Jurnal Teknik Informatika dan Teknologi Informasi
Volume. 5 Nomor. 3 Desember 2025
E-ISSN .: 2827-9387, P-ISSN .: 2827-9379, Hal 401-416
DOI: https://doi.org/10.55606/jutiti.v5i3.6338
Tersedia: https://journalshub.org/index.php/jutiti

Klasifikasi Indikator Kesehatan Diabetes Menggunakan Algoritma Random Forest

Haura Syahla[1*], Haris Izzudin[2], Fariz Aditya Pratama[3], Beni Rahmatullah[4], Ahmad Jurnaidi Wahidin[5], Ika Kurniawati[6]
[1-6] Teknik Informatika, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika, Indonesia
Email: hsyahla28@gmail.com [1], harissholeh1753@gmail.com [2], Farizadityapratama72@gmail.com [3], Beni.brh@bsi.ac.id [4], Ahmad.ajn@bsi.ac.id [5], ika.iki@nusamandiri.ac.id [6]

**Penulis Korespondensi: hsyahla28@gmail.com*

Abstract. *Diabetes continues to rise as a global health concern, highlighting the need for analytical methods that can assist in earlier and more accurate detection. This study aims to classify diabetes conditions using the Random Forest algorithm implemented through the Orange Data Mining platform. The dataset used contains various health-related attributes such as glucose levels, blood pressure, body mass index, age, and other clinical indicators associated with diabetes risk. Random Forest was selected due to its ability to produce stable models, handle large and complex datasets, and minimize overfitting by combining multiple decision trees. The research process includes data preprocessing, splitting the dataset into training and testing portions, building the Random Forest model, and evaluating its performance using metrics such as accuracy, precision, recall, F1-score, and confusion matrix. The results indicate that Random Forest delivers strong and consistent performance in classifying diabetes conditions based on the given health indicators. These findings suggest that employing data mining techniques especially Random Forest within Orange—can serve as a practical and reliable approach to support medical analysis and assist healthcare practitioners in achieving earlier and more accurate diabetes detection.*

Keywords*: Classification; Data Mining; Diabetes; Orange; Random Forest.*

Abstrak. Diabetes menjadi salah satu masalah kesehatan yang terus meningkat dan membutuhkan upaya deteksi yang lebih cepat serta akurat agar risiko komplikasi dapat diminimalkan. Penelitian ini bertujuan mengklasifikasikan kondisi diabetes menggunakan algoritma Random Forest melalui aplikasi Orange Data Mining sebagai alat bantu analisis. Dataset yang digunakan berisi berbagai informasi kesehatan, seperti kadar glukosa, tekanan darah, indeks massa tubuh, usia, dan beberapa parameter lain yang berpengaruh terhadap risiko munculnya diabetes. Algoritma Random Forest dipilih karena dikenal mampu menghasilkan model yang stabil, bekerja dengan baik pada data berukuran besar, serta efektif dalam mengurangi overfitting melalui proses penggabungan banyak pohon keputusan. Tahapan penelitian dilakukan mulai dari pembersihan dan persiapan data, pembagian data menjadi training dan testing, pembangunan model, hingga evaluasi performa menggunakan akurasi, presisi, recall, F1-score, serta confusion matrix. Berdasarkan hasil pengujian, model Random Forest menunjukkan kinerja yang kuat dan konsisten dalam memprediksi kondisi diabetes berdasarkan indikator kesehatan yang tersedia. Temuan ini menunjukkan bahwa pemanfaatan teknologi data mining, khususnya Random Forest pada platform Orange, dapat menjadi solusi praktis untuk mendukung proses analisis medis dan membantu tenaga kesehatan dalam mendeteksi diabetes secara lebih dini dan tepat sasaran.

Kata Kunci: Data Mining; Diabetes; Klasifikasi; Orange; Random Forest.

1. LATAR BELAKANG

Diabetes merupakan salah satu penyakit kronis yang jumlah penderitanya terus bertambah setiap tahun dan kini menjadi salah satu masalah kesehatan yang paling banyak mendapat perhatian (Sahgal, 2024). Penyakit ini dapat menimbulkan beragam komplikasi serius jika tidak terdeteksi sejak dini, sehingga proses identifikasi risiko sangat penting untuk dilakukan secara cepat dan tepat (Olina et al., 2024). Di tengah perkembangan teknologi, metode data mining dan pembelajaran mesin mulai banyak dimanfaatkan untuk membantu

Naskah Masuk: 21 Oktober 2025; Revisi: 15 November 2025; Diterima: 08 Desember 2025; Terbit: 10 Desember 2025

tenaga medis dalam menganalisis data kesehatan (Sari et al., 2024). Berbagai penelitian terdahulu telah menggunakan algoritma seperti Decision Tree, Logistic Regression, dan Support Vector Machine untuk memprediksi diabetes (Mandias & Manoppo, 2025). namun beberapa di antaranya masih memiliki kelemahan, terutama dalam menghadapi data yang kompleks, berjumlah besar, atau memiliki banyak variabel yang saling berkaitan (Homepage et al., 2024).

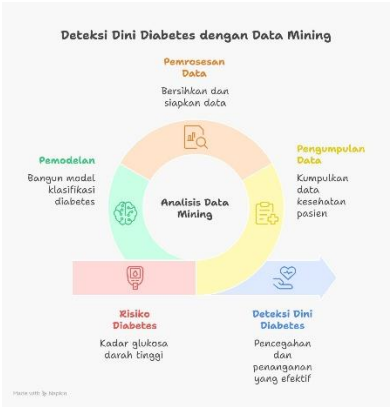
Kesenjangan tersebut menunjukkan perlunya metode yang mampu menghasilkan prediksi lebih stabil dan akurat. Random Forest menjadi salah satu algoritma yang menawarkan keunggulan tersebut melalui pendekatan ensemble yang menggabungkan banyak pohon keputusan, sehingga dapat mengurangi risiko overfitting dan meningkatkan performa model secara keseluruhan. Selain itu, belum banyak penelitian yang mengeksplorasi penggunaan platform Orange Data Mining secara mendalam, padahal aplikasi ini memiliki keunggulan berupa workflow visual yang memudahkan proses analisis tanpa memerlukan kemampuan pemrograman (Muharrom, 2023).

