

## Analisa Prediksi Penyakit HIV Menggunakan Algoritma Random Forest

Muhammad Nabilla Fatorohman<sup>1</sup>, Karlena Indriani<sup>2</sup>, Monikka Nur Winnarto<sup>3\*</sup>

<sup>1,2,3</sup>Universitas Bina Sarana Informatika

e-mail: <sup>1</sup>nabilbta26@gmail.com, <sup>2</sup>karlena@bsi.ac, <sup>3</sup>monikka.mnt@bsi.ac.id

Diterima	Direvisi	Disetujui
01-11-2024	06-11-2024	06-12-2024

**Abstrak** - *Human Immunodeficiency Virus* (HIV) adalah infeksi yang menyerang sistem kekebalan tubuh dengan menargetkan sel darah putih, sehingga melemahkan sistem kekebalan. Akibatnya, penderita lebih rentan terhadap penyakit seperti tuberkulosis, infeksi, dan beberapa jenis kanker. HIV ditularkan melalui cairan tubuh orang yang terinfeksi, termasuk darah, ASI, air mani, dan cairan vagina, tetapi tidak menyebar melalui ciuman, pelukan, atau berbagi makanan. Penularan juga dapat terjadi dari ibu ke bayi. Penyakit yang disebabkan oleh HIV, yaitu AIDS, memiliki tingkat kematian yang mendekati 100 persen dan telah memberikan dampak besar pada keluarga serta masyarakat. Penderita AIDS menghadapi tantangan besar dalam menjalani aktivitas harian, bahkan untuk mendapatkan kepastian status kesehatan mereka. Perlu adanya metode mendeteksi HIV secara dini, dengan mengetahui status kesehatannya, mereka memiliki kesempatan untuk merencanakan pengobatan dan perawatan yang diperlukan, serta berkontribusi pada keberlangsungan hidup yang sehat di masyarakat. Penelitian ini bertujuan untuk menganalisis prediksi perkembangan penyakit HIV pada individu menggunakan algoritma *Random Forest* melalui aplikasi RapidMiner. Hasil pengujian data dengan algoritma *Random Forest* menunjukkan akurasi sebesar 92,86%, yang menandakan model prediksi yang baik. Nilai *recall* sebesar 91,43% menunjukkan tingkat keberhasilan sistem dalam mengidentifikasi informasi secara benar, sedangkan presisi sebesar 94,12% menunjukkan ketepatan sistem dalam memberikan informasi yang akurat. *F1-Score* sebesar 86,50% menunjukkan keseimbangan antara presisi dan *recall*, menggambarkan model yang efektif dalam analisis prediksi HIV.

Kata Kunci : HIV, Prediksi, *Random Forest*, RapidMiner

### Abstracts

*Human Immunodeficiency Virus (HIV) is an infection that attacks the immune system by targeting white blood cells, weakening the immune system. As a result, sufferers are more susceptible to diseases such as tuberculosis, infections, and some types of cancer. HIV is transmitted through the body fluids of an infected person, including blood, breast milk, semen, and vaginal fluids, but is not spread through kissing, hugging, or sharing food. Transmission can also occur from mother to baby. The disease caused by HIV, namely AIDS, has a mortality rate approaching 100 percent and has had a major impact on families and communities. AIDS sufferers face major challenges in carrying out daily activities, even to get certainty about their health status. There needs to be a method for detecting HIV early, by knowing their health status, they have the opportunity to plan the necessary treatment and care, and contribute to a healthy life in the community. This study aims to analyze the prediction of HIV disease development in individuals using the Random Forest algorithm through the RapidMiner application. The results of data testing with the Random Forest algorithm showed an accuracy of 92.86%, which indicates a good prediction model. The recall value of 91.43% indicates the success rate of the system in correctly identifying information, while the precision of 94.12% indicates the accuracy of the system in providing accurate information. The F1-Score of 86.50% indicates a balance between precision and recall, describing an effective model in HIV prediction analysis.*

Keywords : HIV, Prediction, *Random Forest*, Rapid Miner

### PENDAHULUAN

*Human immunodeficiency infection* (HIV) merupakan virus yang mengkontaminasi pada sistem kekebalan tubuh. HIV menyerang sel darah putih untuk melemahkan sistem kekebalan tubuh. Sehingga, individu menjadi lebih mudah terkena infeksi penyakit seperti tuberkulosis dan beberapa

jenis kanker. HIV adalah penyakit yang ditularkan melalui cairan tubuh individu yang sudah terinfeksi, seperti darah, air susu ibu, sperma, cairan vagina pada wanita, berciuman, berpelukan dan berbagi makanan. Selain itu, penyakit ini bisa ditularkan melalui ibu kepada anak (WHO, 2023).

