
Prediksi Pasien Terkena Penyakit *HIV* dengan Algoritma *K-Nearest Neighbor (KNN)* di RSUD Dr. Chasbullah Abdul Madjid Kota Bekasi

Abdul Aziz Firdaus¹, Sanudin Sanudin², Aceng Badruzzaman³

¹²³ Universitas Pelita Bangsa

Email: abdul.aziz03@mhs.pelitabangsa.ac.id¹, sanudin@pelitabangsa.ac.id²,
aceng_badruzzaman@pelitabangsa.ac.id³

Diterima	11	April	2025
Disetujui	13	Mei	2025
Dipublish	02	Juni	2025

Abstract

HIV/AIDS is a global health concern present in almost every part of the world. The advancement of information system technology has helped solve various problems across different fields, one of which is through the application of data mining. The utilization of data mining is not only implemented in the healthcare sector but also in the technology industry. One method to predict patients who potentially have *HIV/AIDS* is by using Machine Learning (ML). ML aims to train models with algorithms capable of performing statistical analysis using Supervised Learning techniques to generate accurate predictions. Prediction is one of the most important statistical elements in the decision-making process. This research uses the *K-Nearest Neighbor* algorithm, which classifies data based on the majority class of *K* nearest neighbors. The algorithm is combined with the *SMOTETomek* technique as a resampling method to address data containing noise and class imbalance problems. The dataset used to train the *K-Nearest Neighbor* model comes from the Voluntary Counselling and Testing (VCT) unit with a total of 2,205 data points. The disease testing prediction results are then processed and visualized in a website format. Based on testing conducted using Confusion Matrix, the model's performance measurement results show and Accuracy value of 97.96%, Precision of 78.61%, Recall of 98.88%, and f1-score of 84.45%. The results indicate that the use of machine learning is quite effective for implementation in *HIV/AIDS* disease prediction.

Keywords: *Prediction, HIV/AIDS, Machine Learning, K-Nearest Neighbor, Website.*

Abstrak

Penyakit *HIV/AIDS* merupakan masalah kesehatan global yang terdapat di hampir seluruh dunia. Perkembangan teknologi sistem informasi telah membantu menyelesaikan berbagai masalah di berbagai bidang, salah satunya melalui penerapan *data mining*. Pemanfaatan *data mining* tidak hanya digunakan di sektor kesehatan tetapi juga di industri teknologi. Salah satu metode untuk memprediksi pasien yang berpotensi terkena penyakit *HIV/AIDS* adalah dengan menggunakan *Machine Learning (ML)*. *ML* bertujuan untuk melatih model dengan algoritma yang mampu melakukan analisis statistik menggunakan teknik *Supervised Learning* untuk menghasilkan prediksi yang akurat. Prediksi merupakan salah satu unsur statistika yang sangat penting dalam proses pengambilan keputusan. Penelitian ini menggunakan algoritma *K-Nearest Neighbor* yang mengklasifikasikan data berdasarkan kelas mayoritas dari *K* tetangga terdekat. Algoritma tersebut dikombinasikan dengan teknik *SMOTETomek* sebagai metode *resampling* untuk mengatasi data yang mengandung *noise* dan masalah ketidakseimbangan kelas. Dataset yang digunakan untuk melatih model *K-Nearest Neighbor* berasal dari unit *Voluntary Counselling and Testing (VCT)* dengan total 2.205 data. Hasil prediksi pengujian penyakit kemudian diolah dan divisualisasikan dalam bentuk *website*. Berdasarkan pengujian yang



dilakukan menggunakan *Confusion Matrix*, hasil pengukuran performa model menunjukkan nilai *Accuracy* sebesar 97.96%, *Precision* sebesar 78.61%, *Recall* sebesar 98.88%, dan *f1-score* sebesar 84.45%. Hasil ini menunjukkan bahwa penggunaan *machine learning* cukup efektif untuk diimplementasikan dalam prediksi penyakit *HIV/AIDS*.

Kata kunci: Prediksi, HIV/AIDS, Machine Learning, K-Nearest Neighbor, Website.

Pendahuluan

Kemajuan teknologi informasi telah memberikan dampak signifikan dalam bidang medis, menghasilkan perkembangan pesat di sektor administrasi rumah sakit, perawatan medis, dan penelitian ilmu kesehatan. Meskipun ada minat global yang besar terhadap perawatan kesehatan berbasis teknologi karena potensinya meningkatkan kesejahteraan manusia, implementasi yang efektif memerlukan pemberdayaan para profesional kesehatan dan keterlibatan aktif seluruh subsistem di sektor kesehatan. Untuk mewujudkan sistem layanan kesehatan yang adil, berfokus, dan terjangkau, diperlukan komitmen serius dan kerja sama antara pemerintah pusat dan daerah, sehingga dapat mengarah pada prosedur kesehatan yang lebih tepat di masa depan (Zahra et al., 2021).

HIV/AIDS telah menjadi masalah kesehatan global yang kritis, dengan peningkatan kasus dari 36,6 juta pada tahun 2002 menjadi 39,4 juta pada tahun 2004, termasuk 8,2 juta kasus di Asia. Di Indonesia, virus ini telah menyebar ke 80% provinsi dan kota sejak pertama kali teridentifikasi di Bali tahun 1987, dengan data terbaru dari Kementerian Kesehatan menunjukkan 10.203 infeksi baru antara juli hingga september 2014, sementara di Kabupaten Tasikmalaya pada 2021 tercatat 160 pengidap dengan 34 kematian. Peningkatan kasus di kalangan remaja, yang diperparah oleh pergaulan bebas dan minimnya pendidikan tentang kesehatan reproduksi di sekolah, menjadi keprihatinan serius meskipun virus ini dapat ditransmisikan melalui darah, air susu ibu, dan kontak seksual serta dapat dihambat dengan obat *antiretroviral*. Faktor budaya yang masih

menganggap pendidikan seks sebagai hal tabu menyebabkan 10-12% remaja di Indonesia masih kurang memiliki pengetahuan seksual yang memadai, menunjukkan bahwa inisiatif sosialisasi tentang perilaku penularan *HIV/AIDS* belum cukup efektif meningkatkan pemahaman mereka (S. Rohmah, 2024).

Hal ini sejalan dengan penelitian (Ismayati et al., 2023) bahwasannya berbagai cara pencegahan dan pengobatan telah banyak dilakukan oleh pemerintah Indonesia dalam upaya penurunan kasus *HIV/AIDS* di Indonesia, khususnya terhadap remaja, yaitu melalui peningkatan pengetahuan yang diharapkan dapat berdampak pada tumbuhnya perilaku yang mampu menghindari tertularnya *HIV/AIDS*.