Dengan mempertimbangkan hal tersebut, penelitian ini dilakukan untuk menerapkan algoritma Random Forest pada dataset diabetes menggunakan Orange Data Mining. Tujuan utamanya adalah mengevaluasi sejauh mana model ini mampu mengklasifikasikan risiko diabetes dengan akurat. Hasil penelitian diharapkan dapat memberikan kontribusi terhadap pemanfaatan teknologi data dalam bidang kesehatan, khususnya dalam mendukung proses deteksi dini diabetes secara lebih efisien dan informatif.

2. KAJIAN TEORITIS

Diabetes dan Faktor Risiko

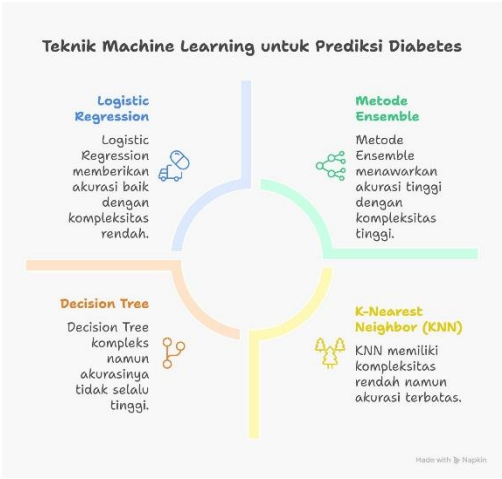
Diabetes merupakan gangguan metabolik yang ditandai oleh tingginya kadar glukosa dalam darah akibat gangguan pada produksi atau fungsi insulin (Putu et al., 2024). Berbagai faktor dapat memengaruhi risiko terjadinya diabetes, seperti kadar glukosa, tekanan darah, indeks massa tubuh, usia, riwayat keluarga, serta pola hidup (Mahmudah et al., 2025). Dalam penelitian data mining, informasi kesehatan biasanya dimanfaatkan untuk membantu memprediksi kemungkinan terjadinya penyakit, termasuk diabetes. Pemahaman mengenai faktor-faktor risiko ini menjadi dasar dalam pembangunan model klasifikasi untuk membantu proses deteksi dini.



Gambar 1. Alur Deteksi Data Diabetes.

Data Mining dan Machine Learning dalam Kesehatan

Data mining merupakan proses pengolahan dan analisis data untuk menemukan pola, informasi tersembunyi, atau hubungan antarvariabel (Sembiring & Sembiring, 2023). Dalam bidang kesehatan, data mining digunakan untuk mendukung keputusan klinis, memprediksi penyakit, hingga mengidentifikasi faktor risiko. Teknik machine learning, terutama model klasifikasi, sangat bermanfaat dalam mengolah data medis yang biasanya memiliki kompleksitas tinggi. Metode klasifikasi seperti Decision Tree, K-Nearest Neighbor, Logistic Regression, hingga teknik ensemble sering digunakan dalam penelitian prediksi diabetes (Huda et al., 2024).

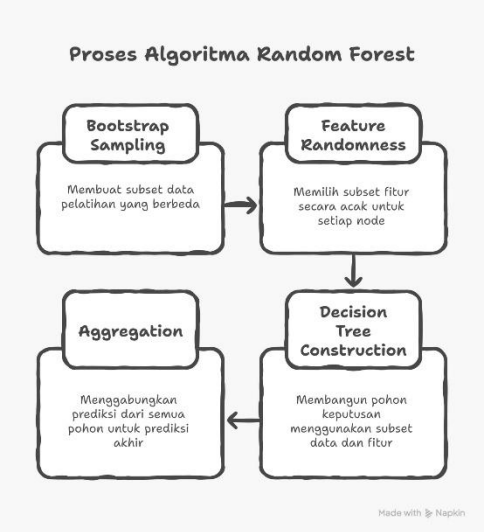


Gambar 2. Teknik Machine Learning untuk Prediksi Diabetes.

Algoritma Random Forest

Random Forest merupakan algoritma ensemble learning yang menggabungkan banyak pohon keputusan (decision tree) untuk menghasilkan prediksi yang lebih stabil dan akurat. Setiap pohon dilatih pada subset data yang berbeda dengan teknik bootstrap, sedangkan pemilihan fitur dilakukan secara acak pada setiap node (Iskandar et al., 2024). Kombinasi banyak pohon ini membuat Random Forest lebih tahan terhadap overfitting dibandingkan

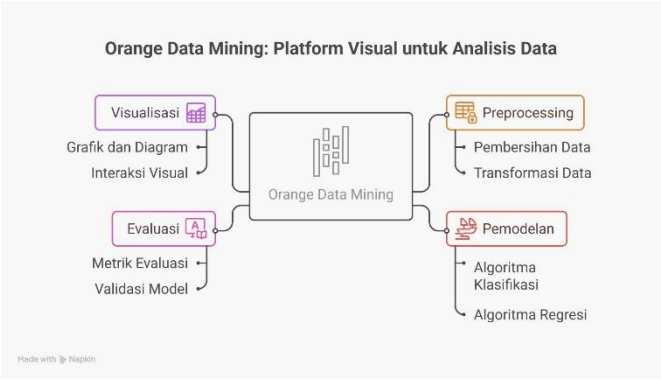
metode tunggal, serta mampu menangani data berdimensi tinggi (Inonu et al., 2025). Keunggulan inilah yang membuat Random Forest populer dalam penelitian klasifikasi medis, termasuk prediksi diabetes (Salsabil et al., 2024).



Gambar 3. Proses Algoritma Random Forest.

Platform Orange Data Mining

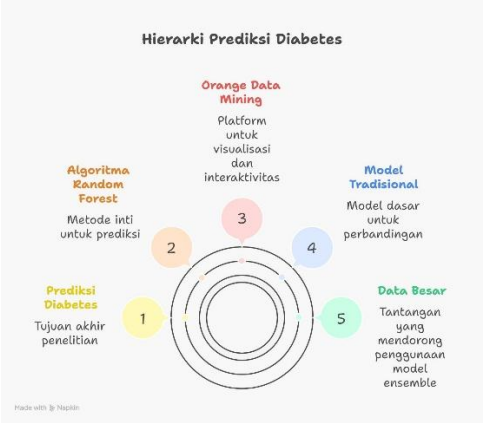
Orange Data Mining adalah perangkat lunak analisis data berbasis visual workflow yang menyediakan berbagai komponen untuk preprocessing, pemodelan, evaluasi, dan visualisasi data (Muharrom, 2023). Platform ini banyak digunakan dalam penelitian karena mudah dioperasikan tanpa memerlukan kemampuan pemrograman, namun tetap menawarkan algoritma dan fitur analisis yang lengkap (Pranadjaya et al., 2024). Penggunaan Orange memungkinkan peneliti membangun pipeline analisis secara cepat dan transparan, sehingga memudahkan proses eksperimen dan validasi model.



