split Data pada RapidMiner dengan tujuan memisahkan dataset menjadi data latih dan data uji. Proses ini diperlukan agar model yang dibangun dapat dievaluasi menggunakan data yang belum pernah digunakan dalam tahap pelatihan. Dataset dibagi menjadi dua bagian dengan rasio 0,9 dan 0,1, di mana 90% data digunakan sebagai data latih dan 10% sebagai data uji. Data latih digunakan untuk membangun model klasifikasi menggunakan algoritma Decision Tree setelah terlebih dahulu dilakukan penentuan atribut target melalui operator Set Role. Selanjutnya, model yang telah terbentuk diterapkan pada data uji menggunakan operator Apply Model. Hasil prediksi kemudian dievaluasi menggunakan operator Performance untuk mengetahui kinerja model. Dengan adanya pembagian data ini, evaluasi yang dilakukan dapat menggambarkan kemampuan model dalam melakukan prediksi terhadap data baru secara lebih objektif dan valid.

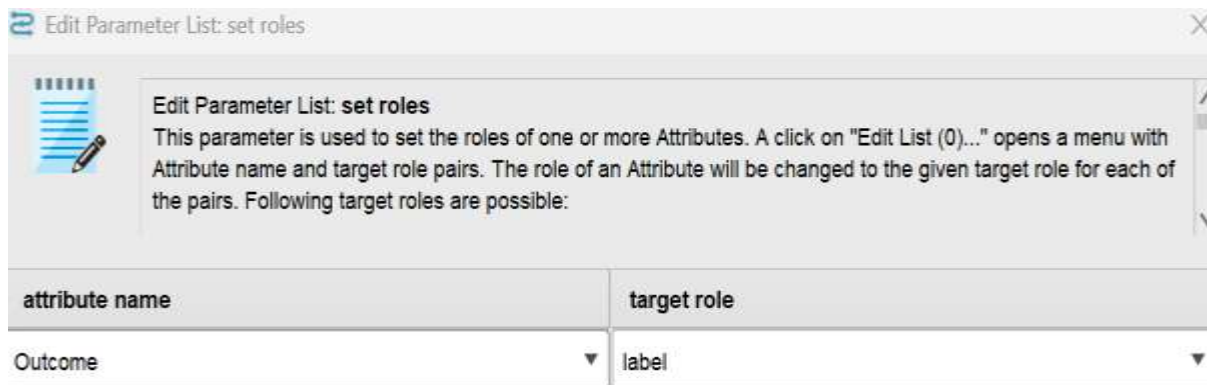


Gambar2. Proses pembagian dataset(Split Data)

Set Role(penentuan atribut)

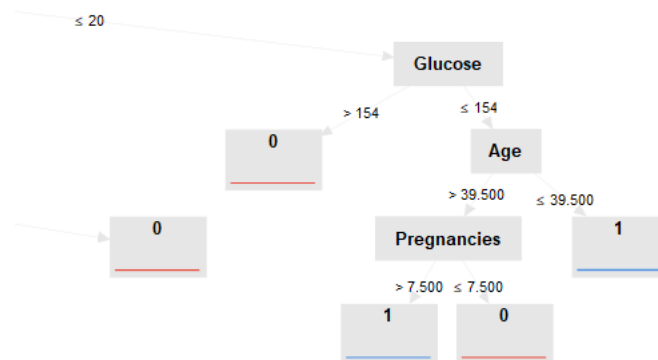
Tahap penentuan atribut (*set role*) merupakan salah satu tahapan penting dalam proses *data preparation* yang bertujuan untuk menetapkan peran masing-masing atribut dalam dataset sebelum dilakukan proses pemodelan. Penentuan peran atribut ini dilakukan menggunakan operator Set Role pada perangkat lunak RapidMiner Studio. Dalam penelitian ini, atribut Outcome ditetapkan sebagai variabel target (label) karena merepresentasikan hasil klasifikasi kondisi pasien, yaitu pasien yang terindikasi diabetes dan pasien yang tidak terindikasi diabetes. Sementara itu, atribut lainnya, yaitu Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, dan Age, ditetapkan sebagai variabel prediktor (fitur) yang digunakan sebagai input dalam proses pembentukan model klasifikasi. Penentuan atribut yang tepat sangat berpengaruh terhadap kinerja algoritma data mining. Kesalahan dalam menetapkan peran atribut dapat menyebabkan proses pembelajaran model tidak berjalan

secara optimal dan berpotensi menurunkan tingkat akurasi hasil klasifikasi. Oleh karena itu, tahap *set role* menjadi langkah yang krusial untuk memastikan bahwa struktur data telah sesuai dengan kebutuhan algoritma yang digunakan. Hasil dari tahap penentuan atribut menunjukkan bahwa dataset telah memiliki pembagian peran atribut yang jelas antara variabel target dan variabel prediktor, sehingga data siap digunakan pada tahap pemodelan dan evaluasi selanjutnya.