Penularan *HIV/AIDS* di kelompok remaja dipengaruhi oleh minimnya akses terhadap keakuratan dan kebenaran informasi. Karena itu, edukasi seksual menjadi komponen penting yang mampu membantu para remaja dalam mengatasi berbagai tantangan hidup yang berakar dari dorongan seksual mereka. Pemahaman mengenai cara penularan dan pencegahan *AIDS* merupakan alat utama dalam memerangi *HIV/AIDS*, mengingat belum tersedianya pengobatan atau vaksin yang efektif untuk menghentikan penyebaran infeksi ini (Gusrianti et al., 2023). Mayoritas kasus *HIV/AIDS* di Kabupaten Bekasi, menurut data statistik tahun 2021, terjadi pada kelompok usia produktif, terutama pada rentang usia 25-49 tahun (169 kasus), diikuti oleh kelompok usia 20-24 tahun (56 kasus), dan remaja (14 kasus). Kasus laki-laki lebih banyak dibandingkan perempuan, menurut temuan studi oleh (Juhaefah et al., 2020).



Provinsi Jawa barat menempati posisi kedua dengan jumlah kasus *HIV/AIDS* tertinggi di Indonesia, dengan RSUD Kota Bekasi berfungsi sebagai fasilitas rujukan utama untuk area bekasi dan sekitarnya. Berdasarkan pernyataan Kusnanto Saidi selaku Kepala Dinas Kesehatan Kota Bekasi, data hingga oktober 2016 mencatat sebanyak 3.434 individu terdiagnosis HIV dan 1.206 orang dengan *AIDS* di RSUD Kota Bekasi. Distribusi faktor risiko penularan meliputi hubungan seksual sesama jenis (5%), hubungan seksual beda jenis (42,9%), orientasi seksual ganda (5%), transmisi dari ibu ke anak (4%), penggunaan jarum suntik tidak steril (44,5%), pembuatan tato (0,1%), dan berbagai faktor lainnya (Aprianingsih et al., 2021).

Remaja laki-laki merupakan mayoritas pasien *HIV*, dengan rasio 30,8% untuk perempuan dan 69,2% untuk laki-laki. Berbeda dengan perempuan, yang lebih mungkin tertular dari pasangan seksualnya, banyak remaja laki-laki yang terlibat dalam hubungan seksual yang berbahaya, menggunakan narkoba suntik, dan LSL (lelaki yang berhubungan seks dengan lelaki). Meningkatnya angka *HIV/AIDS* dan hilangnya masa produktif mereka dan mengidap penyakit ini akan berdampak pada hilangnya usia produktif di Indonesia jika permasalahan yang dihadapi remaja tidak segera diatasi (Pangaribuan et al., 2021).

Penggunaan *data mining* adalah salah satu dari sekian banyak masalah yang telah diselesaikan dengan kemajuan teknologi sistem informasi. Praktik mengidentifikasi pola atau informasi berharga lainnya dalam kumpulan data yang sangat besar dan rumit dikenal sebagai *data mining*. *KNN* adalah salah satu teknik *data mining* yang berhubungan dengan klasifikasi. Menemukan nilai data *K* yang paling dekat dari data yang akan diklasifikasikan adalah bagaimana metode ini beroperasi. Data tersebut kemudian dikategorikan sesuai dengan kelas

mayoritas dari data yang paling dekat (Cahyanti et al., 2020).

Dalam penelitian sebelumnya yang berjudul “Optimasi Algoritma *K-Nearest Neighbor (KNN)* Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver”(Zulaikhah et al., 2022), hasil penelitian tersebut menunjukkan bahwa penerapan *machine learning* dengan teknik pemodelan *decision tree* menghasilkan akurasi yang sangat baik. Dalam situasi ini, para pengambil keputusan dapat mencoba meramalkan jumlah kasus penyakit sebagai simulasi untuk membuat tujuan dalam mencapai tingkat pelayanan kesehatan yang ada. Praktik pengambilan keputusan berdasarkan data tidak diragukan lagi dan konsisten dengan inisiatif untuk transformasi digital.

Penelitian kedua, berjudul “Penerapan Metode *Fuzzy K-Nearest Neighbor* Pada Klasifikasi Penyakit Menular Seksual Pria” (Siburian et al., 2020) menggunakan prosedur inisialisasi *fuzzy* sebagai komponen kunci dalam proses klasifikasi. Metode penelitian membandingkan antara pendekatan *fuzzy KNN* dan algoritma *KNN* klasik. Melalui validasi silang dengan 10 kali pengujian terhadap 60 dataset, penelitian menemukan nilai akurasi optimal 91,67% ketika parameter *K* bernilai 5. Selanjutnya, evaluasi performa dilakukan menggunakan *confusion matrix* untuk menganalisis *precision* dan *recall* pada sampel uji sebanyak 30 data.

Penelitian ketiga, berjudul “Penerapan Algoritma *K-Nearest Neighbor (KNN)* Untuk Memprediksi Penyakit Stroke”, dilakukan oleh (Maskuri et al., 2022.). Algoritma *KNN* dapat memprediksi penyakit stroke berdasarkan faktor-faktor seperti jenis kelamin, usia, hipertensi, riwayat penyakit jantung, status pernikahan, jenis pekerjaan, jenis tempat tinggal, rata-rata kadar glukosa,



indeks masa tubuh, dan status sebagai perokok. Hal ini ditunjukkan oleh hasil penelitian prediksi klasifikasi pada data penyakit stroke, yang mencakup 80 data pelatihan dan 20 data uji. Algoritma ini mencapai tingkat akurasi 95% dengan nilai $k = 9$.

Studi ini bertujuan untuk menilai kerja algoritma *K-Nearest Neighbor (KNN)* dalam mengklasifikasikan risiko penyakit *HIV/AIDS* melalui sistem berbasis web, dengan fokus pada peningkatan akurasi diagnosis dan efisiensi pengelolaan data medis. Manfaat utama penelitian ini adalah memberikan alat analisis yang dapat membantu tenaga medis mengidentifikasi tingkat risiko pasien secara lebih akurat, mengoptimalkan proses klasifikasi data, dan mendukung pengambilan keputusan klinis dengan memanfaatkan teknologi *machine learning*, sehingga pada akhirnya dapat meningkatkan kualitas pelayanan kesehatan dalam penanganan kasus *HIV/AIDS*.

Metode Penelitian

A. Objek Penelitian

Objek penelitian yaitu didapatkan dari unit *Voluntary Counselling and Testing (VCT)* terhadap penyakit *HIV/AIDS* di Rumah Sakit Dr.Chasbullah Abdul Madjid Kota Bekasi.