Gambar 4. Fitur Orange Data Mining.

Penelitian Terdahulu yang Relevan

Berbagai penelitian sebelumnya telah memanfaatkan machine learning untuk prediksi diabetes. Studi menggunakan Decision Tree dan Logistic Regression menunjukkan hasil yang cukup baik, namun sering mengalami keterbatasan akurasi pada data besar (Sutarman et al., 2024). Penelitian lain menunjukkan bahwa model ensemble seperti Random Forest dan Gradient Boosting cenderung memberikan akurasi lebih tinggi dan kestabilan yang lebih baik (Dwipa Jaya, 2023). Meskipun begitu, masih terbatas penelitian yang mengevaluasi performa Random Forest menggunakan platform Orange Data Mining secara komprehensif. Hal ini menunjukkan adanya ruang kontribusi dan landasan teoritis bagi penelitian ini untuk mengisi gap tersebut.



Gambar 5. Hierarki Prediksi Diabetes.

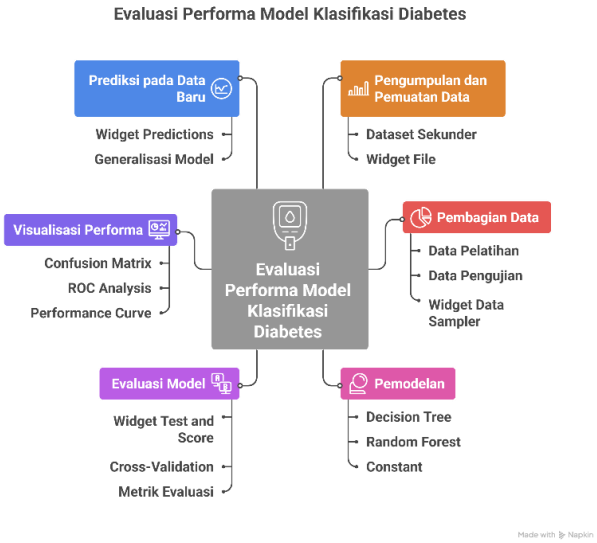
3. METODE PENELITIAN

Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan analisis data serta pemodelan klasifikasi berbasis machine learning. Tujuan utamanya adalah menilai kinerja algoritma Decision Tree dan Random Forest dalam mengelompokkan indikator kesehatan yang berkaitan dengan diabetes. Seluruh proses dilakukan menggunakan Orange Data Mining, yang memudahkan pengolahan data secara visual tanpa perlu melakukan pemrograman manual.

Proses penelitian diawali dengan memuat dataset melalui widget File, kemudian membaginya menjadi data pelatihan dan pengujian menggunakan Data Sampler agar evaluasi model berjalan objektif. Data tersebut selanjutnya dialirkan ke tiga algoritma, yaitu Tree, Random Forest, dan Constant. Tree dan Random Forest menjadi model utama yang diuji, sementara Constant berfungsi sebagai baseline untuk melihat perbedaan performa dibandingkan model yang lebih kompleks.

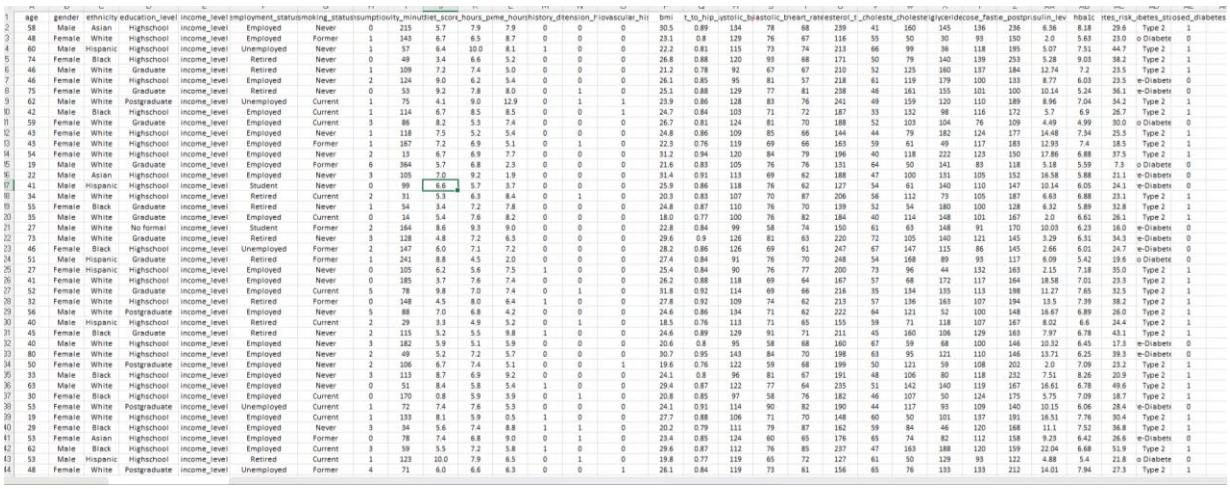
Setiap model dievaluasi menggunakan widget Test and Score melalui metode cross-validation, sehingga menghasilkan metrik seperti akurasi, precision, recall, F1-score, dan AUC (Jurnal & Teknologi, n.d.). Hasil evaluasi ditampilkan melalui Confusion Matrix, ROC Analysis, dan Performance Curve sebagai gambaran menyeluruh dari performa model. Pada tahap akhir, widget Predictions digunakan untuk menguji kemampuan model dalam melakukan prediksi terhadap data baru sehingga dapat dipastikan bahwa model tidak hanya bekerja baik pada data pelatihan, tetapi juga memiliki kemampuan generalisasi yang baik.



Gambar 6. Evaluasi Model Klasifikasi Diabetes.