Gambar2. Proses Set Role(penentuan atribut)

Decision Tree (Pembangunan Model Klasifikasi)



Gambar3. Struktur pohon keputusan hasil pembentukan model klasifikasi

Pada gambar tersebut terlihat bahwa atribut Glucose menjadi node akar (root node). Hal ini menunjukkan bahwa atribut *Glucose* memiliki nilai *information gain* tertinggi sehingga dipilih sebagai pemisah utama dalam pembentukan pohon keputusan. Data kemudian dibagi berdasarkan kondisi $Glucose \leq 154$ dan $Glucose > 154$.

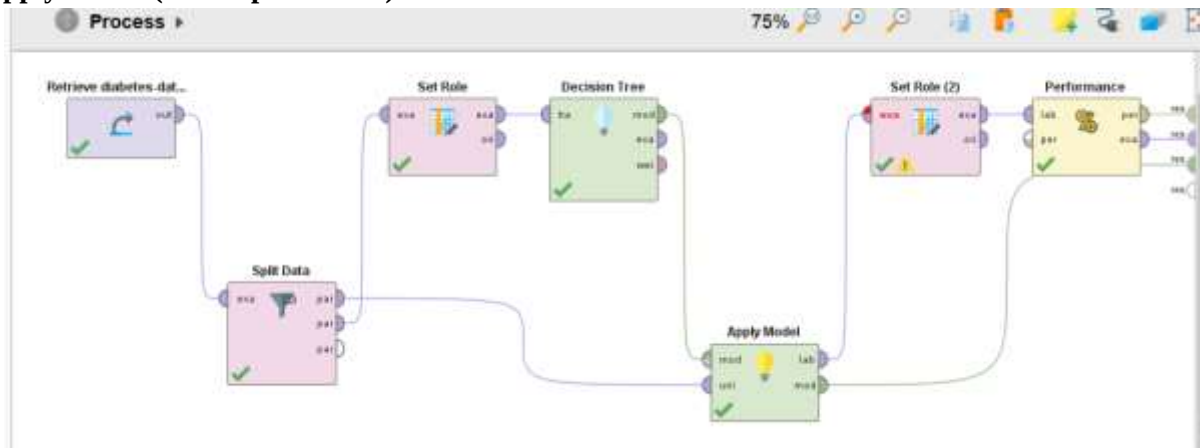
Untuk cabang Glucose > 154 , model langsung mengklasifikasikan data ke dalam kelas 0 (non-diabetes), yang ditunjukkan oleh node daun berlabel 0. Ini berarti, berdasarkan data latih, nilai Glucose yang melebihi batas tersebut secara dominan diasosiasikan dengan kelas tertentu sesuai pola yang dipelajari model.

Sementara itu, pada cabang Glucose ≤ 154 , proses klasifikasi dilanjutkan dengan atribut Age sebagai node berikutnya. Atribut Age membagi data menjadi dua kondisi, yaitu Age ≤ 39.5 dan Age > 39.5 . Pada kondisi Age ≤ 39.5 , data langsung diklasifikasikan ke dalam kelas 1 (diabetes), sedangkan pada Age > 39.5 , pemisahan dilanjutkan menggunakan atribut Pregnancies.

Atribut Pregnancies berfungsi sebagai node lanjutan yang membagi data berdasarkan jumlah kehamilan, dengan ambang batas ≤ 7.5 dan > 7.5 . Kedua kondisi tersebut menghasilkan node daun dengan kelas yang berbeda, yaitu kelas 0 dan kelas 1, yang merepresentasikan hasil akhir prediksi model.

