

Komparasi Algoritma Decision Tree dan Random Forest Untuk Prediksi Penyakit Berdasarkan Gejala Pasien

Comparison of Decision Tree and Random Forest Algorithms for Disease Prediction Based on Patient Symptoms

Sunarto Taliki
Prodi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
atotaliki@gmail.com

Serwin
Prodi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
erwindnsn.ui@gmail.com

Haditsah Annur*
Prodi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
haditsah99@gmail.com*

Mohamad Rayhan A Ismail
Prodi Teknik Informatika
Universitas Ichsan Gorontalo
Gorontalo, Indonesia
mohamadrayhanis-
mail@gmail.com

Diterima : Maret 2026
Disetujui : Mei 2026
Dipublikasi : Juli 2026

Abstrak— Prinsip dasar kesehatan masyarakat merupakan pondasi penting dalam upaya melindungi dan meningkatkan kesehatan masyarakat, dengan perhatian utama pada pencegahan penyakit. Di era transformasi dibutuhkan suatu ilmu yang dapat melakukan prediksi, seperti penggunaan algoritma *decision tree* dan *random forest* yang memiliki kemampuan untuk menghasilkan prediksi akurat melalui proses pengolahan data dengan membentuk pohon keputusan. Penelitian ini menggunakan dataset publik tentang peduli sehat, yang dikumpulkan melalui <https://www.kaggle.com/datasets/krismonosadi/peduli-sehat-dataset> dengan jumlah data 4921 record. Masalah penelitian ini mengetahui algoritma manakah yang memberikan performa terbaik berdasarkan dataset sebanyak 130 gejala penyakit. Sehingga penelitian ini bertujuan mengeksplorasi teknik data mining menggunakan pendekatan perbandingan performa antara algoritma *decision tree* dan algoritma *random forest* dalam melakukan prediksi. Analisis data dilakukan menggunakan pengujian *split validation* 70:30 untuk mengetahui performa terbaik untuk prediksi penyakit. Hasil penelitian menunjukkan bahwa algoritma *decision tree* menghasilkan performa yaitu akurasi 94,17%, *precision* 95,04 % dan *recall* 94,55 %, sedangkan performa algoritma *random forest* yaitu akurasi 44,65%, *precision* 45,88% dan *recall* 46,67%. Sehingga studi ini membuktikan bahwa algoritma *decision tree* lebih efektif untuk dataset yang memiliki banyak fitur gejala namun berpola linear dibandingkan algoritma *random forest*, selain itu juga hasil yang didapatkan dapat memberikan kontribusi ilmiah, khususnya dalam bidang ilmu *machine learning*, dan sebagai rujukan pengembangan ilmu di bidang kesehatan.

Kata Kunci— *Prediksi; Penyakit; Perbandingan; Decision Tree; Random Forest*

Abstract— *Basic principles of public health form a crucial foundation in efforts to protect and improve public health, with a primary focus on disease prevention. In the era of transformation, a science capable of prediction is needed, such as the use of decision tree and random forest algorithms, which have the ability to generate accurate predictions through data processing by forming decision trees. This research utilizes a public dataset on "Peduli*

Sehat," collected from <https://www.kaggle.com/datasets/krismonosadi/peduli-sehat-dataset>, consisting of 4,921 records. The research problem is to determine which algorithm provides the best performance based on a dataset of 130 disease symptoms. Therefore, this study aims to explore data mining techniques using a comparative performance approach between the decision tree algorithm and the random forest algorithm in making predictions. Data analysis was conducted using a 70:30 validation split test to identify the best performance for disease prediction. The research results show that the decision tree algorithm achieved a performance of 94.17% accuracy, 95.04% precision, and 94.55% recall, while the random forest algorithm performance was 44.65% accuracy, 45.88% precision, and 46.67% recall. Therefore, this study proves that the decision tree algorithm is more effective for datasets with many symptom features but linear patterns compared to the random forest algorithm. Additionally, the results obtained can provide scientific contributions, particularly in the field of machine learning, and serve as a reference for scientific development in the health sector.