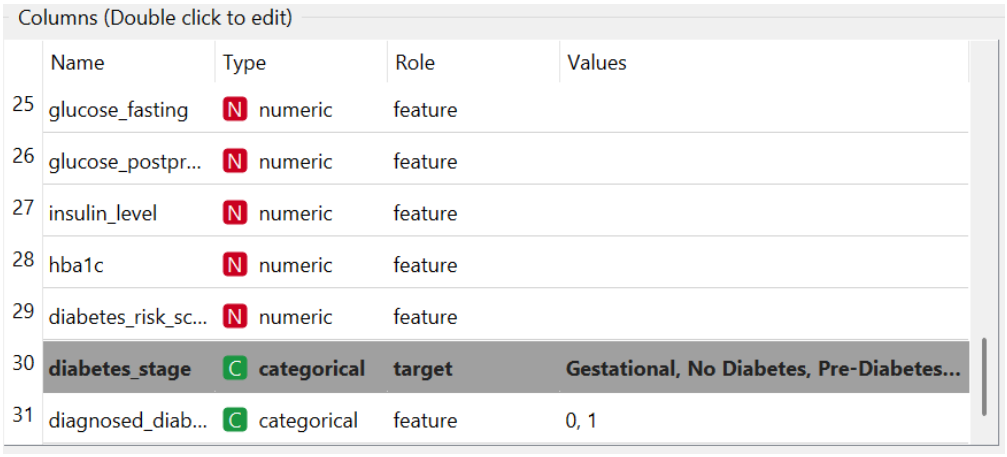
Pengumpulan Data

Dataset yang digunakan dalam penelitian ini berasal dari platform Kaggle, yang dikenal sebagai penyedia data terbuka untuk analisis data dan machine learning, berikut ini alamat dataset : https://www.kaggle.com/datasets/mohankrishnathalla/diabetes-health-indicators-dataset.



Gambar 7. Dataset Penelitian.

Data tersebut berisi berbagai indikator kesehatan yang berkaitan dengan risiko diabetes, seperti kadar glukosa darah, tekanan darah, usia, indeks massa tubuh (BMI), serta atribut pendukung lainnya. Dataset diunduh dalam format CSV dan diimpor ke aplikasi Orange Data Mining melalui widget File.


Columns (Double click to edit)

	Name	Type	Role	Values
25	glucose_fasting	numeric	feature	
26	glucose_postpr...	numeric	feature	
27	insulin_level	numeric	feature	
28	hba1c	numeric	feature	
29	diabetes_risk_sc...	numeric	feature	
30	**diabetes_stage**	**categorical**	**target**	**Gestational, No Diabetes, Pre-Diabetes...**
31	diagnosed_diab...	categorical	feature	0, 1

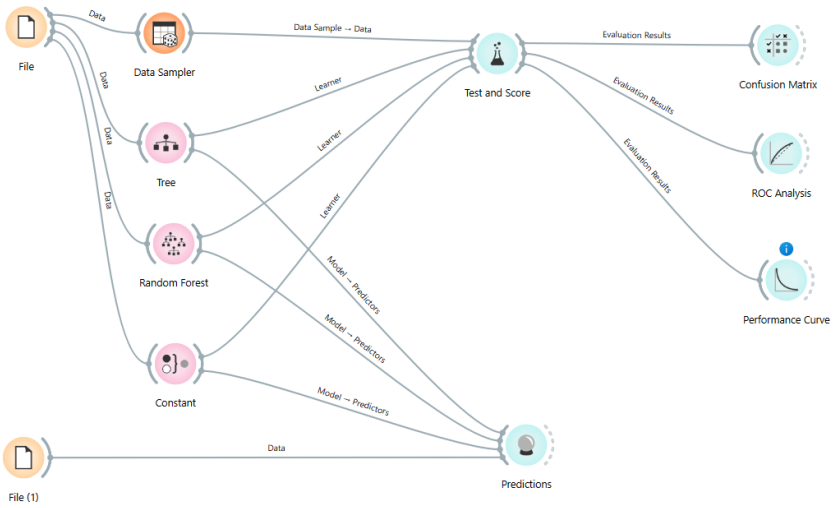
Gambar 8. Format Columns.

Setelah data berhasil dimuat, proses verifikasi dilakukan pada menu Columns untuk memastikan setiap atribut berada pada tipe dan peran yang benar. Pada tahap ini, variabel seperti glucose_fasting, glucose_postprandial, insulin_level, hba1c, dan diabetes_risk_score teridentifikasi sebagai fitur bertipe numerik, sedangkan beberapa atribut lain seperti diagnosed_diabetes terdeteksi sebagai fitur kategorikal. Variabel diabetes_stage ditetapkan sebagai target klasifikasi dengan nilai kategori seperti Gestational, No Diabetes, dan Pre-Diabetes. Pemeriksaan ini juga memastikan bahwa tidak ada kesalahan penempatan role serta seluruh tipe data sudah sesuai dengan kebutuhan pemodelan. Setelah seluruh atribut terdefinisi dengan benar, dataset kemudian siap digunakan untuk proses pelatihan, pengujian, dan evaluasi performa model klasifikasi dalam penelitian ini.

Analisa Data

Analisis data pada penelitian ini dilakukan menggunakan tiga algoritma klasifikasi, yaitu Decision Tree, Random Forest, dan Constant Model, yang seluruh prosesnya dibangun melalui workflow Orange Data Mining. Dataset terlebih dahulu dimuat dan dibagi menjadi data latih dan uji menggunakan Data Sampler, kemudian masing-masing model diterapkan dengan konfigurasi yang konsisten agar hasil evaluasi tetap objektif. Decision Tree menyusun struktur pohon berdasarkan atribut paling informatif, lalu performanya dinilai menggunakan Test & Score, Confusion Matrix, dan ROC untuk melihat kemampuan pemisahan kelas (Hozairi et al., 2021). Random Forest menjadi model utama karena memanfaatkan banyak pohon keputusan melalui teknik bootstrap, sehingga memberikan prediksi yang lebih stabil dan umumnya

menghasilkan akurasi paling tinggi (Hidayah & Rosadi, 2024). Sebagai pembanding, Constant Model digunakan sebagai baseline yang hanya menebak kelas mayoritas tanpa proses pembelajaran, sehingga skor akurasi maupun AUC relatif rendah dan mendekati prediksi acak. Seluruh hasil model kemudian divisualisasikan melalui Predictions untuk melihat kecenderungan keluaran pada data uji. Secara ringkas, analisis ini menggambarkan efektivitas masing-masing algoritma dalam mengklasifikasikan data diabetes dan menunjukkan keunggulan Random Forest sebagai model terbaik dalam workflow penelitian.