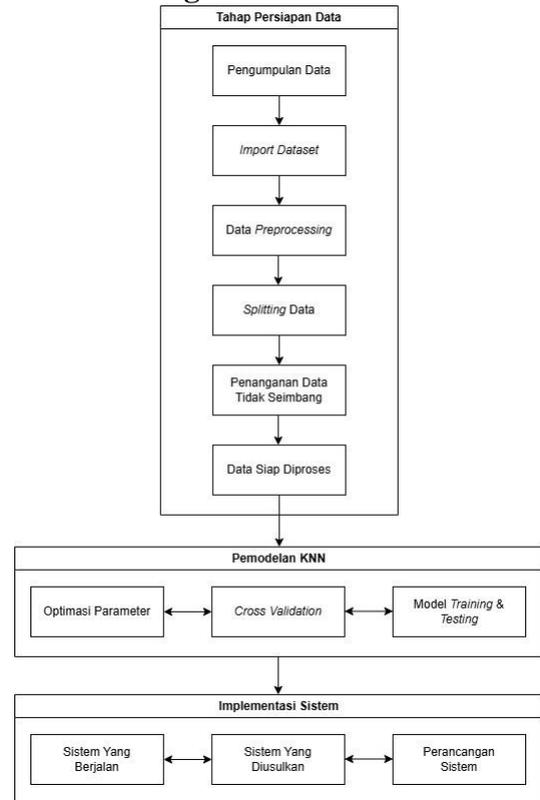
B. Jenis Data

Sumber informasi yang digunakan dalam penelitian. Data primer merupakan sumber data yang diperlukan untuk penelitian ini. Data primer adalah informasi yang langsung dikumpulkan oleh peneliti dari sumber data. *Focus Group Discussion (FGD)*, observasi, dan wawancara adalah beberapa metode utama yang digunakan untuk mengumpulkan data.

C. Metode Yang Digunakan

Dalam penelitian ini, penulis melakukan beberapa tahapan. Diagram alir tahapan penelitian dapat ditunjukkan pada gambar sebagai berikut ini:

Gambar 1. Diagram Alir Penelitian



Sumber: Diolah oleh peneliti (2025)

1. Tahap Persiapan Data

a. Pengumpulan Data

Dataset awal penelitian *HIV/AIDS* terdiri dari 2.205 data dan 9 atribut. Pada data tersebut terdapat data kosong dan atribut tanggal yang tidak dapat dianalisis. Peneliti melakukan pembersihan data secara manual menggunakan *Microsoft Excel*. Proses pembersihan dilakukan dengan menghapus tiga atribut tidak relevan; Tanggal Konfirmasi *HIV*, Tanggal Masuk Perawatan, dan Tanggal Mulai ART. Setelah pembersihan, dataset disederhanakan menjadi 5 atribut yang dapat digunakan untuk analisis, contohnya sebagai tabel dibawah ini:

Tabel 1. Atribut Yang Digunakan

Atribut	Deskripsi
Umur	Usia pasien kunjungan
Jenis Kelamin	Jenis kelamin pasien



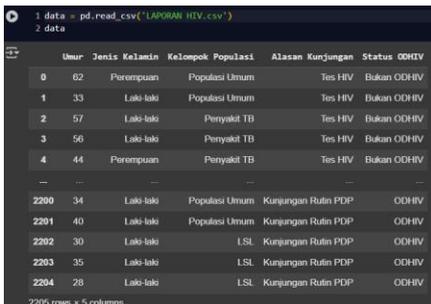
	(laki-laki atau perempuan)	
Kelompok Populasi	Sekelompok individu dengan risiko atau prevalensi <i>HIV/AIDS</i> berdasarkan karakteristik tertentu.	
Alasan Kunjungan	Motif atau tujuan pasien melakukan kunjungan, seperti pemeriksaan kesehatan, konsultasi, atau pengobatan.	
Status <i>ODHIV</i>	Hasil status pengecekan apakah pasien tersebut teridentifikasi terkena penyakit <i>HIV/AIDS</i> .	

Sumber: Diolah oleh peneliti (2025)

b. *Import Dataset*

Dataset penelitian pada awalnya dimasukkan ke dalam *Google Colab* dengan nama "LAPORAN HIV.csv". setelah mengimpor dataset, kolom-kolom yang akan digunakan dan ditampilkan pada sistem dipilih. Import dataset ditunjukkan pada gambar dibawah ini:

Gambar 2. *Import Dataset*



Sumber: Diolah oleh peneliti (2025)

Dalam penelitian ini, menganalisis dataset menggunakan 5 kolom data yaitu Umur, Jenis Kelamin, Kelompok Populasi, Alasan Kunjungan, dan Status *ODHIV*.

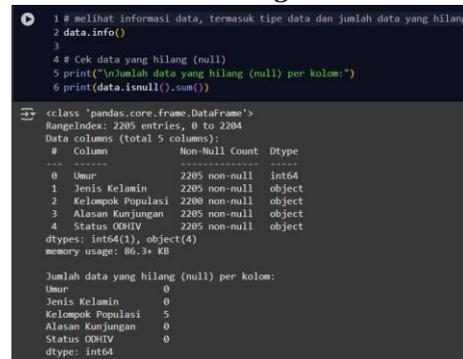
c. *Data Preprocessing*

Data Preprocessing yang merupakan pembersihan dan pemrosesan data (Agung et al., 2023). Pada tahapan ini sebelum masuk ketahap modeling dataset terlebih dahulu diproses melalui tahap-tahap data *cleaning*, data *labelencoder*.

1) *Data Cleaning*

Pada proses ini dilakukan pengecekan missing value, memastikan data tidak ada yang hilang. Cek missing value ditunjukkan pada gambar dibawah ini:

Gambar 3. *Data Cleaning*



Sumber: Diolah oleh peneliti (2025)

Dataset memiliki 2.205 baris dengan 5 kolom, dimana hanya kolom Kelompok Populasi yang memiliki 5 data yang hilang (*null*).

2) *Data LabelEncoder*

Pada proses *preprocessing* menggunakan *LabelEncoder*, pada *machine learning* data seperti ini tidak dapat diproses karena kolom labelnya yaitu berbentuk *string* atau termasuk kategorikal data, maka harus diubah menjadi data numerik dengan label encoder pada data Y yaitu *Train_Y* dan *Test_Y*. *LabelEncoder* adalah berfungsi untuk mengubah setiap nilai dalam kolom menjadi angka berurutan (Latifah et al., 2019).