Keywords—*Prediction; Disease; Comparison; Decision Tree; Random Forest*

I. PENDAHULUAN

Kesehatan tubuh manusia merupakan sesuatu yang penting dalam melakukan aktifitas sehari-hari. Virus, bakteri atau masalah jaringan di dalam tubuh manusia bisa menyebabkan penyakit. Selain itu, penyakit juga bisa terjadi karena gaya hidup yang tidak sehat [1]. Sakit merupakan kondisi dimana tubuh berada dikondisi tidak normal yang disebabkan oleh unsur yang berbeda berasal dari dalam atau luar tubuh [2]. Penyakit merupakan salah satu yang menjadi masalah yang dialami banyak manusia, baik penyakit yang berkategori ringan maupun penyakit yang berkategori kronis, sehingga masyarakat perlu menjaga kondisi kesehatan dengan benar. Selain itu juga penyakit ditandai dengan gejala-gejala tertentu [3]. Prinsip-prinsip dasar kesehatan masyarakat menjadi dasar utama dalam upaya melindungi dan meningkatkan kesehatan masyarakat, dengan fokus pada

pencegahan penyebaran penyakit, mendorong pola hidup sehat, mewujudkan keadilan sosial, menggunakan pendekatan yang didasarkan pada bukti, serta bekerja sama antar berbagai sektor agar dapat menciptakan masyarakat yang lebih sehat dan lebih baik di masa depan [4]. Di era transformasi digital dibutuhkan suatu ilmu yang dapat melakukan prediksi yang memberikan dukungan untuk mengetahui penyakit yang dialami berdasarkan gejala yang dirasakan pasien. Prediksi merupakan proses perkiraan secara sistematis tentang kejadian masa depan yang dapat diprediksi berdasarkan informasi masa lalu agar kesalahan bisa diminimalkan [5].

Data mining adalah cara dan alat khusus yang digunakan untuk menganalisa data, mencari pola yang tersembunyi, serta menghasilkan informasi yang membantu dalam mengambil keputusan dan memahami data [6]. Data mining sangat berkaitan dengan analisis data serta penggunaan perangkat lunak untuk mencari pola-pola yang tersembunyi dalam sebuah kumpulan data, dua algoritma data mining yang dapat digunakan untuk prediksi adalah algoritma *decision tree* dan algoritma *random forest* yang merupakan algoritma pohon keputusan [7]. Algoritma *decision tree* merupakan algoritma yang mampu memodelkan hubungan kompleks antar variabel *input* dan *output* dengan cara yang mudah dipahami [8]. Selain itu juga, algoritma *decision tree* juga bisa bekerja dengan data yang memiliki atribut berupa angka atau kategori, serta mampu menampilkan pohon keputusan dalam bentuk visual yang mudah dimengerti [9]. Algoritma *random forest* merupakan salah satu algoritma *machine learning* yang memiliki kemampuan untuk menghasilkan prediksi yang akurat melalui proses pengolahan data dengan membentuk beberapa pohon keputusan [10], selain itu juga algoritma *random forest* dapat mengelola data berdimensi tinggi dan kemampuannya menilai pentingnya fitur menjadikan pilihan ideal untuk melakukan prediksi [11].

Masalah penelitian ini adalah algoritma manakah yang memberikan performa terbaik dengan menggunakan dataset sebanyak 130 gejala. Maka dari itu kedua algoritma tersebut akan digunakan sebagai perbandingan mendapatkan hasil performa terbaik dalam prediksi penyakit berdasarkan gejala pasien. Tujuan penelitian ini mengeksplorasi teknik data mining dengan membandingkan performa algoritma *decision tree* dan algoritma *random forest* dalam melakukan prediksi. Manfaat penelitian ini memberikan kontribusi ilmiah, khususnya dalam bidang ilmu *machine learning*. Selain itu juga sebagai rujukan pengembangan keilmuan di bidang kesehatan dalam mencegah penyakit yang lebih akurat dan efisien.