Nursalam dan Kurniawati N.D didalam jurnal yang berjudul “Studi Kualitatif Strategi Koping

Penderita HIV AIDS di Kota Bandung” Mengemukakan “HIV dan AIDS adalah penyakit mematikan yang disebabkan oleh infeksi virus *human immunodeficiency virus* yang menyerang sistem kekebalan tubuh dengan menghancurkan sel CD4. Bentuk infeksi HIV yang paling parah dikenal sebagai *acquired immunodeficiency syndrome* (AIDS). Pada tahap ini, kemampuan tubuh untuk melawan infeksi sangat lemah, sehingga sangat rentan terhadap berbagai infeksi. Sebagian besar pasien meninggal akibat komplikasi terkait penyakit ini.” (Salami et al., 2021).

Penyakit yang disebabkan oleh HIV memiliki tingkat kematian yang mendekati 100 persen. AIDS telah melenyapkan keluarga dan menyebabkan menderitanya seseorang yang tak terhitung jumlahnya di wilayah yang paling parah, termasuk wilayah yang miskin di dunia ini, AIDS, namun seringkali situasi ini menjadi penghalang bagi mereka untuk beraktivitas untuk mendapatkan kepastian status kesehatan mereka. Terlepas dari apakah mereka dinyatakan positif atau negatif HIV, pemeriksaan status kesehatan berbasis VCT harus dilakukan. Tes ini dilakukan untuk individu yang siap, sadar atau melakukannya dengan sukarela. Jika ada kemungkinan yang tinggi terhadap status HIV positif, individu tersebut siap untuk mengakui kondisinya dan seharusnya memiliki pilihan untuk benar-benar mencoba yang dianggap sebagai perbaikan untuk dirinya dan orang lain, demi keberlangsungan hidup di masyarakat (Suarnianti & Haskas, 2021).

Menurut *Global HIV Epidemic Report 2019* dari *United Nations Programme on HIV/AIDS* (UNAIDS), 38 juta orang di seluruh dunia hidup dengan HIV pada tahun 2019, dan 7,1 juta orang tidak mengetahui bahwa mereka terinfeksi HIV. Epidemio HIV merupakan masalah global yang besar dan merupakan tantangan kesehatan masyarakat, termasuk di Indonesia. Epidemio HIV merupakan masalah global yang serius dan tantangan kesehatan masyarakat di negara industri dan negara berkembang, termasuk Indonesia.

Suzana Murni berpendapat “Jumlah kasus HIV/AIDS sebagian besar kasus penyebabnya terkait dengan transfusi darah, hubungan seks heteroseksual atau tidak aman, penggunaan narkoba dengan jarum suntik atau tindik yang tidak steril sehingga HIV/AIDS mudah ditularkan.” (Azizah et al., 2024).

Dalam mengolah data prediksi kita bisa memilih metode yang relevan salah satunya adalah algoritma *Random Forest*. Devella, S., Yohannes, Y., dan Rahmawati, F. N. mengemukakan “*Random Forest* adalah algoritma yang menggunakan metode partisi biner rekursif untuk menentukan simpul akhir dalam struktur pohon berdasarkan pohon klasifikasi dan regresi. Algoritma *Random Forest* memiliki beberapa keunggulan, termasuk kemampuan mencapai kesalahan yang relatif kecil, klasifikasi yang baik, efisiensi pemrosesan untuk jumlah data