Gambar 9. Workflow Klasifikasi Diabetes.

4. HASIL DAN PEMBAHASAN

Pengolahan data pada penelitian ini dilakukan menggunakan workflow analisis berbasis machine learning pada aplikasi Orange Data Mining, yang melibatkan tahapan pemuatan data, pemisahan dataset, pembangunan model, hingga evaluasi performa menggunakan berbagai metrik. Tiga algoritma utama yang digunakan yaitu Decision Tree, Random Forest, dan Constant Model sebagai model pembanding (baseline). Seluruh proses analisis mengikuti alur kerja pada workflow, dimulai dari File → Data Sampler → (Decision Tree / Random Forest / Constant Model) → Test & Score → Confusion Matrix → ROC Analysis → Predictions. Tahapan ini memungkinkan proses pemodelan dilakukan secara sistematis serta memudahkan peneliti dalam membandingkan performa antar algoritma secara objektif.

Evaluation results for target (None, show average over classes)

Model	AUC	CA	F1	Prec	Recall	MCC
Random Forest	-1.692	0.996	0.994	0.992	0.996	0.992
Tree	-1.697	0.996	0.994	0.992	0.996	0.992
Constant	-0.843	0.598	0.448	0.358	0.598	0.000

Gambar 10. Hasil Tes & Score.

Hasil evaluasi menunjukkan bahwa Random Forest dan Decision Tree memiliki performa yang hampir sama dan sangat tinggi, dengan akurasi, F1-score, precision, recall, dan MCC berada pada rentang 0.992–0.996. Nilai ini mencerminkan kemampuan prediksi yang sangat stabil dan kuat, sementara AUC yang tampil negatif merupakan karakteristik visualisasi Orange pada kasus multi-kelas dan tidak menunjukkan penurunan performa. Di sisi lain, Constant Model memiliki kinerja yang jauh lebih rendah dengan akurasi 0.598 dan metrik lain yang juga rendah, menandakan bahwa model ini hanya mengandalkan tebakan kelas mayoritas. Dengan demikian, Random Forest dan Decision Tree terbukti jauh lebih unggul dibandingkan model baseline.

Pengolahan Menggunakan Decision Tree

Model Decision Tree dibangun menggunakan widget *Tree* untuk membentuk struktur pohon keputusan berdasarkan atribut yang paling informatif dalam membedakan kelas diabetes. Dataset yang telah dibagi melalui *Data Sampler* kemudian dilatih menggunakan algoritma ini, dan hasilnya dievaluasi melalui *Test & Score*. Hasil evaluasi ini selanjutnya diteruskan ke widget Confusion Matrix, ROC Analysis, dan Performance Curve untuk melihat kinerja model dari berbagai sisi.

Confusion Matrix

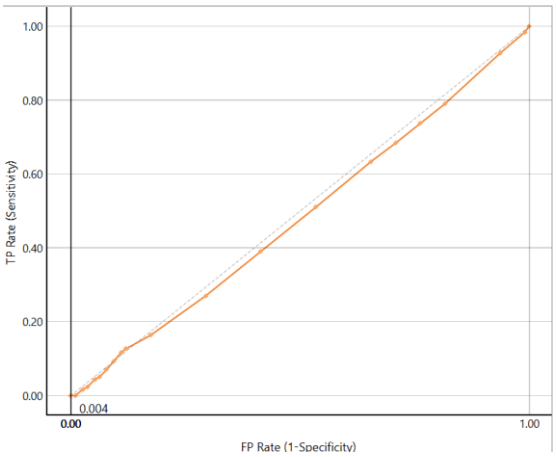
Predicted

Actual	Gestational	No Diabetes	Pre-Diabetes	Type 1	Type 2	Σ
Gestational	0	31	87	0	182	**300**
No Diabetes	0	8330	0	0	0	**8330**
Pre-Diabetes	0	0	33440	0	0	**33440**
Type 1	0	23	40	0	67	**130**
Type 2	0	0	0	0	62800	**62800**
Σ	**0**	**8384**	**33567**	**0**	**63049**	**105000**

Gambar 11. Hasil Confusion Matrix Decision Tree.

Hasil evaluasi ditinjau lewat Confusion Matrix, yang pada gambar menunjukkan bahwa model sangat tepat dalam mengenali kelas **Pre-Diabetes** (33.440 prediksi benar) dan **Type 2** (62.800 prediksi benar). Sebaliknya, kelas **Gestational** banyak salah diprediksi terutama sebagai **Type 2**, kelas **No Diabetes** hanya terklasifikasi benar pada 8.330 kasus, dan kelas **Type 1** masih tercampur dengan prediksi ke Pre-Diabetes dan Type 2. Analisis kemampuan pemisahan tiap kelas dilanjutkan melalui ROC Analysis, sementara widget Predictions digunakan untuk menampilkan hasil prediksi akhir pada data uji.

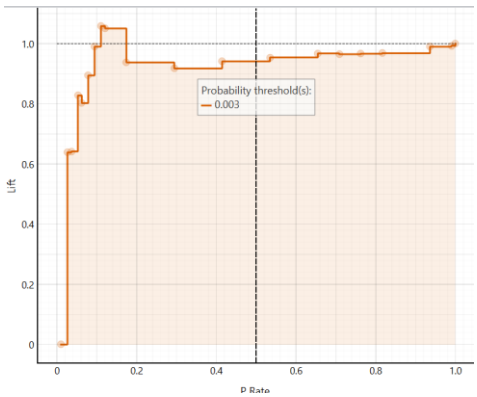
ROC Analysis



Gambar 12. Hasil ROC Analysis Decision Tree.

Hasil evaluasi ditinjau dengan ROC Analysis, yang menunjukan Kurva Decision Tree berada dekat dengan garis diagonal, menunjukkan performa yang lemah dalam mendeteksi kelas Gestational. Model ini kurang mampu memisahkan kelas karena data Gestational yang sedikit dan kecenderungan Tree untuk overfitting.