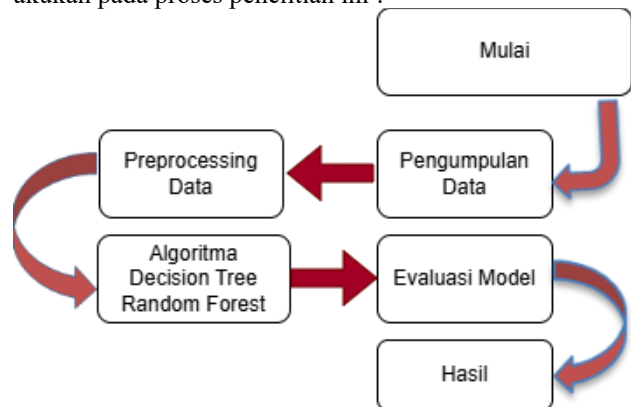
Beberapa peneliti sebelumnya masih terbatas pada penggunaan satu algoritma dan satu jenis penyakit tanpa membandingkan secara menyeluruh, terutama pada dataset yang memiliki banyak fitur. Penelitian yang dilakukan oleh Lailin Nafi'ah dan Zaehol Fatah [12], yang berjudul tentang Implementasi Algoritma *decision tree* untuk pendeteksian penyakit jantung. Hasil penelitian tersebut menjelaskan algoritma *decision tree* mencapai tingkat akurasi yang cukup baik dan sangat tinggi sebesar 80 hingga 90 %. Dhini Septhya dkk [13] yang berjudul Implementasi algoritma *decision tree* dan *support vector machine* untuk klasifikasi penyakit kanker paru-paru. Hasil penelitian tersebut diperoleh pengujian algoritma paling tinggi pada *decision tree* awal dengan rasio 80:20

bernilai 56,7 dan pendekatan penelitian ditambah *forward selection* 53,2. Selain itu juga penelitian yang dilakukan oleh Arifin Yusuf dkk. [14], yang berjudul Penerapan data mining dalam analisis prediksi kanker paru menggunakan algoritma *random forest*. Hasil penelitian tersebut menjelaskan bahwa algoritma *random forest* telah mendapatkan akurasi terbesar senilai 90,61%. Penelitian Shafira A Putri dkk [15], yang berjudul Penerapan *machine learning* algoritma *random forest* untuk prediksi penyakit jantung. Penelitian ini menghasilkan nilai 87,7%.

II. METODE

Model penelitian ini menggunakan metode kuantitatif, dataset yang digunakan tentang peduli sehat, dikumpulkan melalui platform kaggle.com, dengan jumlah data 4921 *record*. Data ini memiliki 130 jenis gejala, seperti batuk, demam, sakit kepala, dll. Serta memiliki diagnosa/*prognosis* penyakit yang dikaitkan dengan kombinasi gejala tersebut seperti demam berdarah, infeksi jamur, alergi, flu, tipis, dll.

Gambar 1. Menjelaskan langkah-langkah yang dilakukan pada proses penelitian ini :



Gambar 1. Tahapan Metode Penelitian

A. Pengumpulan Data

Data penelitian ini dapat diunduh di url <https://www.kaggle.com/datasets/krismonosadi/peduli-sehat-dataset>, sebanyak 4921 Data.

B. Preprocessing Data

Pada tahap *preprocessing data*, dilakukan beberapa langkah penting untuk menyediakan data dengan kondisi yang siap digunakan.