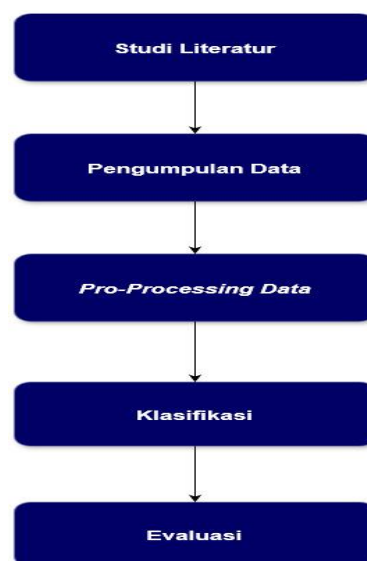
besar, dan metode yang efektif untuk memperkirakan data yang hilang. *Random Forest* menghasilkan beberapa pohon independen, di mana subsetnya dipilih secara acak melalui *bootstrapping* dari sampel yang dilatih dan variabel *input* di setiap *Node*.” (Pamuji & Ramadhan, 2021).

*Random Forest* merupakan algoritma machine learning yang digunakan dalam mengklasifikasikan kumpulan pada data yang besar. Karena kemampuannya, algoritma *Random Forest* bisa diterapkan pada berbagai dimensi serta skala dengan kinerja yang tinggi. *Random Forest* dapat digunakan di berbagai industri seperti perbankan, perawatan kesehatan, analisis keuangan, *e-commerce*, dan lainnya. Algoritma *Random Forest* bekerja dengan menghasilkan beberapa pohon keputusan lalu menggabungkannya untuk menghasilkan prediksi yang lebih stabil dan lebih akurat. Hutan yang dihasilkan oleh algoritma *Random Forest* terdiri dari kumpulan pohon keputusan yang biasanya akan dilatih menggunakan metode *bagging*, yang bertujuan dalam menghasilkan model *machine learning* yang lebih baik (Sudrajat et al., 2022).

Klasifikasi adalah proses pada *Data Mining* memiliki tujuan untuk mencari pola-pola yang digunakan didalam data yang sangat besar dan sangat kecil (Hafizan & Putri, 2020).

Karena latar belakang itu akan dilakukan penelitian dalam menganalisa prediksi penyakit HIV menggunakan algoritma *Random Forest* yang bertujuan untuk mengembangkan model analisa dan menguji keefektifan algoritma *Random Forest* dalam memprediksi penyakit HIV serta memvalidasi hasil prediksi penyakit HIV agar bisa menjadi salah satu langkah dini untuk mencegah penyakit HIV di masyarakat.

## METODE PENELITIAN



Sumber : Hasil Penelitian (2024)

Gambar 1. Tahapan Metode Penelitian

A. Studi Literatur

Tujuan dari Studi Literatur adalah untuk mengkaji penelitian-penelitian sebelumnya untuk mendapatkan teori dan informasi yang relevan yang dapat digunakan dalam penelitian ini. Studi literatur harus relevan dengan HIV dan relevan dengan algoritma *Random Forest*. Studi literatur dilakukan dengan cara mencari informasi dari berbagai literatur buku dan jurnal ilmiah yang berhubungan dengan penyakit HIV dan metode algoritma *Random Forest*.

B. Pengumpulan Data

Dataset yang digunakan untuk prediksi penyakit HIV diambil dari sumber dataset yang sudah terpercaya yaitu Kaggle dengan link <https://www.kaggle.com/datasets/ishigamisenku10/hiv-prediction>. Dari link tersebut yang berjumlah 698 data didalam nya dengan 10 *attribute* yakni *Age*, *Martial Status*, *STD*, *Educational Background*, *HIV Test in Past Year*, *Aids Education*, *Places of Seeking Sex Partners*, *Sexual Orientation*, *Drug Taking* dan *Result*.

C. Pre-Processing Data

*Pro-Processing* data merupakan pengolahan data awal, dimana semua data akan dicek untuk menghindari seperti data-data ganda, data yang tidak sesuai yang bisa menyebabkan eror pada saat pengolahan data dan data-data yang kosong. Lalu pada proses ini isi dari atribut umur akan dikelompokkan yang diubah dari *binominal* menjadi *polynomial*.