Performance Curve



Gambar 13. Hasil Performance Curve Decision Tree.

Hasil evalusai ditinjadengan Performance Curve, yang menunjukanKurva Decision Tree terlihat mendekati nilai lift = 1, yang berarti performanya hanya sedikit lebih baik daripada tebakan acak. Lift yang rendah dan fluktuatif menunjukkan bahwa Decision Tree tidak mampu

menangkap pola yang cukup kuat untuk mendeteksi kelas Gestational yang jumlahnya sangat kecil. Area under the curve 0.921 menunjukkan performa yang lemah dan hampir tidak memberikan keuntungan dibandingkan baseline.

Pengolahan Menggunakan Random Forest

Random Forest diterapkan sebagai model utama karena kemampuannya menggabungkan banyak pohon keputusan sehingga menghasilkan prediksi yang lebih stabil dan tahan terhadap overfitting (Putra, 2025). Evaluasi melalui Test & Score, Confusion Matrix, ROC Analysis, dan Performance Curve menunjukkan bahwa model ini bekerja sangat baik.

Confusion Matrix

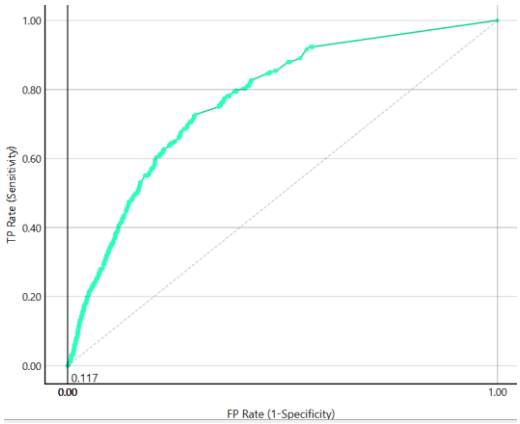
Predicted

Actual	Gestational	No Diabetes	Pre-Diabetes	Type 1	Type 2	Σ
Gestational	0	31	87	0	182	**300**
No Diabetes	0	8323	7	0	0	**8330**
Pre-Diabetes	0	0	33440	0	0	**33440**
Type 1	0	23	40	0	67	**130**
Type 2	0	0	0	0	62800	**62800**
Σ	**0**	**8377**	**33574**	**0**	**63049**	**105000**

Gambar 14. Hasil Confusion Matrix Random Forest.

Berdasarkan Confusion Matrix, Random Forest mampu mengklasifikasikan kelas besar seperti Pre-Diabetes (33.440 benar) dan Type 2 (62.800 benar) hampir tanpa kesalahan, serta mengidentifikasi No Diabetes secara konsisten dengan 8.323 prediksi akurat. Meski kelas minor seperti Gestational dan Type 1 masih mengalami sedikit salah klasifikasi, performanya tetap jauh lebih unggul dibanding Decision Tree maupun Constant Model. Secara keseluruhan, Random Forest terbukti paling efektif dalam menangkap pola kompleks pada dataset dan memberikan akurasi prediksi yang sangat tinggi.

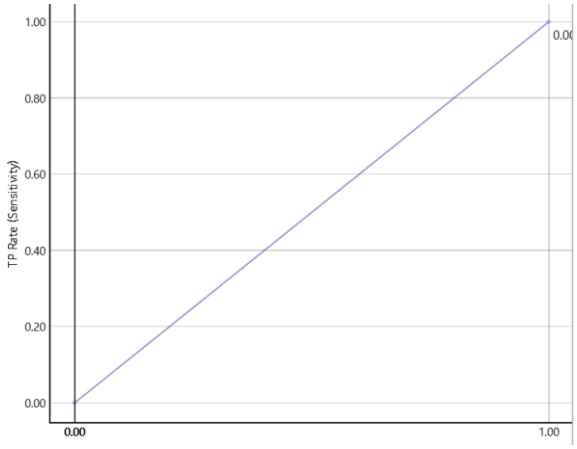
ROC Analysis



Gambar 15. ROC Analysis Random Forest.

Berdasarkan ROC Analysis Kurva ROC Random Forest tampak naik tajam dan berada jauh di atas garis diagonal, menunjukkan kemampuan yang jauh lebih baik dalam membedakan kelas Gestational. Model ini memiliki sensitivitas yang lebih stabil dan akurat meskipun jumlah data Gestational sangat kecil.

Performance Curve



Gambar 16. Performance Curve Random Forest.

Berdasarkan Performance Curve, Lift Curve Random Forest naik sangat tinggi di awal karena model ini mampu mengenali kelas Gestational yang jumlahnya sangat sedikit. Nilai lift kemudian menurun secara bertahap namun tetap berada di atas baseline, menunjukkan performa yang konsisten lebih baik. Pola ini sejalan dengan hasil Confusion Matrix yang menunjukkan bahwa Random Forest bekerja paling efektif.

Pengolahan Menggunakan Constant Model

Constant Model digunakan sebagai model baseline yang hanya memprediksi satu kelas, yaitu kelas mayoritas. Evaluasi melalui Test & Score, Confusion Matrix, ROC Analysis, dan Performance Curve.

Confusion Matrix

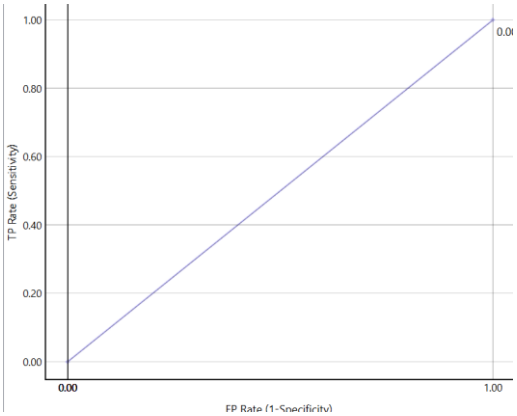
Predicted

Actual	Gestational	No Diabetes	Pre-Diabetes	Type 1	Type 2	Σ
Gestational	0	0	0	0	300	**300**
No Diabetes	0	0	0	0	8330	**8330**
Pre-Diabetes	0	0	0	0	33440	**33440**
Type 1	0	0	0	0	130	**130**
Type 2	0	0	0	0	62800	**62800**
Σ	**0**	**0**	**0**	**0**	**105000**	**105000**

Gambar 17. Hasil Confusion Matrix Constan Model.