- Tahap pertama adalah pemisahan data berdasarkan variabel, dengan cara mengelompokkan data sesuai variabel yang terdapat dalam dataset penelitian.
- Tahap kedua adalah memilih atribut yang relevan untuk model, proses seleksi atribut ini untuk mengurangi kompleksitas data, serta menghasilkan model analisis yang lebih optimal dan akurat.
- Tahap ketiga adalah pemberian nama dan tipe data pada atribut, setiap atribut diberikan nama yang jelas, serta ditentukan tipe datanya, seperti *numerik* atau *kategorikal*.
- Tahap keempat adalah melakukan pembagian dataset dengan rasio 70 : 30, terdiri dari data latih (*training data*) dan data uji (*testing data*) yang digunakan untuk mengevaluasi kinerja model

C. Algoritma Decision Tree

Setelah dilakukan *preprocessing data*, kemudian dianalisis menggunakan algoritma *decision tree*, yang akan digunakan untuk membuat model klasifikasi dengan bentuk struktur pohon keputusan yang menunjukkan hubungan antara data dan hasil klasifikasi. Adapun beberapa langkah, yaitu :

- Menentukan dataset yang digunakan terdiri dari beberapa atribut dan satu atribut target atau kelas yang diprediksi.
- Menentukan atribut akar (*root node*), pemilihan atribut dilakukan menggunakan metode perhitungan *Gain Ratio*. Atribut dengan nilai terbaik akan menjadi pemisah pertama dalam pohon.
- Menentukan *Criterion: Gain Ratio, Max Depth : 20*, dan *confidence : 0,25*
- Membagi data menjadi cabang, setiap nilai atribut akan membentuk cabang baru dalam pohon keputusan.
- Membentuk node baru, pada setiap cabang, algoritma kembali memilih atribut terbaik hingga terbentuk node-node baru
- Membentuk *leaf node*, node terakhir disebut *leaf node* yang berisi hasil prediksi.

D. Algoritma Random Forest

Selanjutnya data akan dilolah menggunakan Algoritma *random forest* yang merupakan pengembangan dari *decision tree* yang menggunakan banyak pohon keputusan untuk meningkatkan akurasi. Adapun langkah, yaitu :

- Menentukan dataset yang digunakan sama seperti pada *decision tree*, yaitu data yang memiliki atribut dan label kelas.
- Menggunakan teknik *bootstrap sampling*, pengambilan sampel secara acak, setiap subset data akan digunakan untuk membangun satu pohon keputusan.
- Memilih atribut secara acak, setiap node pohon hanya memilih sebagian atribut secara acak untuk menentukan pemisahan data.
- Menentukan *Criterion: Gain Ratio, Max Depth : 20*, dan *number of trees : 30*.
- Membangun banyak pohon keputusan, setiap pohon menghasilkan prediksi masing-masing. Kelas yang paling banyak dipilih oleh pohon keputusan, akan menjadi hasil prediksi.

E. Evaluasi Model

- *Accuracy* menunjukkan presentase data yang diprediksi dengan benar oleh model. Menggunakan rumus : $Accuracy = (TP+TN)/(TP+TN+FP+FN)$ (1)
- *Precision*, mengukur ketepatan model dalam memprediksi kelas positif.
- *Recall*, mengevaluasi kemampuan model dalam menemukan semua data yang masuk ke dalam kategori positif.

F. Hasil Perbandingan

Tahap ini dilakukan setelah proses pemodelan dan evaluasi model, akan dilakukan analisis perbedaan performa antara algoritma *decision tree* dan algoritma *random forest*.

III. HASIL DAN PEMBAHASAN

A. Pengumpulan Data

Data penelitian ini dikumpulkan melalui kaggle.com dan dapat di unduh melalui <https://www.kaggle.com/datasets/krismonosadi/peduli-sehat-dataset>, sebanyak 4921 *record* dan 130 jenis gejala penyakit. Berikut contoh data yang telah diperoleh :

[illegible]

Gambar 2. Contoh Dataset

B. Analisis Data

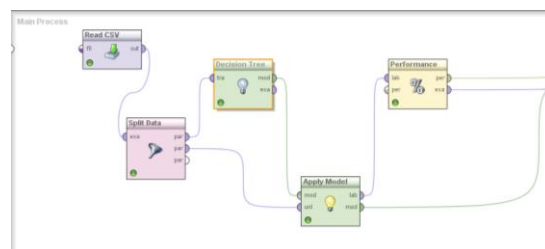
Penelitian ini, analisa data dilakukan dengan dua kali pengujian, yaitu dengan membandingkan dua algoritma *machine learning*, yaitu *decision tree* dan *random forest*, dengan pengujian *split validation* 70:30 untuk mengetahui performa terbaik untuk prediksi penyakit.

Terdapat beberapa tahapan yang perlu dilakukan :

- **Pemrosesan Data**
Tahapan ini dimulai dengan data yang dipilih menggunakan *tools* RapidMiner, selanjutnya data dibersihkan dengan menghapus data atribut yang sama.
- **Pembagian Data**
Pembagian data menjadi 2 bagian, yaitu data latih (*training data*), yang digunakan untuk melatih model dan data uji (*testing data*) yang digunakan untuk menguji seberapa baik model bekerja

C. Pengujian Algoritma *Decision Tree*

Tahapan ini dilakukan pengujian algoritma *decision tree* dengan *Split Validation* untuk data latih dan uji. Berikut model proses algoritma *decision tree* :



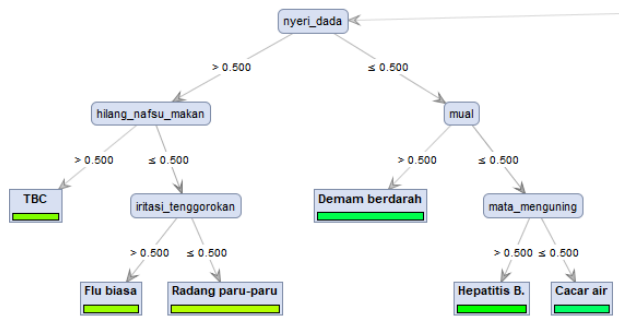
Gambar 3. Desain Model Algoritma Decision Tree

Gambar 3 menjelaskan tentang desain model menggunakan algoritma *decision tree*. Alur utama proses mengambil data dari file CSV, setelah itu menggunakan *split validation* untuk membagi data menjadi data latih dan data

Hasil pengolahan algoritma *decision tree* terlihat pada gambar di bawah ini :

- Hasil Pohon Keputusan

Dari hasil pengujian *tools* RapidMiner, diperoleh gambar pohon keputusan yang menunjukkan pendekatan penelitian, salah satunya seperti pada gambar 4 di bawah ini.



Gambar 4. Hasil Algoritma *Decision Tree*

Gambar 4. Hasil pohon keputusan memiliki atribut gejala rasa tidak enak badan sebagai akar (*root*), menjelaskan :

1. Jika gejala rasa tidak enak badan dirasakan (>0.500), proses dilanjutkan ke gejala nyeri dada (>0.500), dan dilanjutkan ke gejala hilang nafsu makan, maka pasien diprediksi menderita penyakit TBC.
2. Jika pada gejala hilang nafsu makan tidak dirasakan (≤ 0.500), akan dilanjutkan pemeriksaan gejala iritasi tenggorokan, dimana jika merasakan gejala iritasi tenggorokan (>0.500), maka hasil prediksi mengarah menderita penyakit Flu biasa, sedangkan jika tidak merasakan iritasi tenggorokan (≤ 0.500) maka diprediksi menderita penyakit Radang paru-paru.
3. Jika gejala rasa tidak enak badan dirasakan (>0.500), proses dilanjutkan ke gejala nyeri dada, apabila tidak merasakan gejala nyeri dada (≤ 0.500), maka model akan mengevaluasi atribut gejala mual. Jika mual (>0.500) maka diprediksi menderita penyakit Demam berdarah.
4. Jika tidak merasakan mual (≤ 0.500), maka akan dilanjutkan ke atribut gejala mata menguning. Pada tahap ini, jika mata menguning (>0.500), maka hasil prediksi menderita penyakit Hepatitis B, sedangkan jika mata tidak menguning (≤ 0.500), maka hasil prediksi menderita penyakit Cacar air.

- Hasil Akurasi

[illegible]

Gambar 5. Hasil Akurasi Algoritma *Decision Tree*

Hasil pengujian *performance* yang dilakukan menggunakan algoritma *decision tree* menunjukkan bahwa hasil akurasi 94,17 %.

- Hasil *Precision*

[illegible]

Gambar 6. Hasil *Precision* Algoritma *Decision Tree*

Hasil pengujian *performance* yang dilakukan menggunakan algoritma *decision tree* mendapatkan hasil *precision* 95,04 %.

- Hasil *Recall*

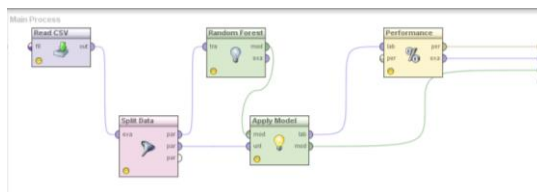
[illegible]

Gambar 7. Hasil *Recall* Algoritma *Decision Tree*

Hasil pengujian *performance* yang dilakukan menggunakan algoritma *decision tree* mendapatkan hasil *recall* 94,55 %.

D. Pengujian Algoritma *Random Forest*

Tahapan ini dilakukan pengujian menggunakan algoritma *random forest* dengan *split validation* untuk data latih dan data uji. Berikut model proses algoritma *random forest* :



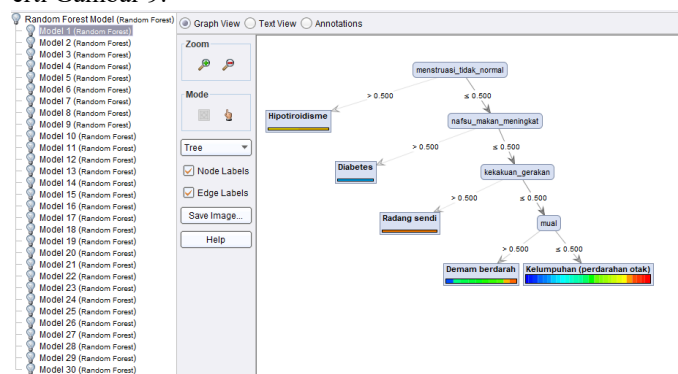
Gambar 8. Desain Model Algoritma *Random Forest*

Gambar 8 menjelaskan tentang desain model menggunakan algoritma *random forest*. Alur utama proses mengambil data dari *file* CSV, setelah itu menggunakan *split validation* digunakan untuk membagi data menjadi data latih dan data uji, kemudian data latih diproses pada tahap algoritma *random forest* dengan menghasilkan 30 model bentuk pohon keputusan, disetiap model menghasilkan gejala dan hasil prediksi penyakit. Kemudian model yang telah terbentuk diterapkan pada data uji melalui proses *apply model* sehingga dapat menghasilkan prediksi, pada tahap akhir *performance* mengevaluasi hasil prediksi yang diperoleh nilai kinerja seperti akurasi, *precision* dan *recall*.

Hasil pengolahan algoritma *random forest* terlihat pada gambar di bawah ini :

- Hasil Pohon Keputusan

Hasil pengujian *tools* RapidMiner, didapatkan 30 model pohon keputusan, salah satunya model 1 (*random forest*) seperti Gambar 9.



Gambar 9. Hasil Model 1 *Random Forest*

Gambar 9. Hasil pohon keputusan pada model 1 memiliki atribut gejala menstruasi tidak normal sebagai akar (*root*), menjelaskan :

1. Jika gejala mentruasi tidak normal dirasakan ($>0,500$), maka diprediksi menderita penyakit Hipotiroidisme.
2. Jika gejala mentruasi tidak normal tidak dirasakan ($\leq 0,500$), akan dilanjutkan pemeriksaan gejala nafsu makan meningkat, apabila gejala nafsu makan meningkat dirasakan ($>0,500$), maka diprediksi menderita penyakit Diabetes.
3. Jika gejala nafsu makan meningkat tidak dirasakan ($\leq 0,500$), akan dilanjutkan pemeriksaan gejala kekakuan gerakan, apabila gejala kekakuan gerakan dirasakan ($>0,500$), maka diprediksi menderita penyakit Radang sendi.
4. Jika gejala kekakuan gerakan tidak dirasakan ($\leq 0,500$), akan dilanjutkan pemeriksaan gejala mual, apabila gejala mual dirasakan ($>0,500$), maka diprediksi menderita penyakit Demam Berdarah, sedangkan gejala mual tidak dirasakan ($\leq 0,500$), maka diprediksi menderita penyakit Kelumpuhan (Pendarahan Otak).

- Hasil Akurasi

[illegible]

Gambar 10. Hasil Akurasi Algoritma *Random Forest*

Hasil pengujian *performance* yang dilakukan menggunakan algoritma *random forest* mendapatkan akurasi 44,65 %.

- Hasil *Precision*

[illegible]

Gambar 11. Hasil *Precision* Algoritma *Random Forest*

Hasil pengujian *performance* yang dilakukan menggunakan algoritma *random forest* mendapatkan *precision* 45,88 %.

- Hasil *Recall*

[illegible]

Gambar 12. Hasil *Recall* Algoritma *Random Forest*

Hasil pengujian *performance* yang dilakukan menggunakan algoritma *random forest* mendapatkan *recall* 46,67 %.

E. Pembahasan

Berdasarkan hasil akurasi, *precision* dan *recall* yang didapatkan dari pengujian algoritma *decision tree* dan *random forest* telah memberikan hasil perbandingan :

Tabel 1. Perbandingan *Accuracy*, *Precision* dan *Recall*

	Algoritma <i>Decision Tree</i>	Algoritma <i>Random Forest</i>
Accuracy	94,17 %	44,65 %
Precision	95,04 %	45,88 %

	Algoritma <i>Decision Tree</i>	Algoritma <i>Random Forest</i>
Recall	94,55 %	46,67 %

Berdasarkan Tabel 1. Hasil evaluasi performa yang telah dilakukan di dapatkan, bahwa hasil performa dari algoritma *decision tree* lebih baik dikarenakan mampu memahami pola data dengan baik dan sesuai dengan karakteristik dataset yang digunakan, sedangkan algoritma *random forest* memiliki performa yang lebih rendah karena jumlah fitur gejala penyakit yang mencapai 130, sehingga menyebabkan dimensi fitur menjadi terlalu besar yang bisa mengurangi kemampuan model jika tidak dilakukan fitur yang tepat dan penggunaan parameter tidak optimal, sehingga menurunkan performa.

IV. KESIMPULAN

Melalui hasil pengujian algoritma *decision tree* dan *random forest*, menghasilkan perbedaan hasil menunjukkan bahwa algoritma *decision tree* memiliki hasil lebih tinggi, yaitu akurasi 94,17%, *precision* 95,04 % dan *recall* 94,55 %. Studi ini membuktikan bahwa algoritma *decision tree* lebih efektif untuk dataset dengan fitur gejala yang banyak namun berpola linear. Sehingga hasil penelitian ini dapat memberikan kontribusi ilmiah, khususnya dalam bidang ilmu *machine learning*. Selain itu juga sebagai rujukan pengembangan keilmuan di bidang kesehatan dalam mencegah penyakit yang lebih akurat dan efisien. Penelitian selanjutnya diharapkan dapat mencoba teknik *Hyperparameter Tuning* untuk meningkatkan performa *Random Forest*.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Lembaga Penelitian dan Pengabdian Masyarakat Universitas Ichsan Gorontalo karena telah mendukung penelitian ini melalui program Hibah Penelitian Kompetitif Dosen 2026.

REFERENSI

- [1] S. Matondang, "Sistem Pakar Diagnosis Penyakit Umum Menggunakan Metode Forward Chaining," *Remik Ris. dan E-Jurnal Manaj. Inform. Komput.*, vol. 7, pp. 1712–1723, 2023.
- [2] A. L. Kartika, M. Orisa, D. Rudhistiar, T. Informatika, D. Shafer, and P. Umum, "Aplikasi Sistem Pakar Deteksi Penyakit Umum Berbasis Android," vol. 7, no. 5, pp. 3068–3074, 2023.
- [3] M. J. Suwardy, "Pengembangan Aplikasi Panduan Perawatan dan Gejala Pengidap Penyakit Umum Ringan Menggunakan Adobe Flash," vol. 3, no. 1, pp. 62–67, 2024.
- [4] A. M. Multazam and E. Andayanie, *Kesehatan*

Masyarakat. 2024.

- [5] M. Muhrizal, A. Haditsah, W. Yunus, and S. S. Panna, "Prediksi jumlah produksi kopra menggunakan metode regresi linear berganda pada umkm mandiri desa lion," vol. 4, no. 2, pp. 141–149, 2025.
- [6] M. A. Salwa, "Optimasi Model Algoritma Klasifikasi Data Mining Menggunakan Metode Feature Selection Untuk Prediksi Penyakit Stroke," vol. 26, no. 1, pp. 11–20, 2025.
- [7] E. Sabna and O. Dewi, "Prediksi Penyakit Stroke menggunakan Algoritma Decision Tree dan Naïve Bayes," vol. 4, no. 3, pp. 1294–1299, 2025.
- [8] U. Maltuf and Z. Fatah, "Penerapan Algoritma Decision Tree Untuk Klasifikasi Komsumsi Energi Listrik Rumah Tangga Dengan Penggunaan Rapid Miner," *J. Ilmia Multidisplin Ilmu*, vol. 2, no. 1, pp. 38–45, 2025.
- [9] Y. Nuryamin and F. Risyda, "Analisis Prediksi Penyakit Diabetes Menggunakan Metode Decision Tree C4 . 5 Dan Naive Bayes," vol. 5, pp. 234–242, 2025.
- [10] J. Kuswanto and L. Hakim, "Penerapan Algoritma Random Forest untuk memprediksi Performa Akademik Mahasiswa," vol. 5, no. 1, pp. 262–270, 2025.
- [11] M. A. Kurniawan, G. Z. Syauqi, M. Safriyanti, F. U. Azmie, and A. Setiawan, "Prediksi Pendapatan Penjualan di Indomaret Menggunakan Algoritma Random Forest Regression," vol. 9675, pp. 93–99, 2025.
- [12] L. Nafi and Z. Fatah, "Implementasi Algoritma Decision Tree Untuk Pendeteksian Penyakit Jantung," *Jusifor J. Sist. Inf. dan Inform.*, vol. 3, no. 2, pp. 160–165, 2024.
- [13] D. Septhya, K. Rahayu, S. Rabbani, and V. Fitria, "Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification Implementasi Algoritma Decision Tree dan Support Vector Machine untuk Klasifikasi Penyakit Kanker Paru," vol. 3, no. April, pp. 15–19, 2023.
- [14] A. Y. Permana, H. N. Fazri, and M. F. N. Athoilah, "Penerapan Data Mining Dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma Random Forest," vol. 3, no. 2, 2023.
- [15] S. A. Putri, N. Selayanti, and M. Kristanaya, "Penerapan Machine Learning Algoritma Random Forest Untuk Prediksi Penyakit Jantung," vol. 2024, no. Senada, 2024.