D. Klasifikasi

Pada klasifikasi, dataset akan dibagi menjadi dua yaitu data *training* dan data *testing*, dengan perbandingan 90:10 yang merupakan hasil pembagian data dengan akurasi terbaik. Data *training* akan akan untuk membangun model algoritma *Random Forest* secara keseluruhan melalui aplikasi RapidMiner yakni dengan cara setelah memasukan semua operator maka akan digunakan operator *Random Forest* yaitu operator yang akan menghitung dan mengolah otomatis dengan *Result* sebagai label atributnya dan akan mengolah dataset tersebut secara otomatis dengan hitungan algoritma *Random Forest* yang akan menghasilkan beberapa grafik *Tree* beserta deskripsinya.

E. Evaluasi

Selanjutnya akan dilakukan validasi dan pengujian pada data *testing* dengan algoritma *Random Forest* pada RapidMiner untuk membentuk sebuah model *Confusion Matrix* untuk mengetahui tingkat akurasi, presisi, *recall* dan kurva ROC serta menghitung manual *F-1 Score*.

**HASIL DAN PEMBAHASAN**

Penelitian ini menggunakan dataset yang bersumber dari web dataset yang terpercaya yakni

Kaggle. Dataset prediksi HIV yang bersumber dari link

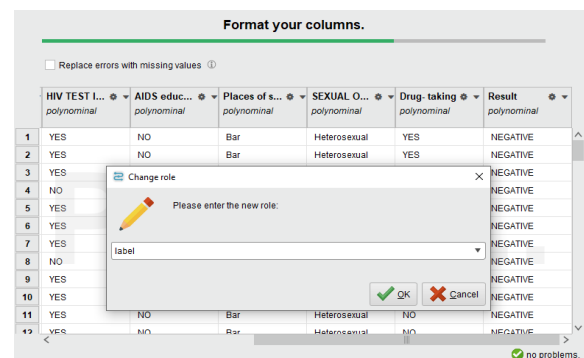
<https://www.kaggle.com/datasets/ishigamisenku10/hiv-prediction> memiliki data yang berjumlah 698 data dan mempunyai 10 *attribute* didalamnya. Berikut keterangan 10 *attribute* dataset prediksi HIV pada tabel berikut.

Tabel 1. *Feature Dataset*

<i>Feature</i>	<b>Keterangan</b>
<i>Age</i>	Pengelompokan Umur yang Terdiri dari Teenager, Adult dan Elderly (Kemenkes RI, 2024)
<i>Marital Status</i>	Status Pernikahan
<i>STD</i>	Status Penyakit Menular Seksual
<i>Educational Background</i>	Latar Belakang Pendidikan
<i>HIV Test in Past Year</i>	Status Tes HIV Dalam Setahun Terakhir
<i>AIDS education</i>	Status Edukasi Tentang AIDS
<i>Places of seeking sex partners</i>	Tempat Mencari Pasangan Sex
<i>SEXUAL ORIENTATION</i>	Orientasi Seksual
<i>Drug- taking</i>	Status Penggunaan Jarum
<i>Result</i>	Hasil Prediksi

Sumber : Hasil Penelitian (2024)

Pada Tabel 1 terlihat ada atribut yang dimodifikasi yaitu *atribut Age*. Pada *atribut Age* dilakukan modifikasi dari *Binominal* menjadi *Polynomial* yang bertujuan untuk mengelompokkan usia berdasarkan kategorinya. Selanjutnya akan dilakukan *Pre-Processing* data sebagai berikut:



Sumber : Hasil Penelitian (2024)

Gambar 2. Pre-Processing Data

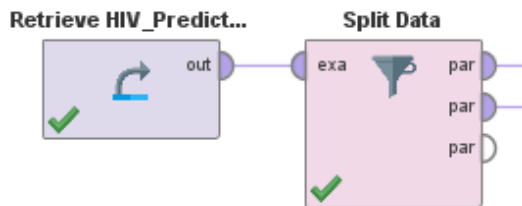
Pada proses pengolahan data di aplikasi RapidMiner, RapidMiner akan secara otomatis melakukan *Pre-Processing* data untuk mendeteksi data yang kosong, isi kolom yang berbeda dan akan otomatis memberi centang *no problems* jika data sudah sesuai, serta

pada tahap ini juga atribut *Result* akan dijadikan label untuk penentuan keputusan pada prediksi penyakit HIV.