Pada Confusion Matrix terlihat bahwa seluruh data—termasuk Gestational, No Diabetes, Pre-Diabetes, maupun Type 1—dipetakan menjadi Type 2, sehingga tidak ada klasifikasi benar pada keempat kelas lainnya. Kondisi ini menghasilkan akurasi yang rendah dan performa yang tidak informatif karena model tidak melakukan pembelajaran apa pun. Meski sangat sederhana, model ini tetap penting sebagai pembanding untuk menunjukkan bahwa Random Forest dan Decision Tree memberikan peningkatan performa yang signifikan dalam proses klasifikasi.

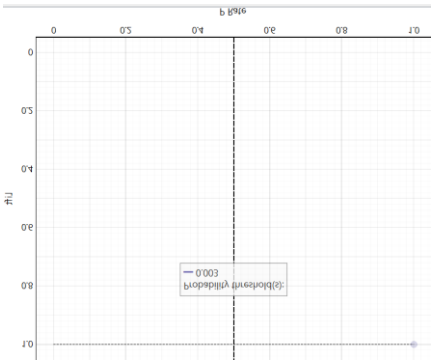
ROC Analysis



Gambar 18. ROC Analysis Constan Model.

Berdasarkan Kurva Constant tepat mengikuti garis diagonal, menandakan performa setara prediksi acak. Model ini berfungsi sebagai pembanding dasar untuk melihat apakah model lain bekerja lebih baik dari baseline.

Performance Curve

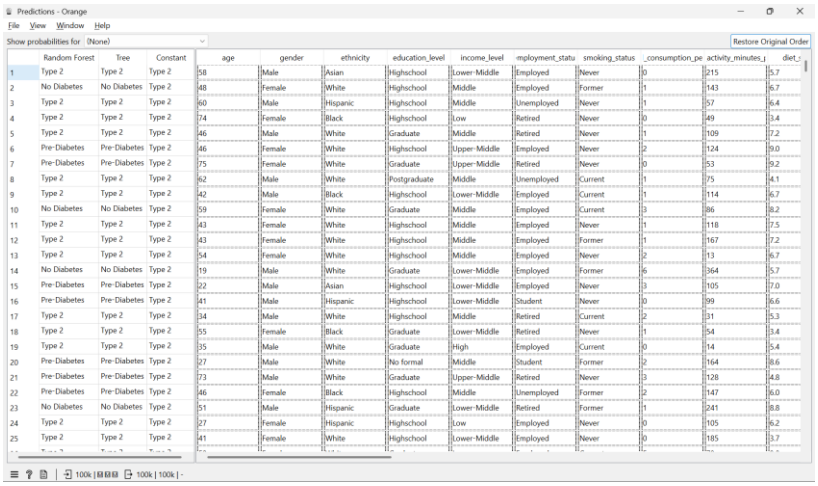


Gambar 19. Hasil Performance Curve Constan Model.

Model Constant membentuk garis datar pada lift = 1 sepanjang kurva. Hal ini wajar karena model ini tidak melakukan proses klasifikasi dan hanya menjadi pembanding dasar untuk melihat apakah model lain memberikan peningkatan performa.

Interpretasi Hasil Prediksi

Pada tahap akhir, widget Predictions digunakan untuk menampilkan keluaran dari setiap model terhadap data baru yang tidak digunakan selama proses pelatihan. Hasil prediksi memperlihatkan bagaimana ketiga algoritma Random Forest, Decision Tree, dan Constant Model, memberikan klasifikasi untuk setiap entri berdasarkan variabel input yang tersedia.



Gambar 20. Hasil Prediksi.

Dari keluaran tersebut terlihat bahwa Random Forest mampu menghasilkan prediksi yang lebih konsisten dan selaras dengan pola data, dibandingkan dengan Decision Tree yang cenderung mengalami variasi keputusan, serta Constant Model yang hanya berfungsi sebagai tolok ukur dasar tanpa kemampuan prediktif yang signifikan. Temuan ini menegaskan bahwa Random Forest merupakan metode yang paling andal dalam memetakan tahap diabetes pada dataset penelitian ini, berkat akurasi yang lebih tinggi serta performa generalisasi yang lebih baik terhadap data baru.

5. KESIMPULAN DAN SARAN

Penelitian ini bertujuan mengelompokkan tahapan diabetes menggunakan algoritma Random Forest, Decision Tree, dan Constant Model melalui platform Orange Data Mining. Dari seluruh proses evaluasi yang dilakukan, Random Forest terbukti memberikan performa paling kuat dan stabil dalam memprediksi kelas diabetes dibandingkan dua model lainnya. Decision Tree masih menunjukkan kinerja yang cukup baik namun cenderung kurang konsisten pada kelas dengan jumlah data kecil, sedangkan Constant Model hanya berfungsi sebagai acuan dasar karena tidak mempelajari pola data dan selalu memprediksi kelas mayoritas. Berdasarkan temuan tersebut, dapat disimpulkan bahwa Random Forest merupakan metode yang paling efektif untuk klasifikasi tahap diabetes pada dataset penelitian ini, meskipun generalisasinya

tetap perlu dilakukan dengan hati-hati mengingat penelitian hanya menggunakan satu sumber data dan satu lingkungan analisis.

Untuk pengembangan penelitian selanjutnya, disarankan agar dilakukan pengujian menggunakan algoritma lain seperti Gradient Boosting, XGBoost, atau model berbasis jaringan saraf guna mengetahui apakah terdapat pendekatan yang mampu mengungguli Random Forest. Peningkatan kualitas data melalui teknik penyeimbangan seperti SMOTE juga dapat dilakukan agar model lebih optimal dalam mengenali kelas minor yang jumlahnya terbatas, seperti Gestational dan Type 1. Selain itu, penggunaan dataset dari sumber yang lebih beragam maupun data klinis langsung dari institusi kesehatan akan membantu menghasilkan model yang lebih representatif dan relevan pada kondisi di lapangan..