Contohnya:

$$\begin{aligned} \text{Negative} &= 0 \\ \text{Positive} &= 1 \end{aligned}$$

Proses codingan untuk mengubah kategorikal ke numerikal:

Gambar 4. *Proses Code LabelEncoder*



Sumber: Diolah oleh peneliti (2025)



Hasil data *labelencoder* sebagai berikut ini:

Gambar 5. Hasil Data LabelEncoder

A	B	C	D	E	F
No	Umur	Jenis Kelamin	Kelompok Populasi	Alasan Kunjungan	Status ODHIV
1	62	1	21	4	1
2	33	0	21	4	1
3	57	0	17	4	1
4	56	0	17	4	1
5	44	1	17	4	1
6	61	0	21	4	1
7	65	0	21	4	1
8	51	1	21	4	1
9	19	1	4	6	1
10	25	1	4	6	1

Sumber: Diolah oleh peneliti (2025)

d. *Splitting Data*

Setelah itu, data dibagi menjadi data *training* dan *testing*. 20% dari data digunakan untuk *testing*, sedangkan 80% sisanya digunakan untuk *training*. X_{train} , X_{test} , Y_{train} , dan Y_{test} adalah empat komponen utama yang dihasilkan dalam prosedur ini. Sementara X_{test} digunakan untuk *testing*, X_{train} terdiri dari variabel independen atau fitur yang digunakan untuk pelatihan. Sementara itu, Y_{test} digunakan pada tahap *testing*, dan Y_{train} adalah variabel dependen atau target yang digunakan pada tahap *training*. Jumlah data X dan Y setelah operasi pembagian data adalah 2.205. Dari jumlah ini, 1.764 data (atau 80% dari total) digunakan untuk pelatihan (X_{train} dan Y_{train}). Tabel Jumlah presentase data *training* dan *testing* sebagai berikut ini:

Tabel 2. Jumlah Presentase Data Training dan Testing

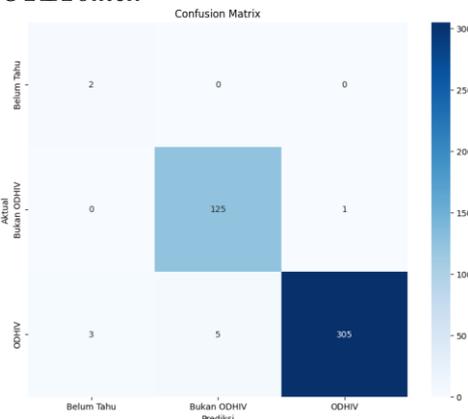
	Presentase	Jumlah X	Jumlah Y
Data <u>Training</u>	80%	1764	441
Data <u>Testing</u>	20%	1764	441

Sumber: Diolah oleh peneliti (2025)

e. Penanganan Data Tidak Seimbang
 Penanganan data yang tidak seimbang terjadi ketika pada tahap *confusion matrix*, label Status ODHIV yang memiliki nilai “Belum Tahu” memiliki data yang tidak seimbang. Oleh karena itu diperlukan proses penyeimbangan data menggunakan teknik

SMOTETomek. *SMOTETomek* adalah teknik *resampling* yang bertujuan untuk mengatasi data yang mengandung *noise* dan masalah ketidakseimbangan kelas. Gambar berikut ini merupakan contohnya:

Gambar 6. Confusion Matrix Tanpa SMOTETomek



Sumber: Diolah oleh peneliti (2025)

Berdasarkan hasil *confusion matrix*, jumlah prediksi untuk status “Belum Tahu” sangat rendah (2 data) karena distribusi data yang tidak seimbang, dengan proporsi kasus “Belum Tahu” jauh lebih kecil dibandingkan “Bukan ODHIV” (125) “ODHIV” (305).

2. Pemodelan KNN

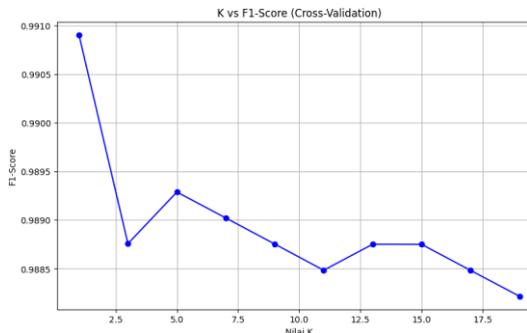
Algoritma *K-Nearest Neighbor (KNN)* adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut (Baharuddin et al., 2019). Algoritma *KNN* termasuk metode yang menggunakan algoritma *supervised learning* bertujuan untuk mengidentifikasi pola-pola baru dalam data dengan cara mengaitkan pola yang telah ditemukan sebelumnya dengan data baru.

Proses pemodelan adalah proses pembentukan model klasifikasi dilakukan dengan menerapkan algoritma atau metode yang diinginkan. Pada penelitian ini, algoritma yang akan diimplementasikan adalah *KNN*. Optimasi parameter nilai K



(tetangga terdekat) terhadap nilai *f1-score* menggunakan teknik *Cross-Validation* pada algoritma *KNN* ditunjukkan sebagai gambar berikut:

Gambar 7. Visualisasi *Cross-Validation*



Sumber: Diolah oleh peneliti (2025)

a. Optimasi Parameter

Parameter yang dioptimasi dalam model ini adalah nilai K pada algoritma *KNN*, dimana setelah dilakukan pengujian dengan *range* nilai K dari 2.5 sampai 20, ditemukan bahwa nilai K optimal adalah 2.5.

b. *Cross-Validation*

Evaluasi performa model menggunakan *f1-score*, dengan hasil terbaik mencapai 0.9908 saat nilai K=2.5, sementara performa terendah terjadi saat K=20 dengan *f1-score* 0.9882.

c. Model *Training & Testing*

Hasil evaluasi model pada *training*, dengan nilai *precision*, *recall*, dan *f1-score* mencapai 1.00 (atau 100%) untuk semua kelas (0,1, dan 2) menghasilkan performa yang sempurna.

Sementara itu, evaluasi pada model *testing* menunjukkan kinerja yang bervariasi untuk setiap kelas. Kelas 0 (Belum Tahu) memiliki *precision* 0.40 dan *recall* 1.00, dengan *f1-score* 0.57, meskipun hanya terdapat 2 sampel. Kelas 1 (Bukan ODHIV) menunjukkan performa yang sangat baik dengan *precision* 0.96, *recall* 0.99, dan *f1-score* 0.98, dengan jumlah sampel 126. Sementara itu, kelas 2 (ODHIV) menunjukkan performa terbaik dengan

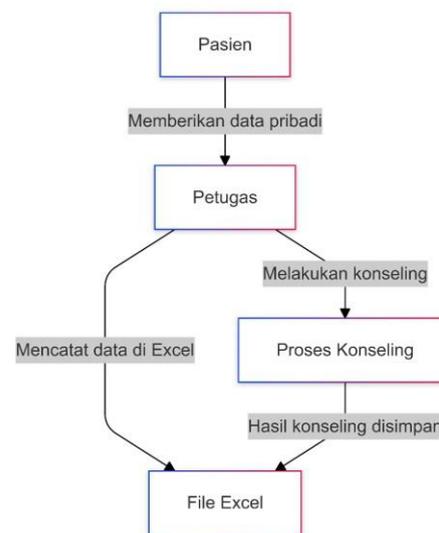
precision 1.00, *recall* 0.97, dan *f1-score* 0.99, dengan 313 sampel. Secara keseluruhan, model *KNN* menunjukkan hasil dengan unggul pada kelas dengan jumlah percobaan yang lebih banyak.