#### A. Pemodelan Random Forest

Pemodelan algoritma *Random Forest* dilakukan oleh *data training* yang bertujuan untuk melatih data training serta membangun model dari algoritma *Random Forest* itu sendiri. Berikut adalah tahapan dari pemodelan tersebut.

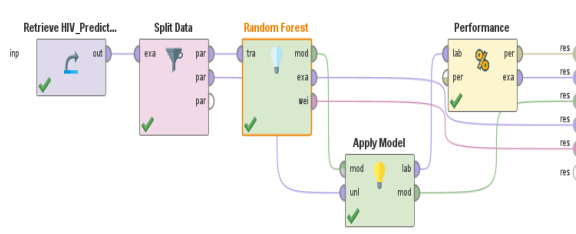
##### 1. Splitting Data



Sumber : Hasil Penelitian (2024)  
Gambar 3. Penggunaan Operator *Split Data*

Pada tahap ini dataset akan dibagi menjadi dua yaitu data *training* dan data *testing* dengan perbandingan 90:10 dan memakai *Shuffled Sampling*. Data *training* akan melatih data untuk membangun sebuah model algoritma *Random Forest*, sedangkan data *testing* untuk melakukan pengujian dan evaluasi untuk mengetahui akurasi, presisi, *recall*, *F1-Score* dan kurva ROC.

##### 2. Klasifikasi Data



Sumber : Hasil Penelitian (2024)  
Gambar 4. Penggunaan Operator Untuk Klasifikasi Data

Pada klasifikasi data, RapidMiner akan membangun sebuah model *Random Forest* dengan data *training* yaitu dengan membuat 100 *tree* atau pohon keputusan dan *testing* akan digunakan untuk pengujian dan evaluasi data melalui operator *performance*.

##### 3. Model Random Forest

Model melalui rumus *Decision Tree* secara acak melalui data *training* yang menghasilkan 100 pohon keputusan untuk menghasilkan evaluasi yang lebih baik.

#### 4. Attribute Weights Random Forest

attribute	weight
STD	0.135
AIDS education	0.095
Places of seeking sex partners	0.048
Marital Staus	0.142
Educational Background	0.141
HIV TEST IN PAST YEAR	0.103
Age	0.145
Drug- taking	0.075
SEXUAL ORIENTATION	0.115

Sumber : Hasil Penelitian (2024)  
Gambar 5. Attribute Weights Random Forest

Pada tahap ini merupakan hasil optimisasi 9 *atribut* dari dataset prediksi HIV melalui algoritma *Random Forest* yang bisa meningkatkan hasil pengujian. *Atribut Age* mendapatkan hasil optimisasi paling besar untuk penyesuaian bobot atau kepentingan relatif dari berbagai *attribute*.

#### B. Evaluasi data

Pada evaluasi data akan menggunakan aplikasi RapidMiner dan perhitungan secara manual untuk menguji *Confusion Matrix* dan melihat grafik kurva ROC pada data yang sudah diklasifikasi dan dimodelkan melalui algoritma *Random Forest*.

##### 1. Pengujian Model

Pengujian model dilakukan untuk mengetahui hasil dari *Confusion Matrix* dari dataset yang sudah dimodelkan melalui algoritma *Random Forest*, seperti tingkat akurasi, presisi, *recall* dan perhitungan manual *F1-Score*, berikut adalah hasil pengujian model dari Rapid Miner.

accuracy: 92.86%

	true NEGATIVE	true POSITIVE	class precision
pred. NEGATIVE	33	3	91.67%
pred. POSITIVE	2	32	94.12%
class recall	94.29%	91.43%	

Sumber : Hasil Penelitian (2024)  
Gambar 6. Hasil Akurasi

precision: 94.12% (positive class: POSITIVE)

	true NEGATIVE	true POSITIVE	class precision
pred. NEGATIVE	33	3	91.67%
pred. POSITIVE	2	32	94.12%
class recall	94.29%	91.43%	

Sumber : Hasil Penelitian (2024)  
Gambar 7. Hasil Presisi

recall: 91.43% (positive class: POSITIVE)

	true NEGATIVE	true POSITIVE	class precision
pred. NEGATIVE	33	3	91.67%
pred. POSITIVE	2	32	94.12%
class recall	94.29%	91.43%	