UCAPAN TERIMA KASIH

Penulis mengucapkan apresiasi yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan selama penelitian ini berlangsung, baik berupa fasilitas, arahan, maupun kesempatan untuk menyelesaikan studi ini. Ketersediaan dataset dari Kaggle serta pemanfaatan platform Orange Data Mining sangat membantu dalam memperlancar proses analisis. Ucapan terima kasih juga penulis sampaikan kepada dosen pembimbing dan institusi pendidikan yang telah memberikan bimbingan akademik serta ruang bagi penulis untuk mengembangkan penelitian ini hingga tuntas.

DAFTAR REFERENSI

Dwipa Jaya, M. K. (2023). Perbandingan Random Forest, Decision Tree, Gradient Boosting, Logistic Regression untuk klasifikasi penyakit jantung. *JNATIA*, *2*(November), 1–5.

Hidayah, L., & Rosadi, M. I. (2024). Penerapan algoritma Random Forest untuk memprediksi jumlah santri baru. *Jurnal Informatika dan Teknik Elektro Terapan*, *12*(3S1). https://doi.org/10.23960/jitet.v12i3s1.5237

Homepage, J., Dwinnie, C., Dwynne, C., Islam, M. J., & Universitas Islam Negeri Sultan Syarif Kasim Riau. (2024). Comparison of machine learning algorithms in diabetes risk classification. *IJATIS: Indonesian Journal of Applied Technology and Innovation Science*, *1*(2), 54–60.

Hozairi, H., Anwari, A., & Alim, S. (2021). Implementasi Orange Data Mining untuk klasifikasi kelulusan mahasiswa dengan model K-Nearest Neighbor, Decision Tree serta Naive Bayes. *Network Engineering Research Operation*, *6*(2), 133. https://doi.org/10.21107/nero.v6i2.237

Huda, R. N., Fitriadi, R., & Wibowo, A. (2024). Optimization product recommendation using K-Means, Agglomerative Clustering and FP-Growth algorithm. *Jurnal Teknik Informatika (JUTIF)*, *5*(4), 953–960. https://doi.org/10.52436/1.jutif.2024.5.4.1901

Inonu, O. Y., Magda, K., & Amarudin, A. (2025). Analisis kinerja algoritma Random Forest dengan model machine learning pada dataset penyakit diabetes. *EXPERT: Jurnal Manajemen Sistem Informasi dan Teknologi*, *15*(1), 1. https://doi.org/10.36448/expert.v15i1.4312

Iskandar, R. F. N., Gutama, D. H., Wijaya, D. P., & Danianti, D. (2024). Klasifikasi menggunakan metode Random Forest untuk awal deteksi Diabetes Melitus Tipe 2. *Jurnal Teknik Industri Terintegrasi*, *7*(3), 1620–1626. https://doi.org/10.31004/jutin.v7i3.26916

Jurnal H., & Teknologi, F. (n.d.). Penerapan Orange Data Mining untuk pembelajaran sistem gambar hewan berbasis machine learning. *X(X)*, 42–49.

Mahmudah, M., Izza, N., Indrawati, L., Paramita, A., & Indriani, D. (2025). The contributing factors to the risk of diabetes mellitus among Indonesian urban workers. *Nurse Media Journal of Nursing*, *15*(1), 98–109. https://doi.org/10.14710/nmjn.v15i1.56916

Mandias, G. F., & Manoppo, I. J. (2025). Analisis komparatif algoritma klasifikasi untuk prediksi diabetes menggunakan pembelajaran mesin. *27*(1), 49–56.

Muharrom, M. (2023). Analisis penggunaan Orange Data Mining untuk prediksi harga USDT/BIDR Binance. *Bulletin of Information Technology (BIT)*, *4*(2), 178–184. https://doi.org/10.47065/bit.v4i2.654

Olina, Y. B., Aisah, S., Setyawati, D., Baidhowy, A. S., Nurkharistna, M., Jihad, A., & Arifianto, N. (2024). Meningkatkan kesadaran hidup sehat melalui skrining deteksi dini penyakit tidak menular di lingkungan Universitas Muhammadiyah Semarang. *4*(1).

Pranadjaya, E., Pangestu, E. S., Sereati, C. O., Octaviani, S., & Darmawan, M. (2024). Perbandingan algoritma machine learning menggunakan Orange Data Mining untuk klasifikasi jenis kendaraan pada sistem tilang digital. *Jurnal Elektro*, *17*(1), 41–47. https://doi.org/10.25170/jurnalelektro.v17i1.5429

Putra, H. (2025). Comparative study of logistic regression, Random Forest, and XGBoost for bank loan approval classification. *9*(5), 2822–2835.

Putu, N., Dharmayanti, D., Darmini, A. A. A. Y., Wayan, N., Dharmapatni, K., & Keperawatan. (2024). Pengetahuan penderita diabetes mellitus tentang pencegahan ulkus diabetik melalui penyuluhan. *Jurnal Abdimas ITEKES*, *3*(2), 70–74. https://ejournal.itekes-bali.ac.id/jai

Sahgal, A. (2024). Опыт аудита обеспечения качества и безопасности медицинской деятельности… *Вестник Росздравнадзора*, *4*(1), 9–15.

Salsabil, M., Lutvi, N., & Eviyanti, A. (2024). Implementasi data mining dalam melakukan prediksi penyakit diabetes menggunakan metode Random Forest dan XGBoost. *Jurnal Ilmiah Komputasi*, *23*(1), 51–58. https://doi.org/10.32409/jikstik.23.1.3507

Sari, Z. D. R., Jasmir, J., & Arvita, Y. (2024). Penerapan data mining untuk prediksi penyakit diabetes. *Jurnal Informatika dan Rekayasa Komputer (JAKAKOM)*, *4*(April), 827–834.

Sembiring, Y. A. B., & Sembiring, E. A. (2023). Implementasi data mining menggunakan algoritma Apriori dalam menentukan persediaan barang. *ADA Journal of Information System Research*, *1*(1), 1–8. https://doi.org/10.64366/adajisr.v1i1.7

Sutarman, S., Siringoringo, R., Arisandi, D., Kurniawan, E., & Nababan, E. B. (2024). Model klasifikasi dengan Logistic Regression dan Recursive Feature Elimination pada data tidak seimbang. *Jurnal Teknologi Informasi dan Ilmu Komputer*, *11*(4), 735–742. https://doi.org/10.25126/jtiik.1148198