

BAB III METODE PENELITIAN

3.1. Objek Penelitian

Objek dalam penelitian ini adalah data terkait penyakit HIV di wilayah Jawa Barat, yang diperoleh melalui penyebaran kuesioner kepada 101 responden. Dataset tersebut memuat informasi mengenai umur, jenis kelamin, serta jawaban dari 16 pertanyaan yang berkaitan dengan gejala-gejala HIV. Penelitian ini difokuskan pada pengembangan model identifikasi HIV dengan menggunakan algoritma *Support Vector Machine* (SVM) berbasis kernel RBF dan *Naïve Bayes*.

3.2. Waktu Penelitian

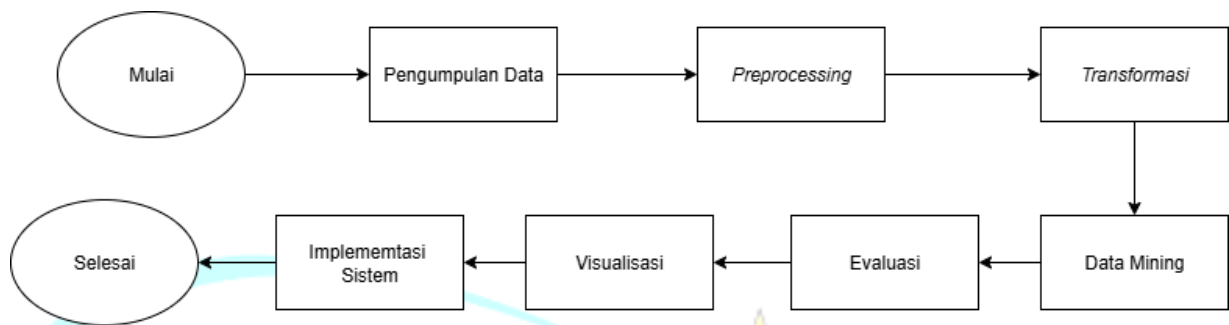
Penelitian ini dirancang dalam kurun waktu 4 bulan, dimulai dari tahap pengumpulan data hingga Hasil evaluasi dengan data kuesioner yang disebarakan.

Tabel 3.1 Waktu Penelitian

No	Tahap Penelitian	November				Desember				Januari				Februari			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Pengumpulan Data	█															
2	Preprocessing					█											
3	Transformasi					█											
4	Data Mining					█											
5	Hasil Evaluasi									█							

3.3. Prosedur Penelitian

Prosedur penelitian ini dirancang untuk memberikan langkah-langkah sistematis dalam mengumpulkan dan menganalisis data, guna mencapai tujuan penelitian yang telah ditetapkan. Berikut gambar flowchart untuk prosedur penelitian.



Gambar 3.1. Prosedur penelitian

Penjelasan Alur:

3.3.1 Pengumpulan Dataset

Dataset yang digunakan dalam penelitian ini merupakan data kuesioner yang berisi 16 pertanyaan mendetail mengenai gejala-gejala yang biasa dialami oleh penderita penyakit HIV. Selain itu, dataset ini juga mencakup informasi demografis responden, seperti jenis kelamin dan rentang umur, yang menjadi variabel penting dalam analisis. Data yang bersifat kategorikal dengan jawaban "Ya" atau "Tidak" ini memberikan dasar yang kaya untuk pengembangan model klasifikasi, di mana algoritma SVM dan *Naïve Bayes* diaplikasikan untuk mengidentifikasi pola dan korelasi antara gejala dengan status HIV.

Kuesioner ini mengumpulkan data tentang gejala dan informasi demografis responden, termasuk jenis kelamin dan rentang usia. Informasi ini penting untuk memahami kerentanan dan gejala HIV. Dataset mencakup kondisi klinis dan latar belakang sosial untuk analisis lebih lanjut. Data bersifat kategorikal dengan jawaban "Ya" atau "Tidak", yang memudahkan pengolahan. Penelitian menggunakan algoritma seperti SVM dan *Naïve Bayes* untuk membangun model prediksi yang menghubungkan gejala dengan status infeksi HIV. Pengumpulan data dilakukan dengan etika, menjaga kerahasiaan, dan memperoleh persetujuan sukarela. Data disimpan aman untuk melindungi informasi pribadi, dengan harapan menghasilkan model akurat untuk deteksi dini HIV secara digital.