Sumber : Hasil Penelitian (2024)  
Gambar 8. Hasil Recall

Pada gambar diatas merupakan hasil prediksi dan *true* yang dihasilkan oleh pemodelan algoritma *Random Forest*. Berikut adalah perhitungan manual untuk *Confusion Matrix*:

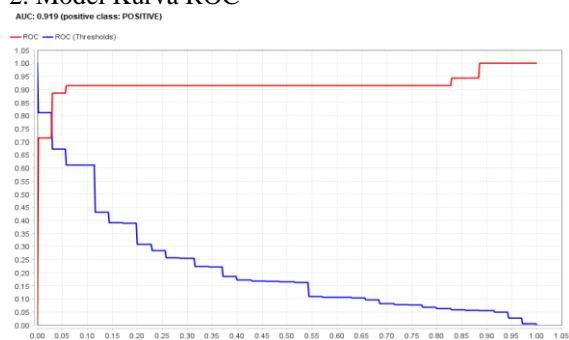
$$\begin{aligned} \text{Akurasi} &= (TP+TN) / (TP+FP+FN+TN) = (32+33) / (32+2+3+33) = 0,92857 \\ \text{Presisi} &= TP / (TP+FP) = 32 / (32+2) = 0,94117 \\ \text{Recall} &= TP / (TP+FN) = 32 / (32+3) = 0,91428 \\ \text{F1-Score} &= 2 * TP / 2 * (TP+FP+FN) = 2 * 32 / 2 * (32+2+3) = 0,86486 \end{aligned}$$

Sumber : Hasil Penelitian (2024)

Gambar 9. Perhitungan *Confusion Matrix*

Bedasarkan hitungan manual diatas, akurasi, *recall* dan presisi didapatkan sesuai dengan pengujian pada RapidMiner. Akurasi mendapatkan 92,86 % yang berarti model menunjukkan dengan baik nilai dengan tingkat kedekatan antara nilai prediksi sistem dengan nilai prediksi manusia, pada *recall* mendapatkan 91,43% yang berarti model dengan baik menunjukkan tingkat keberhasilan atau spesifitas untuk mengetahui kembali sebuah informasi secara benar tentang data yang kelas *negative* maupun *positive*, sedangkan presisi mendapatkan 94,12 % yang berarti model dengan baik menunjukkan ketepatan sistem antara informasi yang diberikan oleh sistem untuk menunjukkan secara benar data kelas *negative* atau kelas *positive* dan yang terakhir *F1-Score* mendapatkan hasil sekitar 86,50 % yang berarti model dengan cukup baik menunjukkan gambaran keseimbangan antara presisi dan *recall*.

## 2. Model Kurva ROC



Sumber : Hasil Penelitian (2024)

Gambar 10. Kurva ROC

Dari gambar diatas didapatkan hasil dari kurva ROC yaitu garis merah yang merupakan kurva ROC menunjukkan garis semakin tinggi diatas diagonal yang menyisakan ruang sedikit diatasnya dan menyisakan ruang yang banyak dibawahnya, sehingga pada *Area Under Curve* atau AUC menghasilkan akurasi sebesar 91,9 %. Pada garis biru didalam kurva ROC merupakan garis referensi atau baseline yang menunjukkan kinerja jika model klasifikasi dilakukan secara acak yang berfungsi untuk membandingkan kinerja model klasifikasi yang telah dibuat. Jika garis merah lebih rendah daripada garis biru maka model klasifikasi yang telah dibuat

dikatakan gagal karena performanya sama dengan tebakan acak.

Berdasarkan keterangan diatas, garis merah yang merupakan kurva ROC jauh berada diatas garis biru sehingga diketahui mengasilkan kinerja klasifikasi yang maksimal dengan akurasi sebesar 91,9%, ini menandakan bahwa akurasi tersebut termasuk dalam kategori *Excellent classification* karena pada perbandingan *True Positive Rate* dan *False Positive Rate* didalamnya, *True Positive Rate* mendapatkan hasil yang lebih besar daripada *False Positive Rate*.