3. Implementasi Sistem

a. Sistem Yang Berjalan

Sistem yang digunakan saat ini masih menggunakan input data secara manual melalui file *excel* antara petugas dan pasien dalam kegiatan konseling terkait penyakit ODHIV. *Flowmap* input data secara manual sebagai gambar berikut:

Gambar 8. Sistem Yang Berjalan



Sumber: Diolah oleh peneliti (2025)

Pasien memberikan data pribadi ke petugas setelah itu melakukan konseling, apakah pasien tersebut mempunyai riwayat ODHIV atau tidak, dari proses ini selesai hasil konseling data disimpan di *Microsoft Excel*.

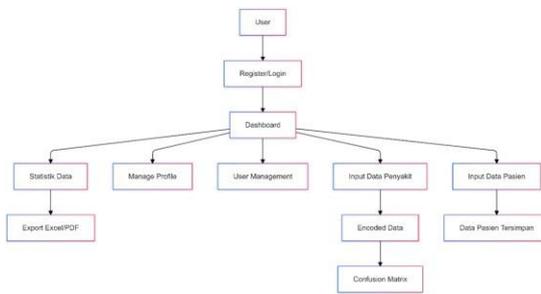
b. Sistem Yang Diusulkan

Program yang akan dikembangkan adalah Sistem Klasifikasi *KNN*, yaitu sebuah aplikasi yang dirancang untuk melakukan pengujian model klasifikasi menggunakan algoritma *KNN*. Selain berfungsi sebagai sistem pengujian klasifikasi *KNN*, aplikasi ini juga diharapkan dapat mendukung dan mempermudah proses konseling secara lebih efektif. Sistem yang diusulkan ditunjukkan



sebagai gambar berikut:

Gambar 9. Sistem Yang Diusulkan



Sumber: Diolah oleh peneliti (2025)

Langkah-langkah sistem yang diusulkan dapat di deskripsikan pada tabel berikut ini:

Tabel 3. Deskripsi Fitur Sistem Aplikasi

Fitur	Deskripsi
Register	User dapat mendaftar akun sebelum ke halaman login.
Login	User dapat mengakses aplikasi berdasarkan akun yang sudah dibuat.
Forgot Password	User dapat mengubah password akun.
Dashboard	Pengguna masuk ke halaman utama setelah login. Menampilkan statistik yang dapat diekspor ke Excel atau PDF.
Manage Profile	Mengelola Profil Pengguna.
User Management	Mengelola pengguna lain dalam sistem.
Data Penyakit	Input form data penyakit yang akan diubah menjadi Encoded Data dan menampilkan Confusion Matrix.
Data Encoded	Pengujian sistem klasifikasi KNN.
Data Pasien	Mengelola data pasien dalam sistem.

Sumber: Diolah oleh peneliti (2025)

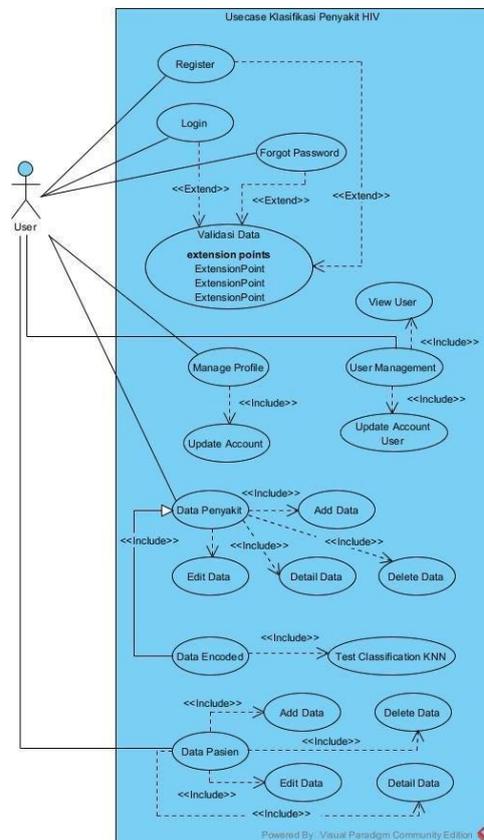
c. Perancangan Sistem

Perancangan sistem ini adalah tahapan perancangan sistem yang akan dibentuk yang dapat berupa penggambaran proses-proses suatu elemen-elemen dari suatu komponen, proses perancangan ini merupakan suatu tahapan awal dari perancangan web persediaan (Setiyanto Rudi et al., 2019). Sistem pengujian klasifikasi KNN dengan menggunakan UML (Unified Modelling Language) diagram dan untuk pembuatan perangkat lunaknya dibuat dengan menggunakan sistem aplikasi database MySQL dan aplikasi Visual Paradigm 17.0 dalam pembuatan UML. Perancangan aplikasi ini menggunakan 4 diagram, diantaranya:

a) Usecase Diagram

Usecase Diagram digunakan oleh penulis menjelaskan proses-proses yang berlangsung pada sebuah sistem.

Gambar 10. Usecase Diagram



Sumber: Diolah oleh peneliti (2025)



Berikut adalah deskripsi pendefinisian *usecase* pada sistem klasifikasi *KNN*:

Tabel 4. Identifikasi *Usecase*

Aktor	Usecase	Deskripsi
User	Register	Proses yang dilakukan oleh petugas/user yang belum terdaftar di dalam aplikasi.
User	Login	Proses yang dilakukan oleh petugas/user yang terdaftar di sistem sebelum masuk ke sistem.
User	Forgot Password	Proses yang dilakukan petugas/user apabila lupa akun sebelum masuk ke dalam sistem.
User	Dashboard	Proses yang dilakukan petugas/user yang dapat melihat data statistik dan dapat <i>export</i> data.
User	Manage Profile	Proses yang dilakukan petugas/user yang dapat mengubah akun secara langsung tanpa melihat user lain yang akses sistem ini.
User	User Management	Proses yang dilakukan petugas/user untuk dapat melihat user lain yang akses dan dapat mengubah data.
User	Data Penyakit	Proses yang dilakukan petugas/user untuk dapat memasukkan data penyakit berdasarkan form inputan tertera pada aplikasi.
User	Data Encoded	Proses yang dilakukan petugas/user untuk melakukan testing

klasifikasi *KNN* dan dapat melihat hasil *confusion matrix* berupa tabel dibawah tabel data.