	B	C	D	E	F	G	H	I	J
1	Jenis Kelamin	Umur	Apakah Anda m	Apakah Anda m	Apakah Anda m	Apakah Anda m	Apakah Anda m	Apakah Anda m	Apakah Anda m
2	Laki-Laki	20-29 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
3	Perempuan	30-39 Tahun	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
4	Laki-Laki	30-39 Tahun	Ya	Tidak	Ya	Tidak	Ya	Tidak	Ya
5	Laki-Laki	20-29 Tahun	Tidak	Tidak	Tidak	Tidak	Ya	Ya	Tidak
6	Laki-Laki	20-29 Tahun	Tidak	Tidak	Tidak	Ya	Ya	Tidak	Tidak
7	Laki-Laki	20-29 Tahun	Tidak	Tidak	Ya	Tidak	Ya	Ya	Tidak
8	Laki-Laki	20-29 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
9	Perempuan	30-39 Tahun	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
10	Laki-Laki	30-39 Tahun	Ya	Tidak	Tidak	Tidak	Ya	Ya	Ya
11	Laki-Laki	30-39 Tahun	Tidak	Ya	Ya	Ya	Ya	Ya	Ya
12	Perempuan	20-29 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
13	Perempuan	20-29 Tahun	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
14	Laki-Laki	20-29 Tahun	Ya	Ya	Tidak	Tidak	Ya	Tidak	Tidak
15	Perempuan	30-39 Tahun	Ya	Ya	Tidak	Tidak	Ya	Tidak	Ya
16	Laki-Laki	40-49 Tahun	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak
17	Laki-Laki	20-29 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
18	Perempuan	30-39 Tahun	Ya	Ya	Ya	Tidak	Ya	Ya	Tidak
19	Laki-Laki	30-39 Tahun	Ya	Ya	Ya	Tidak	Ya	Ya	Ya
20	Perempuan	30-39 Tahun	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Ya
21	Perempuan	20-29 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
22	Perempuan	30-39 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
23	Perempuan	30-39 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
24	Perempuan	20-29 Tahun	Ya	Ya	Ya	Ya	Ya	Ya	Ya
25	Perempuan	30-39 Tahun	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak	Tidak

Gambar 3.2 Dataset Penyakit HIV

3.3.2 Preprocessing dataset

Tahapan ini bertujuan untuk memastikan data yang digunakan bersih dan siap untuk diolah. Beberapa Langkah yang dilakukan:

- 1) Pemeriksaan *Missing Values*: tidak ditemukan data yang hilang (semua bernilai lengkap), sehingga tidak diperlukan proses penghapusan.
- 2) Pengecekan *Outlier*: tidak ditemukan data ekstrem dalam dataset, sehingga tidak diperlukan penanganan *outlier*.

3.3.3 Transformasi

Pada tahap ini, data diubah menjadi format atau struktur yang lebih sesuai untuk analisis dalam identifikasi penyakit HIV. Contoh transformasi yang dilakukan:

- 1) Mengubah data kategorikal menjadi data *numerik* (*encoding*), seperti mengonversi jawaban "Ya/Tidak" pada riwayat penggunaan jarum suntik, faktor risiko, atau gejala menjadi nilai 0 dan 1.
- 2) Membuat fitur baru berdasarkan data yang ada, misalnya menghitung tingkat perubahan jumlah CD4 per bulan atau tingkat pertumbuhan kasus HIV tahunan untuk melihat pola perkembangan penyakit.
- 3) Membagi data menjadi data *training* dan data *testing* dengan proporsi tertentu (misalnya 80% *training* dan 20% *testing*) untuk memastikan model dapat mempelajari pola dari data dan diuji pada data baru.

3.3.4 Data Mining

Pada tahap ini, dilakukan penerapan algoritma SVM dan *Naïve Bayes* untuk identifikasi penyakit HIV.

- 1) *Pseudocode Support Machine Vector* (SVM) kernel RBF

Tabel 3.2 *Pseudocode* identifikasi dengan SVM

Algoritma: Identifikasi menggunakan SVM kernel RBF

Input: Dataset *training* $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$

x_i = fitur input

y_i = label kelas (+1 atau -1)

Output: Model SVM dengan parameter α , b

1. Pilih fungsi kernel RBF: $K(x_i, x) = \exp(-\gamma \|x_i - x\|^2)$
 2. Tentukan parameter C (regularisasi) dan γ (parameter kernel)
 3. Formulasi masalah optimasi:
-

Algoritma: Identifikasi menggunakan SVM kernel RBF

Maksimumkan margin dengan meminimumkan:

$$\min \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i$$

dengan constraints $\mathbf{y}_i (\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i) + \mathbf{b}) \geq 1 - \xi_i \quad i = 1, \dots, n,$

$\xi_i \geq 0 \quad i = 1, \dots, n,$

4. Gunakan algoritma *Quadratic Programming* (QP) untuk menemukan α (*lagrange multiplier*) dan b (bias)

5. Setelah selesai, simpan model α , b , dan *support vectors*

6. Untuk prediksi data baru \mathbf{x} :

$$\text{Hitung } f(\mathbf{x}) = \sum_{i=1}^n \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b$$

Jika $f(\mathbf{x}) \geq 0$, prediksi kelas +1

Jika $f(\mathbf{x}) < 0$, prediksi kelas -1

2) *Pseudocode Naive Bayes*

Tabel 3.3 *Pseudocode* identifikasi dengan *Naive Bayes*

Algoritma: Identifikasi menggunakan *Naive Bayes*

Input: Dataset *training* $\{(\mathbf{x}_1, \mathbf{y}_1), (\mathbf{x}_2, \mathbf{y}_2), \dots, (\mathbf{x}_n, \mathbf{y}_n)\}$

x_i = fitur input (x_1, x_2, \dots, x_m)

y_i = label kelas (kelas dari data)