## KESIMPULAN

Penelitian ini mengembangkan dan mengevaluasi model prediksi penyakit HIV menggunakan algoritma *Random Forest*, yang menunjukkan kinerja unggul dengan akurasi 92,86%, presisi 94,12%, *recall* 91,43%, dan *F1-score* 86,50%. *Random Forest* terbukti lebih efektif dibandingkan metode konvensional dalam mengidentifikasi faktor risiko HIV, memberikan panduan keputusan yang berguna untuk strategi pencegahan dan pengelolaan HIV. Validasi dengan dataset luas memastikan generalisasi hasil, sehingga temuan ini berkontribusi signifikan terhadap pemahaman dan pengelolaan HIV. Penelitian juga mengungkap keterkaitan antara berbagai faktor risiko, memberikan wawasan yang lebih holistik tentang HIV. Hasil ini menyediakan dasar untuk penelitian lanjutan yang dapat memperluas analisa prediksi HIV dengan metode yang lebih efektif atau dataset yang lebih besar, menunjukkan bahwa *Random Forest* adalah dapat menjadi salah satu metode efektif dalam pencegahan dan pengelolaan HIV.

## REFERENSI

- Alpian, D., Krisna, N., & Salamah, U. (2022). Perbandingan Algoritma Naïve Bayes Dan K-Nearest Neighbor Untuk Klasifikasi Berita Hoax Kesehatan Di Media Sosial Twitter. *Jurnal Teknik Informatika Kaputama (JTIK)*, 6(2).
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640. <https://doi.org/10.30865/mib.v5i2.2937>
- Azizah, R. N., Nisak, U. K., & indahyanti, U. (2024). Analisis Jumlah Prediksi Penyebaran HIV/AIDS di Kabupaten Sidoarjo menggunakan Metode Multiple Linier Regression. *Physical Sciences, Life Science and Engineering*, 1(1), 11. <https://doi.org/10.47134/pslse.v1i1.163>
- Buaton, R., Mawengkang, H., Zarlis, M., Effendi, S., Manaoor Hara Pardede, A., Maulita, Y., Fauzi, A., & Novriyenni, N. (2019). Time Series