Node daun (leaf node) yang berlabel 0 dan 1 menunjukkan keputusan akhir klasifikasi, di mana 0 merepresentasikan non-diabetes dan 1 merepresentasikan diabetes. Warna pada node juga membantu mengindikasikan kelas dominan hasil klasifikasi. Berdasarkan proses pelatihan menggunakan data latih, diperoleh model klasifikasi berbasis Decision Tree C4.5 sebagaimana ditampilkan pada Gambar 3, Pembangunan model klasifikasi dalam penelitian ini dilakukan menggunakan algoritma Decision Tree C4.5. Hasil pembentukan model ditampilkan pada Gambar 3, yang memperlihatkan struktur pohon keputusan yang dihasilkan berdasarkan data latih. Struktur pohon keputusan tersebut terdiri atas node akar, node internal, dan node daun yang merepresentasikan proses pengambilan keputusan secara hierarkis.

Apply Model (Penerapan Model)



Gambar4. Alur penerapan model klasifikasi menggunakan operator Apply Model pada RapidMiner

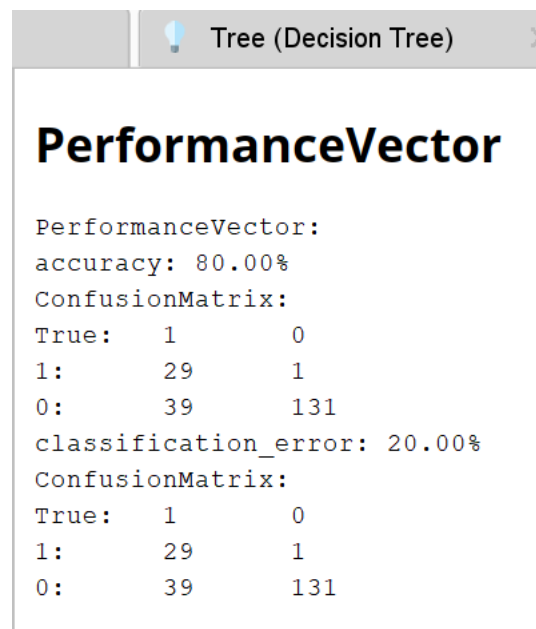
Tahap Apply Model merupakan fase penerapan model klasifikasi yang telah dibangun pada tahap sebelumnya ke dalam data uji (testing data) (Abrori & Fatah, 2025). Berdasarkan alur proses RapidMiner yang ditunjukkan pada Gambar 4, operator Apply Model menerima dua masukan utama, yaitu model Decision Tree hasil pelatihan dan data uji yang berasal dari proses Split Data. Tahap ini bertujuan untuk menguji kemampuan model dalam melakukan prediksi terhadap data yang tidak digunakan selama proses pembentukan model.

Pada proses ini, model Decision Tree yang telah terbentuk diaplikasikan pada setiap instance data uji dengan mengikuti aturan-aturan keputusan yang tersusun dalam struktur pohon. Setiap data uji akan ditelusuri mulai dari root node hingga mencapai leaf node berdasarkan nilai atribut yang dimilikinya. Hasil dari proses tersebut berupa label prediksi kelas (prediction) yang menunjukkan apakah suatu data diklasifikasikan ke dalam kelas diabetes (1) atau non-diabetes (0).

Selain menghasilkan label prediksi, operator Apply Model juga menghasilkan nilai confidence untuk masing-masing kelas. Nilai confidence ini merepresentasikan tingkat keyakinan model terhadap hasil prediksi yang diberikan. Semakin tinggi nilai confidence pada suatu kelas, semakin besar probabilitas data tersebut termasuk ke dalam kelas yang diprediksi. Informasi ini penting dalam konteks analisis medis karena memberikan gambaran tingkat kepastian keputusan yang dihasilkan oleh model klasifikasi (Dika et al., 2025).

Output dari tahap Apply Model berupa dataset yang telah diperkaya dengan atribut tambahan, yaitu label hasil prediksi dan nilai confidence. Dataset hasil penerapan model ini kemudian diteruskan ke tahap Performance untuk dilakukan evaluasi kinerja model secara kuantitatif. Dengan demikian, tahap Apply Model berperan sebagai penghubung antara proses pembentukan model dan proses evaluasi kinerja, serta menjadi dasar dalam menilai kemampuan generalisasi model Decision Tree terhadap data yang belum pernah dilihat sebelumnya

Performance (Evaluasi Kinerja Model)



Gambar5. Ringkasan kinerja model Decision Tree berdasarkan PerformanceVector pada RapidMiner.