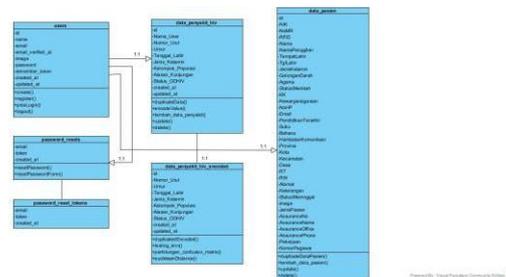
User	Data Pasien	Proses yang dilakukan petugas/user untuk dapat memasukkan data pasien berdasarkan form input tertera pada aplikasi.
------	-------------	---

Sumber: Diolah oleh peneliti (2025)

b) *Class Diagram*

Berikut merupakan gambar *class diagram* dari sistem yang akan dibangun:

Gambar 11. *Class Diagram*



Sumber: Diolah oleh peneliti (2025)

Keterangan kelas dalam *class diagram*:

Tabel 5. Keterangan *Class Diagram*

Nama Kelas	Keterangan
<i>Users</i>	Merupakan kelas yang mengelola data pengguna dalam mengakses aplikasi.
<i>Passwords_Resets</i>	Merupakan kelas untuk melakukan perubahan <i>password</i> pada halaman awal sebuah aplikasi.
<i>Password_Reset_Tokens</i>	Merupakan kelas untuk menerima token setelah melakukan perubahan <i>password</i> .



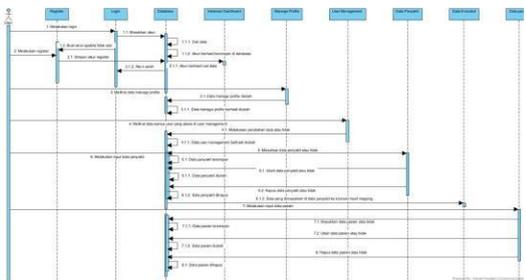
Data_Penyakit_Hiv	Merupakan kelas data yang digunakan untuk proses pengelolaan data penyakit HIV.
Data_Penyakit_Encoded	Merupakan kelas data untuk menampung kelas data yang terkonversi dari tabel database data penyakit hiv.
Data_Pasien	Merupakan kelas data yang digunakan untuk proses pengelolaan data pasien.

Sumber: Diolah oleh peneliti (2025)

c) *Sequence Diagram*

Berikut adalah diagram aktivitas dari sistem informasi klasifikasi penyakit HIV:

Gambar 12. Sequence Diagram



Sumber: Diolah oleh peneliti (2025)

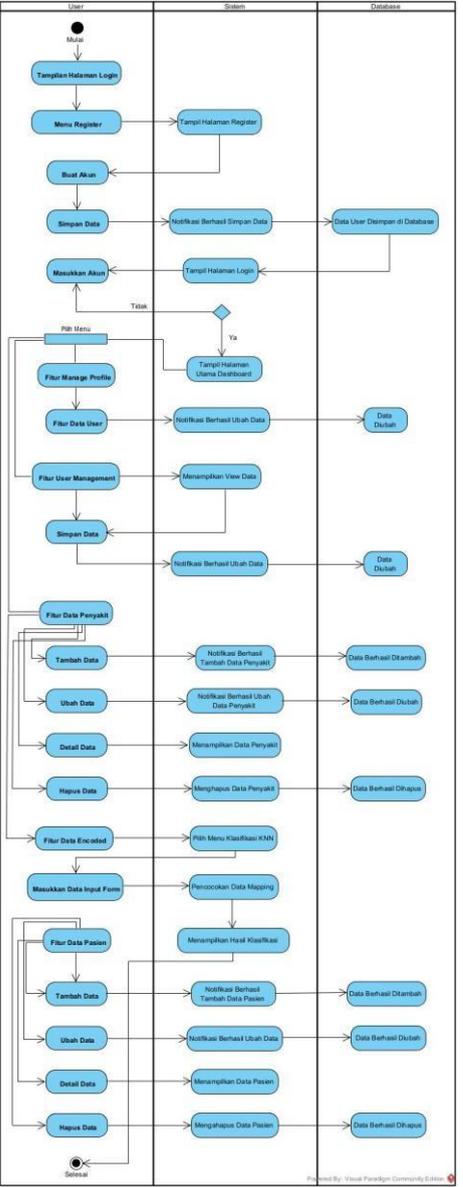
Sequence diagram menggambarkan alur interaksi sistem informasi klasifikasi penyakit HIV yang dimulai dari proses autentikasi (*login/register*), dilanjutkan dengan akses ke fitur utama seperti *manage profile*, *user management*, serta pengelolaan data penyakit dan pasien. Sistem memungkinkan operasi *create, read, update, delete (CRUD)* untuk data penyakit dan pasien, serta fitur konversi hasil *mapping* yang terintegrasi dengan database untuk penyimpanan dan pengambilan data.

d) *Activity Diagram*

Activity diagram menggambarkan untuk

interaksi tertentu antara pengguna dan sistem. Berikut adalah diagram aktivitas dari Aplikasi Klasifikasi Penyakit HIV:

Gambar 13. Activity Diagram



Sumber: Diolah oleh peneliti (2025)

Activity diagram menunjukkan alur sistem informasi klasifikasi penyakit HIV yang mencakup proses autentikasi, manajemen data (*user, penyakit, pasien*), serta fitur klasifikasi menggunakan metode *KNN* dengan percobaan data *mapping*, setiap aktivitas terintegrasi dengan sistem database.



Hasil dan Pembahasan

Pada penjelasan ini akan diuraikan *output* klasifikasi penyakit *HIV* menggunakan algoritma *KNN* berdasarkan dataset yang telah dikumpulkan.

A. Implikasi Ketidakseimbangan Data

Pada tahap sebelumnya, telah dibahas bahwa distribusi status *ODHIV* menunjukkan ketidakseimbangan data, terutama pada status 0 (Belum Tahu). Ketidakseimbangan ini terjadi karena jumlah data asli untuk status 0 hanya terdiri dari dua data, sehingga analisis menjadi kurang representatif untuk kategori tersebut. Berikut distribusi data yang telah diseimbangkan menggunakan teknik *SMOTETomek*, bertujuan untuk mengatasi ketidakseimbangan kelas dalam dataset penyakit *HIV*:

Tabel 6. Distribusi Data Setelah *SMOTETomek*

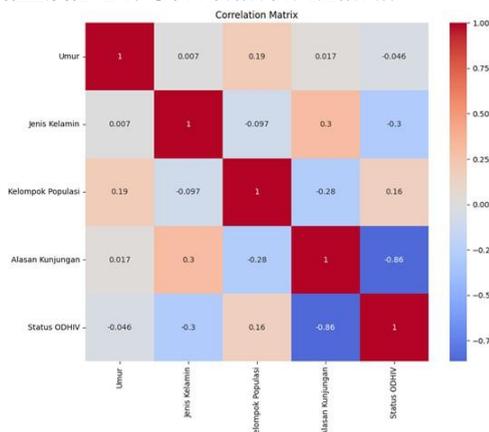
Status ODHIV	Jumlah
1	1249
2	1246
0	1246

Sumber: Diolah oleh peneliti (2025)

B. Korelasi Antar Fitur

Analisis ini membahas korelasi antar lima fitur yaitu: umur, jenis kelamin, kelompok populasi, alasan kunjungan, dan status *ODHIV*. Berikut visualisasi hubungan antar variabel sebagai gambar berikut ini:

Gambar 14. *Correlation Matrix*



Sumber: Diolah oleh peneliti (2025)

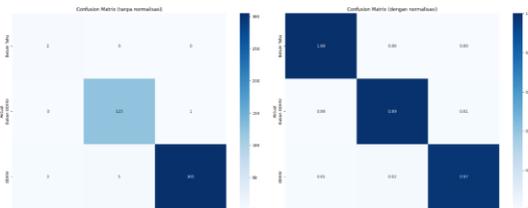
Visualisasi *correlation matrix* menunjukkan hubungan keterkaitan antar variabel dalam dataset. Temuan utamanya adalah muncul ikatan yang sangat kuat dan terbalik antara Status *ODHIV* dengan Alasan Kunjungan (nilai -0.86), yang berarti ketika salah satu variabel meningkat, variabel lainnya cenderung menurun.