- Optimization on Data Mining. *Journal of Physics: Conference Series*, 1235(1), 012014. <https://doi.org/10.1088/1742-6596/1235/1/012014>
- Dharmawan, W. S. (2022). Komparasi Algoritma Klasifikasi Svm-Pso Dan C4.5-Pso Dalam Prediksi Penyakit Jantung. *INFORMATIKA A*, 13(2), 31. <https://doi.org/10.36723/juri.v13i2.301>
- Ferdinan Leo Simanjuntak, K., Carolina Br Barus, A., & Teknologi dan Ilmu Komputer, F. (2021). Implementasi Metode Decision Tree Dan Algoritma C4.5 Untuk Klasifikasi Kepribadian Masyarakat. *Joisie Journal Of Information System And Informatics Engineering*, 5(1), 51–59.
- Hadi, D. A., & Sirodj, D. A. N. (2023). Metode Random Forest untuk Klasifikasi Penyakit Diabetes. *Bandung Conference Series: Statistics*, 3(2), 428–435. <https://doi.org/10.29313/bcss.v3i2.8354>
- Hafizan, H., & Putri, A. N. (2020). Penerapan Metode Klasifikasi Decision Tree Pada Status Gizi Balita Di Kabupaten Simalungun. *KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, 1(2), 68–72. <https://doi.org/10.30645/kesatria.v1i2.23>
- Junaedy, F. Z., Hidayat Insani, R., Santoso, I., Tinggi, S., Komputer, I., Karya Informatika, C., & Jakarta, U. (2023). Komparasi Algoritma Support Vector Machine Dan Naïve Bayes Pada Analisis Sentimen Formula-E Jakarta Tahun 2022. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7. <https://youtu.be/mhAEiHuVFxY>
- Kuncoro, A. T., & Kalifia, A. D. (2024). Analisis Pengaruh Judi Online Terhadap Kesehatan Mental Dengan Menggunakan Visual RapideMiner. *Jurnal Ilmiah Multidisiplin*, 1(3), 189–193. <https://doi.org/10.62017/merdeka>
- M. Fahrul Rizki Aditya, Uce Indahyanti, & Nuril Lutvi Azizah. (2024). Prediksi Penyakit Hipertensi Menggunakan Metode Decision Tree dan Random Forest. *Jurnal Ilmiah Komputasi*, 23(1). <https://doi.org/10.32409/jikstik.23.1.3503>
- Pamuji, F. Y., & Ramadhan, V. P. (2021). Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy. *Jurnal Teknologi Dan Manajemen Informatika*, 7(1), 46–50. <https://doi.org/10.26905/jtmi.v7i1.5982>
- Partogi, Y., & Pasaribu, A. (2022). Perancangan Metode Decision Tree Terhadap Sistem Perpustakaan STMIK Kuwera. *Jurnal Sistem Informasi Dan Teknologi (SINTEK)*, 1(2), 20–25. <https://doi.org/10.56995/sintek.v1i2.4>
- Rayhan Giffary, D., Aditya Darma, T., & Naufal Manaf, R. (2023). Pola Pendidikan Islam dalam Upaya Menurunkan Angka Penyebaran HIV/AIDS pada Remaja di Indonesia. *Journal of Creative Student Research (JCSR)*, 1(5), 384–392. <https://doi.org/10.55606/jcsrpolitama.v1i5>
- Rifkie Primartha. (2021). *Algoritma Machine Learning*. Informatika .
- Said, H., Matondang, N. H., & Irmanda, H. N. (2022). Penerapan Algoritma K-Nearest Neighbor Untuk Memprediksi Kualitas Air Yang Dapat Dikonsumsi. *Techno.Com*, 21(2), 256–267. <https://doi.org/10.33633/tc.v21i2.5901>
- Salami, S., Muvira, A. A., & Yualita, P. (2021). Studi Kualitatif Strategi Koping Penderita HIV AIDS di Kota Bandung. *Faletehan Health Journal*, 8(01), 22–30. <https://doi.org/10.33746/fhj.v8i01.243>
- Suarnianti, S., & Haskas, Y. (2021). Penguatan Intervensi Perilaku terhadap Pencegahan HIV pada Kelompok Berisiko: Sistematis Review. *Jurnal Kesehatan Andalas*, 9(4), 439. <https://doi.org/10.25077/jka.v9i4.1431>
- Sudrajat, D., Purnamasari, A. I., Dikananda, A. R., Kurnia, D. A., & Bahtiar, A. (2022). Klasifikasi Mutu Pembelajaran Hybrid berdasarkan Algoritma C.45, Random Forest dan Naïve Bayes dengan Optimasi Bootsrap Areggating (Bagging) pada masa COVID-19. *JURIKOM (Jurnal Riset Komputer)*, 9(6), 2227. <https://doi.org/10.30865/jurikom.v9i6.5179>
- Suryadi, L., Ngajiyanto, N., Pratiwi, N. E., Ardhy, F., & Riswanto, P. (2022). Penerapan Data Mining Prediksi Penjualan Mebel Terlaris Menggunakan Metode K-Nearest Neighbor(K-Nn) (Studi Kasus : Toko Zerita Meubel). *JUSIM (Jurnal Sistem Informasi Musirawas)*, 7(2), 174–184. <https://doi.org/10.32767/jusim.v7i2.1697>
- Trivusi. (2022, August 16). *Data Splitting: Pengertian, Metode, dan Kegunaannya* . Trivusi.
- Tukino, T. (2019). Penerapan Algoritma C4.5 Untuk Memprediksi Keuntungan Pada PT SMOE Indonesia. *Jurnal Sistem Informasi Bisnis*, 9(1), 39. <https://doi.org/10.21456/vol9iss1pp39-46>
- Tuntun, R., Kusriani, K., & Kusnawi, K. (2022). Analisis Perbandingan Kinerja Algoritma Klasifikasi dengan Menggunakan Metode K-Fold Cross Validation. *Jurnal Media Informatika Budidarma*, 6(4), 2111. <https://doi.org/10.30865/mib.v6i4.4681>
- WHO. (2023, July 13). *HIV and AIDS*. WHO. <https://www.who.int/news-room/fact-sheets/detail/hiv-aids>