Pada gambar 5 menampilkan PerformanceVector yang merupakan ringkasan kinerja model klasifikasi Decision Tree setelah diterapkan pada data uji. Berdasarkan hasil yang ditampilkan, model memperoleh nilai akurasi sebesar 80,00%, yang menunjukkan bahwa model mampu mengklasifikasikan

sebagian besar data uji dengan benar. Nilai ini mengindikasikan bahwa secara umum model memiliki performa yang cukup baik dalam membedakan antara kelas diabetes dan non-diabetes. (Informatika et al., 2024)

Selain itu, classification error sebesar 20,00% menunjukkan proporsi kesalahan prediksi yang masih terjadi. Nilai ini mencerminkan bahwa masih terdapat sejumlah data yang belum dapat diklasifikasikan secara tepat oleh model, sehingga diperlukan analisis lanjutan melalui Confusion Matrix untuk memahami pola kesalahan klasifikasi secara lebih rinci.

accuracy: 80.00%			
	true 1	true 0	class precision
Pred. 1	29	1	96.67%
Pred. 0	39	131	77.06%
Class recall	42.65%	99.24%	

Tabel 2. Hasil Akurasi Confusion Matrix hasil evaluasi kinerja model Decision Tree pada data uji.

Tabel 2 menyajikan Confusion Matrix yang menggambarkan distribusi hasil prediksi model terhadap label aktual. Pada kelas diabetes (1), model berhasil mengklasifikasikan 29 data sebagai True Positive, sementara terdapat 1 data yang salah diklasifikasikan sebagai non-diabetes (False Negative). Hasil ini menghasilkan nilai precision sebesar 96,67%, yang menunjukkan bahwa prediksi diabetes yang dihasilkan model memiliki tingkat ketepatan yang sangat tinggi.

Namun demikian, nilai recall pada kelas diabetes hanya sebesar 42,65%, yang mengindikasikan bahwa model belum mampu mendeteksi seluruh kasus diabetes yang ada dalam data uji. Hal ini menunjukkan adanya kecenderungan model untuk melewati sebagian kasus diabetes, yang menjadi kelemahan penting dalam konteks aplikasi medis.

Sebaliknya, pada kelas non-diabetes (0), model menunjukkan performa yang sangat baik dengan 131 data diklasifikasikan secara benar sebagai True Negative dan 39 data sebagai False Positive. Nilai recall sebesar 99,24% pada kelas non-diabetes menunjukkan bahwa hampir seluruh data non-diabetes berhasil dikenali dengan benar oleh model (Nasution et al., 2024).

KESIMPULAN

Penelitian ini bertujuan untuk menerapkan algoritma Decision Tree C4.5 dalam memprediksi kondisi diabetes dengan memanfaatkan dataset diabetes yang diolah menggunakan perangkat lunak RapidMiner. Berdasarkan latar belakang permasalahan yang menekankan pentingnya model prediksi yang bersifat transparan dan mudah diinterpretasikan pada bidang kesehatan, algoritma Decision Tree dipilih karena kemampuannya dalam menghasilkan aturan klasifikasi yang jelas dan dapat ditelusuri. Proses penelitian dilakukan secara sistematis melalui beberapa tahapan, yaitu pengambilan data, penentuan peran atribut (*set role*), pembagian dataset menjadi data latih dan data uji, pembangunan model klasifikasi menggunakan algoritma Decision Tree, penerapan model pada data uji (*apply model*), serta evaluasi kinerja model menggunakan *Performance*. Alur penelitian ini telah sesuai dengan tahapan standar pengembangan model klasifikasi pada *machine learning*. Hasil pembentukan model klasifikasi menunjukkan bahwa atribut

Glucose berperan sebagai faktor paling dominan dalam proses klasifikasi diabetes, diikuti oleh atribut Age dan Pregnancies, sebagaimana ditunjukkan pada struktur pohon keputusan yang dihasilkan. Struktur pohon keputusan tersebut memperlihatkan bahwa model mampu membentuk aturan *if-then* yang sederhana dan mudah dipahami, sehingga mendukung kebutuhan interpretabilitas dalam analisis data medis. Berdasarkan hasil evaluasi kinerja model, diperoleh nilai akurasi sebesar 80,00%, yang menunjukkan bahwa model Decision Tree memiliki kemampuan klasifikasi yang cukup baik secara keseluruhan. Model menunjukkan performa yang sangat baik dalam mengenali kelas non-diabetes, dengan nilai recall sebesar 99,24%, namun masih memiliki keterbatasan dalam mendeteksi seluruh kasus diabetes, yang tercermin dari nilai recall kelas diabetes sebesar 42,65%. Temuan ini mengindikasikan bahwa meskipun model memiliki tingkat ketepatan prediksi yang tinggi, sensitivitas terhadap kelas positif masih perlu ditingkatkan. Secara keseluruhan, penelitian ini membuktikan bahwa algoritma Decision Tree C4.5 dapat digunakan sebagai model awal yang efektif dan interpretatif dalam prediksi diabetes. Namun demikian, untuk meningkatkan kemampuan model dalam mendeteksi kasus diabetes secara lebih optimal, diperlukan pengembangan lanjutan, seperti penerapan teknik penyeimbangan data, optimasi parameter, atau penggabungan dengan metode klasifikasi lain. Hasil penelitian ini diharapkan dapat menjadi dasar bagi pengembangan sistem prediksi diabetes yang lebih akurat dan seimbang pada penelitian selanjutnya.

DAFTAR PUSTAKA

- Hana, F. M. (2020). *Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5* — Jurnal SISKOM-KB. Penelitian ini menerapkan C4.5 untuk klasifikasi diabetes dengan akurasi tinggi. [Jurnal Tanri Abeng University](#)
- Hermansyah, M. A., Saputra, D. A. A., Dwi Saputra, F., Maulana, M. R., & Arifin, M. (2025). *Penerapan Data Mining dengan Algoritma C4.5 untuk Mengidentifikasi Prediksi Penyakit Diabetes* — JRIIN : Jurnal Riset Informatika dan Inovasi. [Jurnal Mahasiswa](#)
- Devi, E. S. (2025). *Komparasi Algoritma Klasifikasi Naïve Bayes, Decision Tree (C4.5) dan SVM dalam Diagnosa Penyakit Diabetes* — JTINFO : Jurnal Teknik Informatika. [Rumah Jurnal UNISNU](#)
- Afifuddin, A., & Hakim, L. (2025). *Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4.5* — Krisnadana Journal. [Ejournal Sidya Nusa](#)
- Damayanti, D. R., & Purwinarko, A. (2024). Application of C4.5 algorithm using Synthetic Minority Oversampling Technique (SMOTE) for diabetes prediction. *Recursive Journal of Informatics*, 2(1), 18–27. <https://doi.org/10.15294/rji.v2i1.34939>
- Nuryamin, Y., & Risyda, F. (2025). Analisis prediksi penyakit diabetes menggunakan metode Decision Tree C4.5 dan Naïve Bayes. *JSI (Jurnal Sistem Informasi) Universitas Suryadarma*, 12(2), 234–242. <https://doi.org/10.35968/jsi.v12i2.1549>
- Nasution, M. A., Ulumuddin, Z. A., & Fitri, A. (2024). Analisis faktor risiko diabetes melitus menggunakan algoritma C4.5: Implementasi pada aplikasi Orange. *SAINTEK: Jurnal Sains, Teknologi & Komputer*, 1(3). <https://doi.org/10.56495/saintek.v1i3.1345>
- Sari, Z. D. R., Arvita, Y., & Jasmir, J. (2024). Penerapan data mining untuk prediksi penyakit diabetes menggunakan algoritma C4.5. *Jurnal Informatika dan Rekayasa Komputer (JAKAKOM)*, 4(1). <https://doi.org/10.33998/jakakom.2024.4.1.1624>

- Saputra, A. D. M., & Firmansyah, H. (2025). *Implementasi Algoritma Decision Tree untuk Memprediksi Kualitas Udara dan Polusi dengan RapidMiner*. Sudo Jurnal Teknik Informatika. <https://doi.org/10.56211/sudo.v4i3.1145>
- S. Abrori dan Z. Fatah, "Implementasi Metode Decission Tree Dalam Mengklasifikasi Depresi Menggunakan Rapidminer," **Journal of Students' Research in Computer Science (JSRCS)**, vol. 5, no. 2, Nov. 2024. DOI: <https://doi.org/10.31599/vgf7xb32>