Jenis kelamin memiliki hubungan moderat dengan status *ODHIV* (-0.3) dan Alasan kunjungan (0.3). Sementara faktor Umur hampir tidak memiliki pengaruh terhadap variabel lainnya, terlihat dari nilai korelasinya yang sangat kecil (mendekati 0). Dengan kata lain, Status *ODHIV* dan Alasan Kunjungan adalah dua variabel yang paling berkaitan erat dalam dataset ini, sedangkan Umur hampir tidak memiliki pengaruh terhadap variabel lainnya.

C. Hasil Evaluasi Model Menggunakan *Confusion Matrix*

Hasil evaluasi model setelah penyeimbangan data ditunjukkan melalui *confusion matrix* berikut, yang memvisualisasikan akurasi dan kesalahan prediksi:

Gambar 15. Hasil Evaluasi Model *Confusion Matrix*



Sumber: Diolah oleh peneliti (2025)

Model klasifikasi ini menunjukkan performa yang sangat memuaskan dalam mengklasifikasikan status *ODHIV*. Secara keseluruhan, model mencapai akurasi yang tinggi dengan 97-100% untuk setiap kategori, dimana klasifikasi “Belum Tahu” mencapai akurasi sempurna 100% “Bukan *ODHIV*” mencapai 99% dan “*ODHIV*” mencapai 97%. Tingkat kesalahan klasifikasi sangat minimal, dengan kesalahan tertinggi hanya 2% pada



kasus *ODHIV* yang salah diklasifikasikan sebagai bukan *ODHIV*.

D. Hasil Evaluasi Performa Model KNN

Menurut hasil pemeriksaan data yang telah dilakukan, penelitian ini telah melalui beberapa tahapan, yaitu pembersihan data, pemfilteran data, dan penerapan teknik keseimbangan data menggunakan *SMOTETomek*. Pada tahap ini, penelitian akan memaparkan hasil performa model *K-Nearest Neighbor (KNN)* yang telah diimplementasikan, seperti ditunjukkan pada tabel berikut:

Tabel 7. Hasil Evaluasi Model K-Nearest Neighbor (KNN)

Metode Evaluasi	Nilai (%)
<i>Accuracy</i>	97.96
<i>Macro Average Precision</i>	78.61
<i>Macro Average Recall</i>	98.88
<i>Macro Average F1-Score</i>	84.45

Sumber: Diolah oleh peneliti (2025)

Model klasifikasi ini menunjukkan hasil dengan rincian sebagai berikut:

1. Tingkat ketepatan model secara keseluruhan mencapai 97.96% yang berarti dari 2.205 data yang diuji, model mampu memberikan jawaban yang benar untuk 2.203 data. Hal ini menunjukkan bahwa model berhasil melakukan prediksi status *ODHIV*.
2. Tingkat ketepatan dalam mengenali setiap kelompok status mencapai 78.61%. Angka ini menunjukkan bahwa ada beberapa kesalahan disaat melakukan prediksi dimana model terkadang memberikan status yang tidak sesuai dengan kondisi sebenarnya.
3. Kemampuan model dalam menemukan kasus yang sebenarnya mencapai 98.88%. Ini berarti model sangat baik

dalam mengenali status *ODHIV* seseorang sesuai dengan kondisi yang sebenarnya.

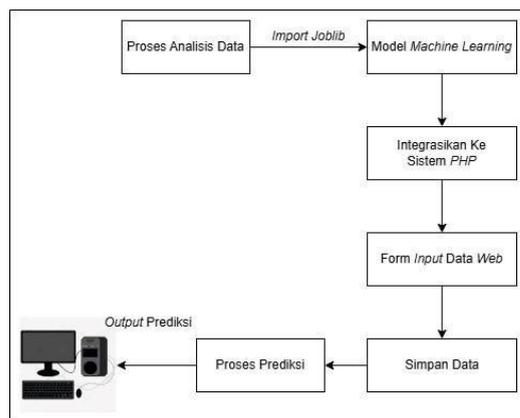
4. Nilai keseimbangan antara ketepatan prediksi dan kemampuan pengenalan mencapai 84.45%. Angka ini menunjukkan bahwa model memiliki kinerja yang seimbang dalam melakukan tugasnya.

Kesimpulannya, model ini sangat efektif dalam mengklasifikasikan status *ODHIV* dengan tingkat prediksi yang sangat akurat.

E. Rancangan Sistem

Sistem ini dirancang untuk menggambarkan proses implementasi model *machine learning* ke dalam sistem web yang menggunakan bahasa pemrograman *PHP*. Model *machine learning* dibuat dan diuji menggunakan *python*, kemudian model tersebut diekspor dan diintegrasikan dengan sistem berbasis *PHP*. Proses ini memungkinkan pengolahan data dan prediksi yang dihasilkan oleh model dan diakses langsung melalui antarmuka *web*.

Gambar 16. Rancangan Sistem



Sumber: Diolah oleh peneliti (2025)

Dari proses analisis data, dilakukan *import jolib* untuk menyimpan model *machine learning*, *jolib* digunakan untuk menyimpan (*serialize*) dan memuat (*deserialize*) model secara efisien, terutama untuk data besar seperti *array NumPy*. Model yang disimpan

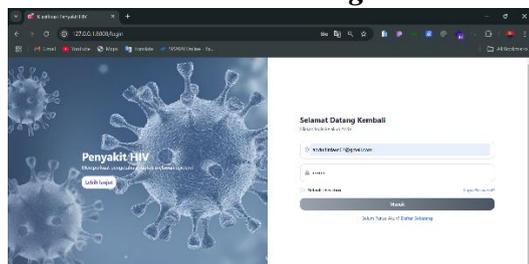


kemudian diintegrasikan dengan *PHP* untuk membuat sistem *web*. Sistem ini memiliki menu *input form*, dimana pengguna mengisi data dan mengirimkannya untuk prediksi. Hasil prediksi ditampilkan pada label utama Status *ODHIV* (0: Belum Tahu, 1: Bukan *ODHIV*, 2: *ODHIV*).

F. Hasil Tampilan Website Testing Klasifikasi KNN

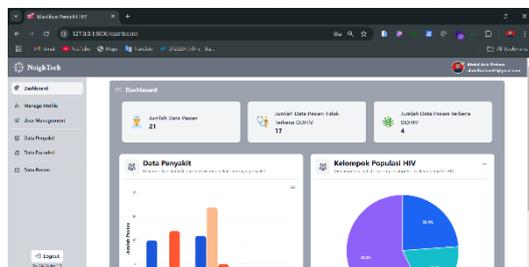
Hasil tampilan website dapat dilihat pada gambar berikut:

Gambar 17. Halaman Login



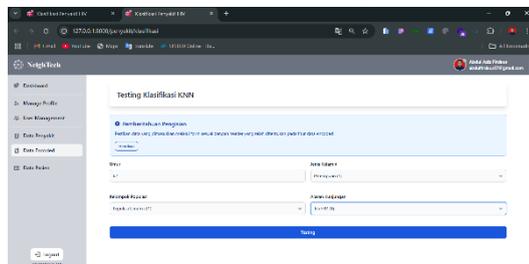
Sumber: Diolah oleh peneliti (2025)

Gambar 18. Halaman Dashboard



Sumber: Diolah oleh peneliti (2025)

Gambar 19. Halaman Klasifikasi KNN



Sumber: Diolah oleh peneliti (2025)

Kesimpulan

Penelitian ini menunjukkan bahwa model *K-Nearest Neighbors* (*KNN*) terbukti efektif dalam mengelompokkan data pasien *HIV/AIDS*, dengan performa yang signifikan

dalam menganalisis berbagai label berdasarkan variabel kunci seperti umur, jenis kelamin, kelompok populasi, dan alasan kunjungan. Hasil pengukuran kinerja model menunjukkan performa yang sangat baik dengan nilai *accuracy* mencapai 97.96%, *precision* 78.61%, *recall* 98.88%, dan *f1-score* 84.45%. Sebagai implementasi dari penelitian ini, telah dikembangkan sebuah *framework website* untuk pengujian klasifikasi *KNN* yang dapat diakses dan digunakan, dilengkapi dengan berbagai fitur didalamnya.

Daftar Pustaka

- Agung, A., Daniswara, A., Kadek, I., & Nuryana, D. (2023). Data Preprocessing Pola Pada Penilaian Mahasiswa Program Profesi Guru. *Journal of Informatics and Computer Science*, 05.
- Aprianingsih, Y., Sianturi, S. R., Tinggi, S., Kesehatan, I., & Carolus, S. (2021). HUBUNGAN KARAKTERISTIK INDIVIDU DENGAN TINGKAT PENGETAHUAN MASYARAKAT TENTANG PENYAKIT HIV/AIDS DI BEKASI.
- Baharuddin, M. M., Azis, H., & Hasanuddin, T. (2019). Analisis Performa Metode K-Nearest Neighbor Untuk Identifikasi Jenis Kaca. *ILKOM Jurnal Ilmiah*, 11(3), 269–274. <https://doi.org/10.33096/ilkom.v11i3.489.269-274>
- Cahyanti, D., Rahmayani, A., & Ainy Husniar, S. (2020). Indonesian Journal of Data and Science Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara. 1(2), 39–43.
- Gusrianti, G., Ulva, F., & Azkha, N. (2023). Efektivitas Penyuluhan Kesehatan Dalam Peningkatan Pengetahuan Remaja Tentang HIV/AIDS. *JIK JURNAL ILMU KESEHATAN*, 7(2), 388. <https://doi.org/10.33757/jik.v7i2.959>
- Ismayati, N., Rifai, A., & Rahayu, T. (2023). Media Informasi Kesehatan untuk Pencegahan HIV/AIDS Yang Disukai



- Generasi Z: Upaya Penurunan Kasus HIV/AIDS di Kalangan Remaja di Indonesia. *Tibandaru: Jurnal Ilmu Perpustakaan Dan Informasi*, 7(1), 54. <https://doi.org/10.30742/tb.v7i1.2824>
- Juhaefah, A., Paramita, S., Kosala, K., & Gunawan, C. A. (2020). GAMBARAN KARAKTERISTIK PASIEN HIV/AIDS YANG MENDAPAT ANTIRETROVIRAL THERAPY (ART). In *Jurnal Medika Karya Ilmiah Kesehatan* (Vol. 5, Issue 1). Online.
- Latifah, R., Setia Wulandari, E., & Priadhana Edi Kreshna, dan. (2019). *Model Decision Tree untuk Prediksi Jadwal Kerja menggunakan Scikit-Learn* (Vol. 16).
- Maskuri, M. N., Sukerti, K., & Herdian Bhakti, R. M. (2022). Penerapan Algoritma K-Nearest Neighbor (KNN) untuk Memprediksi Penyakit Stroke Stroke Disease Predict Using KNN Algorithm. *Jurnal Ilmiah Intech: Information Technology Journal of UMUS*, 4(1).
- Pangaribuan, S. M., Maulidanti, N. N., Siringoringo, L., Rs, A., Cikini, P., & Indonesia, J. (n.d.). Pengetahuan Remaja Tentang Hiv/Aids Di Kelurahan Menteng Jakarta Pusat. *JAKHKJ*, 7(2), 2021.
- S. Rohmah, R. L. Y. I. H. (2024). Hubungan Pengetahuan Siswa Tentang Hiv/AIDS Dengan Sikap Pencegahan Hiv/AIDS Melalui Penerapan Budaya Kagaluan Di Smkn 1 Ciamis. *JKDB: Jurnal Konservasi dan Budaya*, 103–116.
- Setiyanto Rudi, Nurmaesah Nunung, & Rahayu Astuti Sri Nyai. (2019). Jurnal Perancangan Sistem Definisi. *JURNAL SISFOTEK GLOBAL*, 9, 139–139.
- Siburian, N., Cholissodin, I., & Adikara, P. P. (2020). *Penerapan Metode Fuzzy K-Nearest Neighbor Pada Klasifikasi Penyakit Menular Seksual Pria* (Vol. 4, Issue 11). <http://j-ptiik.ub.ac.id>
- Zahra, D., Nurdin, A., Fitria, U., Dinen, K. A., & Kurnia, R. (2021). Pemanfaatan Teknologi Dalam Bidang Kesehatan Masyarakat. In *Public health Journal*. <https://teewanjournal.com/index.php/phj/index>
- Zulaikhah, S. H., Aziz, A., & Harianto, W. (2022). Optimasi Algoritma K-Nearest Neighbor (Knn) Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver. In *Jurnal Mahasiswa Teknik Informatika* (Vol. 6, Issue 2). <https://archive.ics.uci.edu/ml/index.php>