Output: Model *probabilistik Naive Bayes*

Algoritma: Identifikasi menggunakan *Naive Bayes*

1. Hitung prior probability untuk setiap kelas c :

$$P_{(c)} = \frac{N_c}{N}$$

N_c = Jumlah data dengan kelas c

N = Total jumlah data

2. Untuk setiap fitur x_i dan kelas c , hitung *likelihood*:

Algoritma: Identifikasi menggunakan *Naïve Bayes*

$$P(x_i = v|c) = \frac{\text{jumlah data dengan } x_i = v \text{ dan kelas } c}{N_c}$$

(untuk fitur kontinu, gunakan distribusi probabilitas seperti *Gaussian*)

$$P(x_i|c) = \frac{1}{\sqrt{2\pi\sigma_{ic}^2}} \exp\left(-\frac{(x_i - \mu_{ic})^2}{2\pi\sigma_{ic}^2}\right)$$

μ_{ic} = rata – rata fitur x_i untuk kelas c

$\pi\sigma_{ic}^2$ = varians fitur x_i untuk kelas c

3. Model probabilitas kini menjadi:

$$P(c | \mathbf{x}) \propto P(c) \prod_{i=1}^m P(x_i|c)$$

4. Untuk klasifikasi data baru $\mathbf{x} = (x_1, x_2, \dots, x_m)$:

Hitung skor untuk tiap kelas c :

$$\text{score}(c) = P(c) \times \prod_{i=1}^m P(x_i|c)$$

5. Pilih kelas dengan skor tertinggi sebagai prediksi kelas *output*

3.3.5 Evaluasi

Model dievaluasi menggunakan *Confusion Matrix*. Matriks yang menggambarkan kinerja model klasifikasi dengan membandingkan prediksi model terhadap nilai aktual. Matriks ini terdiri dari *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN), yang digunakan untuk menghitung metrik evaluasi seperti *akurasi*, *presisi*, *recall*, dan *F1-score*.

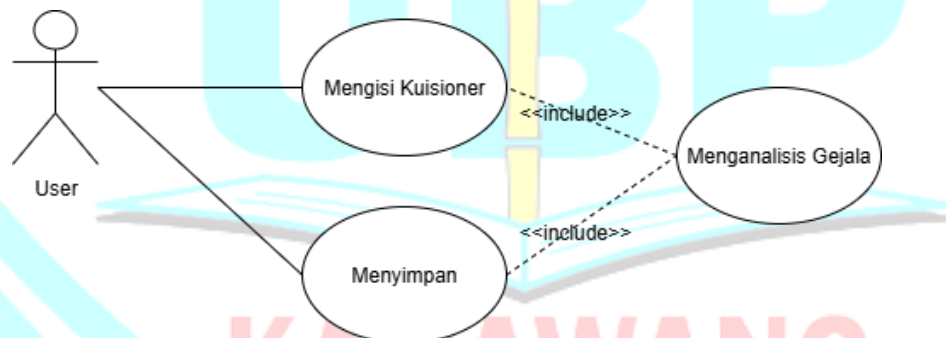
3.3.6 Visualisasi

Hasil evaluasi dan prediksi model divisualisasikan untuk memudahkan interpretasi, antara lain:

- 1) Grafik Distribusi Kelas (*Bar Chart* atau *Pie Chart*): Visualisasi untuk memperlihatkan proporsi distribusi data kelas target (misal pasien dengan dan tanpa gejala HIV) dalam dataset sebagai gambaran awal distribusi data.
- 2) Grafik Perbandingan Akurasi dan Metode Evaluasi Kinerja: Biasanya terdapat grafik batang (*bar chart*) yang menampilkan perbandingan metrik kinerja seperti *akurasi*, *presisi*, *recall*, dan *error rate* antara algoritma SVM dan *Naive Bayes* yang digunakan dalam penelitian

3.4 Desain Implementasi Sistem

Desain implementasi sistem merupakan tahap penting dalam proses pengembangan perangkat lunak yang bertujuan untuk menggambarkan bagaimana sistem akan bekerja sesuai dengan kebutuhan pengguna. Salah satu metode yang digunakan dalam tahap ini adalah *use case diagram* yang bertujuan untuk menggambarkan interaksi antara aktor dengan sistem berdasarkan fungsi-fungsi utama yang tersedia.



Gambar 3.3 Usecase Diagram Pengguna

Berdasarkan *use case diagram* yang ditampilkan pada Gambar 3.3, terdapat satu aktor utama, yaitu Pengguna, yang berinteraksi langsung dengan tiga *use case* utama: Mengisi Kuesioner, Menganalisis Gejala, dan Menyimpan.

- 1) