ANALISIS PERBANDINGAN ALGORITMA *RANDOM FOREST* DAN NAÏVE BAYES DALAM KLASIFIKASI PENYAKIT AIDS

Muhamad Tedi Saprudin[1)], Budi Sudrajat[2)], Hasta Herlan Asymar[3)]
[1,2]Prodi Informatika, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika
[3]Prodi Teknologi Komputer, Fakultas Teknik & Informatika, Universitas Bina Sarana Informatika

Correspondence author: M.T.Saprudin, muhamadtedi078@gmail.com, Jakarta, Indonesia

Abstract

This study aims to compare the performance of the *Random Forest* and Naïve Bayes algorithms in classifying AIDS based on patient data. The *dataset* was obtained from the Kaggle platform and consists of 2,139 data points with 23 attributes, including medical information such as age, weight, haemoglobin level, gender, treatment type, and CD4 and CD8 cell counts. The research method comprises several stages: data collection, *pre-processing*, model building, and evaluation. The model was evaluated using metrics such as accuracy, *precision*, *recall*, F1-score, confusion matrix, and AUC-ROC. The results showed that the *Random Forest* algorithm outperformed Naïve Bayes. *Random Forest* achieved an accuracy of 88.16% and an AUC of 91.25%, while Naïve Bayes only achieved an accuracy of 81.46% and an AUC of 81.40%. These findings indicate that *Random Forest* is better able to handle complex, non-linear medical data and provides more stable, reliable classification results. Therefore, *Random Forest* is recommended as a more effective algorithm for classifying AIDS cases based on the *dataset* used in this study.

Keywords: *Random Forest*, naïve bayes, classification algorithm, aids

Abstrak

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Random Forest* dan Naïve Bayes dalam melakukan klasifikasi penyakit AIDS berdasarkan data pasien. *Dataset* yang digunakan diperoleh dari platform Kaggle dan terdiri atas 2.139 data dengan 23 atribut, yang mencakup informasi medis seperti usia, berat badan, kadar hemoglobin, jenis kelamin, jenis pengobatan, serta jumlah sel CD4 dan CD8. Metode penelitian meliputi beberapa tahapan, yaitu pengumpulan data, pra-pemrosesan, pembentukan model, dan evaluasi. Model dievaluasi menggunakan metrik akurasi, presisi, *recall*, F1-score, confusion matrix, dan AUC-ROC. Hasil penelitian menunjukkan bahwa algoritma *Random Forest* memiliki performa lebih baik dibandingkan Naïve Bayes. *Random Forest* mencapai akurasi sebesar 88,16% dan nilai AUC sebesar 91,25%, sedangkan Naïve Bayes hanya memperoleh akurasi 81,46% dengan AUC 81,40%. Temuan ini menunjukkan bahwa *Random Forest* lebih mampu menangani data medis yang kompleks dan non-linear, serta memberikan hasil klasifikasi yang lebih stabil dan andal. Oleh karena itu, *Random Forest* direkomendasikan sebagai algoritma yang lebih efektif untuk klasifikasi kasus AIDS menggunakan *dataset* pada penelitian ini.

Kata Kunci: *Random Forest, naïve bayes,* algoritma klasifikasi, penyakit aids

JURNAL ELEKTRO & INFORMATIKA
SWADHARMA (JEIS)
VOLUME 06 NOMOR 01 JANUARI 2026
ISSN 2774 – 5775 | eISSN 2774-5767

A. PENDAHULUAN

Penyakit HIV/AIDS (*Human Immunodeficiency Virus / Acquired Immune Deficiency Syndrome*) merupakan salah satu penyakit menular berbahaya yang masih menjadi permasalahan kesehatan global hingga saat ini. Menurut data *World Health Organization*, lebih dari 39 juta orang di seluruh dunia hidup dengan HIV/AIDS, dan sekitar 1,3 juta kasus baru muncul setiap tahunnya (Sari et al., 2024). Laporan *Joint United Nations Programme on HIV/AIDS* (UNAIDS), juga menyebutkan bahwa HIV/AIDS masih menjadi salah satu penyebab utama kematian pada kelompok usia produktif di berbagai negara (Barus et al., 2024).

Di Indonesia, Kementerian Kesehatan Republik Indonesia, melaporkan bahwa jumlah kasus HIV/AIDS terus meningkat setiap tahunnya. Berdasarkan data nasional, hingga tahun 2023 tercatat lebih dari 500 ribu kasus HIV dan 130 ribu kasus AIDS dengan penyebaran tertinggi di provinsi DKI Jakarta, Jawa Timur, dan Papua. Peningkatan ini menunjukkan perlunya strategi yang lebih efektif dalam deteksi dini serta pengendalian penularan penyakit HIV/AIDS (Sari et al., 2024).

HIV (*Human Immunodeficiency Virus*) merupakan virus yang menyerang dan merusak sel darah putih, sehingga mengakibatkan penurunan daya tahan tubuh seseorang. AIDS merupakan tahap di mana kerusakan sistem imun akibat HIV sudah sangat parah, sehingga infeksi ringan sekalipun bisa berkembang menjadi ancaman serius bagi penderitanya. Tahap akhir dari infeksi ini dikenal sebagai AIDS, yaitu kondisi ketika sistem kekebalan tubuh sudah tidak mampu melawan infeksi yang masuk. Deteksi dini terhadap infeksi HIV sangat penting agar pengobatan dapat dilakukan sedini mungkin dan mencegah perkembangan menuju tahap AIDS (Oktavia et al., 2022).

Metode deteksi HIV/AIDS secara konvensional masih mengandalkan pemeriksaan laboratorium, seperti tes antibodi dan tes PCR. Meskipun akurat, metode tersebut memerlukan waktu, biaya, dan fasilitas kesehatan yang memadai (Aini et al., 2025). Dalam konteks ini, kemajuan teknologi informasi khususnya di bidang kecerdasan buatan (*Artificial Intelligence*) dan pembelajaran mesin (*Machine learning*) membuka peluang baru untuk mendeteksi penyakit secara cepat dan efisien dengan memanfaatkan data medis pasien (Handayani, 2024).

Machine learning merupakan cabang kecerdasan buatan yang mampu menganalisis data, mengenali pola, dan membuat prediksi berdasarkan pengalaman dari data sebelumnya. Model klasifikasi *machine learning* telah banyak digunakan untuk mendeteksi berbagai penyakit seperti diabetes (Oktaviana et al., 2024), kanker payudara (Cahyana & Nurlayli, 2023), dan jantung (Samosir et al., 2021). Namun, penelitian yang secara spesifik membahas penerapan algoritma *machine learning* untuk klasifikasi penyakit AIDS masih relatif terbatas.

Beberapa penelitian sebelumnya telah menunjukkan keberhasilan penggunaan algoritma *machine learning* dalam bidang medis. Misalnya, penelitian oleh (Hanifah & Kriswibowo, 2023) menganalisis tren penyebaran penyakit HIV/AIDS di Indonesia dan menemukan pentingnya sistem deteksi digital berbasis data pasien. Penelitian lainnya oleh (Lestari & Homaidi, 2024; Miftakhudin et al., 2025) membandingkan kinerja berbagai algoritma seperti *Random Forest, Naïve Bayes*, dan KNN untuk klasifikasi data medis dan menemukan bahwa *Random Forest* cenderung memberikan hasil yang lebih akurat dan stabil.

Random Forest merupakan metode *ensemble learning* yang menggabungkan beberapa pohon keputusan (*decision tree*) untuk menghasilkan prediksi yang lebih kuat melalui sistem voting. Algoritma ini memiliki keunggulan dalam menangani data yang kompleks dan beragam (Setiawan et al.,

2025). Sementara itu, *Naïve Bayes* adalah algoritma berbasis probabilitas sederhana yang cepat dan efisien, namun performanya menurun ketika terdapat korelasi antar atribut data (Novelan & Aryza, 2025).

Berdasarkan uraian tersebut, penelitian ini dilakukan untuk menganalisis dan membandingkan kinerja algoritma *Random Forest* dan *Naïve Bayes* dalam mengklasifikasikan penyakit AIDS berdasarkan data medis pasien. Penelitian ini bertujuan untuk menentukan algoritma yang lebih efektif dalam mendeteksi status infeksi HIV/AIDS, sehingga dapat mendukung pengembangan sistem deteksi dini berbasis *machine learning* di bidang kesehatan.

B. METODE PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan komparatif yang bertujuan untuk membandingkan performa algoritma *Random Forest* dan *Naïve Bayes* dalam mengklasifikasikan penyakit AIDS berdasarkan indikator kesehatan pasien. Data yang digunakan berasal dari sumber sekunder, yaitu *dataset* yang diperoleh dari situs Kaggle (www.kaggle.com).

Studi Literatur

Tahap awal penelitian ini diawali dengan kegiatan studi literatur yang bertujuan untuk memperoleh landasan teori dan referensi terkait penelitian sebelumnya mengenai penerapan algoritma *machine learning* dalam bidang kesehatan. Studi literatur dilakukan dengan meninjau berbagai sumber seperti jurnal ilmiah, buku, dan artikel daring yang membahas algoritma *Random Forest* dan *Naïve Bayes*. Melalui tahap ini, peneliti memperoleh pemahaman mengenai konsep dasar, karakteristik algoritma, serta hasil penelitian terdahulu yang relevan. Informasi yang diperoleh dari studi literatur menjadi acuan dalam menentukan arah penelitian, rancangan model, serta metode evaluasi yang digunakan.

Pengumpulan Data

Penelitian ini menggunakan *dataset* sekunder yang diperoleh dari situs Kaggle dan berisi data medis pasien terkait HIV/AIDS. *Dataset* ini mencakup berbagai indikator kesehatan seperti usia, kadar hemoglobin, jumlah sel CD4 dan CD8, jenis terapi, serta status infeksi HIV sebagai variabel target. *Dataset* ini dipilih karena bersifat terbuka dan memiliki format yang sesuai untuk penelitian berbasis *machine learning*. Selain itu, data tersebut memiliki indikator kesehatan yang relevan untuk proses klasifikasi dan umum digunakan dalam penelitian prediksi penyakit. Secara keseluruhan, *dataset* terdiri atas 2.139 data pasien dan 23 atribut dengan satu variabel target berupa status infeksi HIV (0 = tidak terinfeksi, 1 = terinfeksi).

Pre-processing Data

Tahap *pre-processing* data dilakukan untuk memastikan bahwa data yang digunakan siap diolah oleh model *machine learning*. Langkah pertama yang dilakukan adalah pembersihan data (*data cleaning*) untuk menghapus nilai kosong (*missing values*) dan data duplikat yang dapat mengganggu hasil pemodelan. Selanjutnya dilakukan transformasi data, yaitu mengubah atribut kategorikal menjadi numerik menggunakan metode Label *Encoding* agar dapat diproses oleh algoritma. Selain itu, dilakukan normalisasi data untuk menyeragamkan skala antar atribut agar model dapat belajar secara optimal. Setelah proses tersebut selesai, *dataset* dibagi menjadi dua bagian yaitu 70% untuk data training dan 30% untuk data testing, dengan tujuan agar model dapat diuji pada data yang belum pernah dilihat sebelumnya.

Membentuk Model

Pada tahap ini, peneliti membangun dua model klasifikasi menggunakan algoritma *Random Forest* dan *Naïve Bayes*. Algoritma *Random Forest* merupakan metode *ensemble learning* berbasis pohon keputusan (*decision tree*) yang membentuk banyak pohon secara acak dan menggabungkan hasil prediksi

JURNAL ELEKTRO & INFORMATIKA
SWADHARMA (JEIS)
VOLUME 06 NOMOR 01 JANUARI 2026
ISSN 2774 – 5775 | eISSN 2774-5767

melalui sistem *majority voting*. Keunggulan *Random Forest* adalah kemampuannya menangani data berukuran besar, mengurangi risiko *overfitting*, serta memberikan hasil prediksi yang stabil. Sementara itu, algoritma *Naïve Bayes* merupakan metode berbasis probabilitas yang menggunakan Teorema Bayes dengan asumsi bahwa setiap fitur bersifat independen. Algoritma ini sederhana, cepat, dan efisien untuk digunakan pada *dataset* berukuran besar dengan banyak atribut.

Pengujian Model

Tahap terakhir adalah pengujian model untuk mengevaluasi performa kedua algoritma dalam melakukan klasifikasi penyakit AIDS. Evaluasi dilakukan menggunakan beberapa metrik pengukuran, yaitu *accuracy* untuk mengukur tingkat ketepatan model secara keseluruhan, *precision* untuk menilai ketepatan prediksi positif, *recall* (*sensitivity*) untuk mengukur kemampuan model dalam mendeteksi kelas positif sebenarnya, serta *F1-score* sebagai rata-rata harmonis antara *precision* dan *recall*. Selain itu, digunakan juga AUC-ROC (*Area Under Curve – Receiver Operating Characteristic*) untuk menilai kemampuan model dalam membedakan antara kelas positif dan negatif.

C. HASIL DAN PEMBAHASAN

Pemahaman data dilakukan untuk mengetahui karakteristik *dataset* sebelum dilakukan proses pemodelan. Tahap ini meliputi pemeriksaan struktur *dataset* untuk mengetahui jumlah baris dan kolom, serta mengevaluasi distribusi kelas target *Infected*, yaitu jumlah data pasien yang terinfeksi HIV (kelas 1) dan yang tidak terinfeksi HIV (kelas 0). Dengan analisis ini, peneliti dapat melihat keseimbangan data serta memahami kondisi awal variabel yang akan digunakan pada tahap pra-pemrosesan dan pemodelan.

Tabel 1. Atribut *Dataset*

NO	NAMA ATRIBUT	KETERANGAN
1	Time	Waktu sejak awal pengamatan atau perawatan
2	Trt	Jenis perawatan atau terapi
3	Age	Usia pasien
4	Wtkg	Berat badan dalam kilogram
5	Hemo	Kadar hemoglobin
6	MSM	Status homoseksual (ya/tidak)
7	Drugs	Penggunaan obat terlarang
8	Karnof	Skor Karnofsky (tingkat kemampuan aktivitas fisik)
9	Oprior	Jumlah pengobatan sebelumnya
10	Z30	Penerimaan zidovudine (AZT) dalam 30 hari terakhir
11	Preanti	Terapi antiretroviral sebelumnya
12	Race	Ras atau etnis pasien
13	Gender	Jenis kelamin
14	Str2	Variabel stratifikasi tambahan
15	Strat	Variabel stratifikasi utama
16	Symptom	Gejala klinis HIV/AIDS
17	Treat	Jenis pengobatan selama penelitian
18	Offtrt	Status penghentian pengobatan
19	Cd40	Jumlah sel CD4 awal
20	Cd420	Jumlah sel CD4 setelah 20 minggu
21	Cd80	Jumlah sel CD8 awal
22	Cd820	Jumlah sel CD8 setelah 20 minggu
23	Infected	Status infeksi HIV (kelas target: terinfeksi (1) / tidak terinfeksi (0))

Tabel 2. Sebelum *Cleaning Dataset*

No	Column	Non-Null Count	Dtype
0	time	2139 non-null	int64
1	trt	2139 non-null	int64
2	age	2139 non-null	int64
3	wtkg	2139 non-null	object
4	hemo	2139 non-null	int64

No	Column	Non-Null Count	Dtype
5	homo	2139 non-null	int64
6	drugs	2139 non-null	int64
7	karnof	2139 non-null	int64
8	oprior	2139 non-null	int64
9	z30	2139 non-null	int64
10	preanti	2139 non-null	int64
11	race	2139 non-null	int64
12	gender	2139 non-null	int64
13	str2	2139 non-null	int64
14	strat	2139 non-null	int64
15	symptom	2139 non-null	int64
16	treat	2139 non-null	int64
17	offtrt	2139 non-null	int64
18	cd40	2139 non-null	int64
19	cd420	2139 non-null	int64
20	cd80	2139 non-null	int64
21	cd820	2139 non-null	int64

Analisis Awal dan Pembersihan Data

Dataset yang digunakan berjumlah 2.139 entri dengan 23 kolom (atribut) yang berisi berbagai indikator kesehatan pasien. Atribut-atribut tersebut meliputi data demografis, hasil pemeriksaan laboratorium, serta status infeksi HIV yang menjadi target klasifikasi.

Berdasarkan hasil analisis awal, seluruh kolom pada *dataset* memiliki nilai lengkap (*non-null count* = 2.139) sehingga tidak ditemukan data yang hilang (*missing value*). Proses data *cleaning* dilakukan untuk memastikan konsistensi tipe data, di mana seluruh atribut telah dikonversi ke format numerik (int64) agar dapat diolah dengan baik oleh model *machine learning*.

Setelah dilakukan proses data cleaning, seluruh data yang tidak relevan, duplikat, serta nilai kosong (missing values) telah dihapus sehingga *dataset* menjadi bersih dan siap digunakan untuk tahap analisis selanjutnya. Informasi mengenai struktur data setelah proses pembersihan dapat dilihat pada Tabel 3, yang menunjukkan jumlah entri, nama kolom, tipe data, serta jumlah nilai non-null pada masing-masing atribut.

Tabel 3. Sesudah Cleaning *Dataset*

No	Kolom	Non- Null Count	Dtype
0	time	2139 non-null	int64
1	trt	2139 non-null	int64
2	age	2139 non-null	int64
3	wtg	2139 non-null	int64
4	hemo	2139 non-null	int64
5	homo	2139 non-null	int64
6	drugs	2139 non-null	int64
7	karno	2139 non-null	int64
8	oprior	2139 non-null	int64
9	z30	2139 non-null	int64
10	preanti	2139 non-null	int64
11	race	2139 non-null	int64
12	gender	2139 non-null	int64
13	str2	2139 non-null	int64
14	strat	2139 non-null	int64
15	symptom	2139 non-null	int64
16	treat	2139 non-null	int64
17	offtrt	2139 non-null	int64
18	cd40	2139 non-null	int64
19	cd420	2139 non-null	int64
20	cd80	2139 non-null	int64
21	cd820	2139 non-null	int64
22	infected	2139 non-null	int64

Dari kedua tabel tersebut, dapat dilihat bahwa tidak terdapat perbedaan jumlah entri sebelum dan sesudah proses pembersihan. Hal ini menunjukkan bahwa data sudah bersih dan siap digunakan untuk proses pemodelan.

Hasil Pengujian Algoritma *Random Forest*

Algoritma *Random Forest* diuji menggunakan data testing sebanyak 642

JURNAL ELEKTRO & INFORMATIKA
SWADHARMA (JEIS)
VOLUME 06 NOMOR 01 JANUARI 2026
ISSN 2774 – 5775 | eISSN 2774-5767

sampel. Hasil pengujian model ini dapat dilihat pada Tabel 4, yang menampilkan hasil evaluasi lengkap dengan metrik akurasi, AUC-ROC, serta nilai *precision*, *recall*, dan f1-score untuk masing-masing kelas.

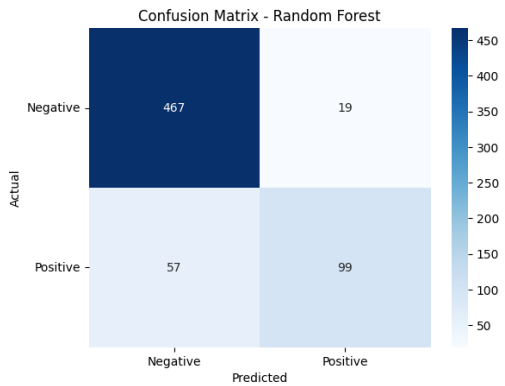
Tabel 4. Hasil Evaluasi *Random Forest*

Metric / Kelas	*Precision*	*Recall*	F1-Score	Support
0	0.89	0.96	0.92	486
1	0.84	0.63	0.72	156
Accuracy			0.88	642
Macro Avg	0.87	0.80	0.82	642
Weighted Avg	0.88	0.88	0.88	642

Berdasarkan hasil tersebut, model *Random Forest* menghasilkan akurasi sebesar 0,8816 (88,16%) dan nilai AUC sebesar 0,9125**.** Nilai ini menunjukkan kemampuan model yang sangat baik dalam membedakan antara pasien yang terinfeksi dan tidak terinfeksi HIV

Dari metrik yang ditampilkan, terlihat bahwa kelas negatif (pasien tidak terinfeksi) memiliki *recall* yang sangat tinggi, yaitu 0,96, menandakan model mampu mengenali hampir seluruh pasien negatif dengan tepat. Namun pada kelas positif, nilai *recall* sebesar 0,63 mengindikasikan bahwa masih ada sebagian pasien terinfeksi yang salah terklasifikasi.

Hasil prediksi lebih lanjut divisualisasikan dalam bentuk confusion matrix pada Gambar 1.



Gambar 1. Confusion Matrix *Random Forest*

Berdasarkan *confusion matrix* tersebut, diketahui bahwa 467 data pasien negatif berhasil diklasifikasi dengan benar, sedangkan hanya 19 data negatif yang salah diklasifikasi sebagai positif. Untuk pasien yang terinfeksi HIV, 99 data terprediksi benar sebagai positif, namun terdapat 57 data positif yang salah diklasifikasi sebagai negatif.

Hasil ini menunjukkan bahwa *Random Forest* bekerja dengan sangat baik dalam mendeteksi pasien yang tidak terinfeksi, namun masih perlu peningkatan sensitivitas terhadap pasien positif.

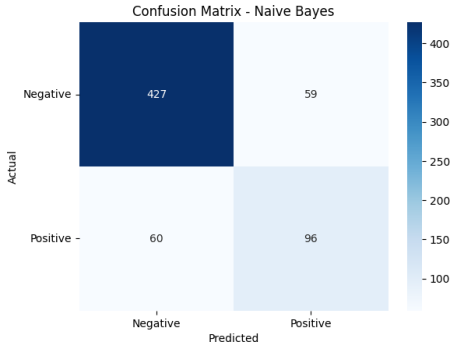
Hasil Pengujian Algoritma *Naïve Bayes*

Pengujian selanjutnya dilakukan menggunakan algoritma *Naïve Bayes* pada *dataset* yang sama. Hasil evaluasi model dapat dilihat pada Tabel 5, yang menampilkan nilai akurasi, AUC-ROC, serta metrik *precision*, *recall*, dan *f1-score* untuk setiap kelas.

Tabel 5. Hasil Evaluasi Naïve Bayes

Metric / Kelas	*Precision*	*Recall*	F1-Score	Support
0	0.88	0.88	0.88	486
1	0.62	0.62	0.62	156
Accuracy			0.81	642
Macro Avg	0.75	0.75	0.75	642
Weighted Avg	0.81	0.81	0.81	642

Berdasarkan hasil tersebut, algoritma *Naïve Bayes* memperoleh akurasi sebesar 0,8146 (81,46%) dengan AUC sebesar 0,8140. Meskipun performanya cukup baik, nilai akurasi ini masih berada di bawah *Random Forest*. Nilai *precision* dan *recall* pada kelas positif (pasien terinfeksi) masing-masing sebesar 0,62, menunjukkan bahwa model masih kesulitan dalam mengenali seluruh pasien yang benar-benar terinfeksi HIV. Detail hasil klasifikasi divisualisasikan pada *confusion matrix* di Gambar 2.

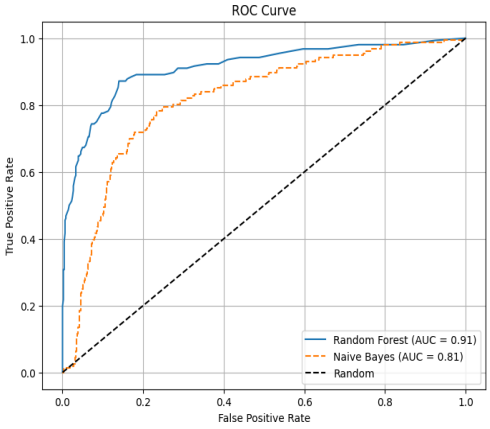


Gambar 2. Confusion Matrix Naïve Bayes

Berdasarkan gambar tersebut, terlihat bahwa 427 pasien negatif berhasil diklasifikasi dengan benar, sedangkan 59 pasien negatif salah diklasifikasi sebagai positif. Untuk kelas positif, 96 pasien terinfeksi berhasil dikenali dengan benar, sementara 60 pasien positif salah diprediksi sebagai negatif. Dari hasil ini dapat disimpulkan bahwa Naïve Bayes memiliki akurasi yang cukup baik, tetapi sensitivitasnya terhadap pasien positif masih lebih rendah dibandingkan *Random Forest*.

Perbandingan Kinerja Kedua Model

Perbandingan performa kedua model ditunjukkan melalui kurva ROC pada Gambar 3. Kurva ini menggambarkan hubungan antara *true positive rate* dan *false positive rate* untuk masing-masing algoritma.



Gambar 3. Plot ROC kedua algoritma

Dari kurva tersebut, terlihat bahwa *Random Forest* memiliki area di bawah kurva (AUC) sebesar 0,91, sedangkan Naïve Bayes sebesar 0,81. Semakin besar nilai AUC, semakin baik kemampuan model dalam membedakan kelas positif dan negatif. Oleh karena itu, dapat disimpulkan bahwa *Random Forest* memiliki performa klasifikasi yang lebih baik dan lebih stabil dibandingkan Naïve Bayes.

D. PENUTUP

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Random Forest* dan Naïve Bayes sama-sama mampu digunakan dalam proses klasifikasi penyakit HIV/AIDS. Keduanya dapat mengenali perbedaan antara pasien yang terinfeksi dan tidak terinfeksi berdasarkan indikator kesehatan yang tersedia pada *dataset*.

Namun demikian, *Random Forest* menunjukkan performa yang lebih baik dibandingkan Naïve Bayes, dengan nilai akurasi sebesar 88,16% dan AUC sebesar 0,91, sedangkan Naïve Bayes memperoleh akurasi 81,46% dengan AUC 0,81. *Random Forest* juga menghasilkan nilai *precision*, *recall*, dan F1-score yang lebih tinggi, terutama pada kelas pasien yang tidak terinfeksi HIV.

Meskipun begitu, Naïve Bayes tetap dapat digunakan sebagai alternatif karena memiliki keunggulan dalam kecepatan komputasi dan efisiensi pemrosesan data. Hasil penelitian ini menunjukkan bahwa selain kinerja model, pemilihan algoritma yang sesuai dan pemahaman terhadap karakteristik data medis juga berperan penting dalam meningkatkan akurasi sistem deteksi penyakit berbasis *machine learning*.

Dengan demikian, penelitian ini menegaskan bahwa algoritma *Random Forest* lebih unggul dan konsisten dalam menghasilkan prediksi status infeksi HIV/AIDS, sedangkan Naïve Bayes tetap relevan sebagai metode pembanding yang sederhana dan cepat.

E. DAFTAR PUSTAKA

Aini, R., Padmarini, A., & Sulistyo, A. (2025). Sensitivity and Specificity of Human Immunodeficiency Virus (HIV) Antibody Testing Using the Immunochromatography Method Rapid Test Compared to ECLIA (Electro Chemiluminescence Immuno Assay). *Jaringan Laboratorium Medis*, *7*(2), 119–126. https://doi.org/10.31983/jlm.v7i2.13620

Barus, D. J., Arwina B., H., & Rajagukguk, D. L. (2024). Gambaran Pengetahuan Remaja Tentang Penyakit HIV/AIDS. *TEKESNOS : Jurnal Teknologi, Kesehatan Dan Ilmu Sosial*, *6*(2), 63–70. https://e-journal.sari-mutiara.ac.id/index.php/tekesnos/article/view/5754

Cahyana, C. W., & Nurlayli, A. (2023). Analisis Performa Logistic Regression, Naïve Bayes, dan Random Forest sebagai Algoritma Pendeteksi Kanker Payudara. *INSERT: Information System and Emerging Technology Journal*, *4*(1), 51–64. https://doi.org/10.23887/insert.v4i1.62362

Handayani, O. P. (2024). Revolusi Kesehatan Digital: Peran AI dan Pembelajaran Mesin dalam Diagnosa Perawatan. *KORISA : Jurnal Kolaborasi Riset Sarjana*, *1*(2), 1–17. https://ejournal.uhb.ac.id/index.php/korisa/article/view/1773

Hanifah, L., & Kriswibowo, A. (2023). Kebijakan Penanggulangan HIV/Aids dalam Perspektif Health Policy Triangle Analysis di Kota Surabaya. *MANHAJ: Jurnal Hukum Dan Pranata Sosial Islam*, *5*(1), 961–970. https://doi.org/10.37680/almanhaj.v5i1.2827

Lestari, I. I., & Homaidi, A. (2024). Komparasi Algoritma Naive Bayes Dan Random Forest Pada Klasifikasi Kanker Payudara. *Gudang Jurnal Multidisiplin Ilmu*, *2*(12), 778–785. https://doi.org/10.59435/gjmi.v2i12.1206

Miftakhudin, A., Santoso, N. A., & Santoso, B. A. (2025). Komparasi Algoritma KNN dan Random Forest untuk Diagnosa Penyakit Jantung Koroner (Studi Kasus: RSUD Dr. Soeselo Slawi). *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(3), 2962–2971. https://doi.org/10.31004/riggs.v4i3.2424

Novelan, M. S., & Aryza, S. (2025). *Belajar Mengenal Dasar Machine Learning*. Payakumbuh : PT. Serasi Media Teknologi.

Oktavia, C., Suheti, T., Husni, A., & Melianingsih, L. (2022). Gambaran Pengetahuan Dan Sikap Remaja Tentang Pencegahan HIV/AIDS. *Jurnal Keperawatan Indonesia Florence Nightingale*, *2*(1), 37–43. https://doi.org/10.34011/jkifn.v2i1.97

Oktaviana, A., Wijaya, D. P., Pramuntadi, A., & Heksaputra, D. (2024). Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (K-NN). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(3), 812–818. https://doi.org/10.57152/malcom.v4i3.1268

Samosir, A., Hasibuan, M. S., Justino, W. E., & Hariyono, T. (2021). Komparasi Algoritma Random Forest, Naive Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung. *Prosiding Seminar Nasional Darmajaya*, 214–222.

Sari, I. D., Rustiana, N., & Damayanti, A. E. (2024). Hubungan Karakteristik Demografi Dengan Pengetahuan Masyarakat Tentang HIV (Human Immunodeficiency Virus)/ Di RW 02 Kelurahan Pinang Ranti Jakarta-Timur. *Jurnal Farmasi IKIFA*, *3*(3), 121–131. https://epik.ikifa.ac.id/jfi/article/view/237

Setiawan, A., Susanto, B., Wijaya, R. W. N., Tita, F., Indrajaya, D., Leipary, H., & Kurniawan, T. A. D. (2025). *Pengantar Data Mining*. Yogyakarta : Diandra Kreatif.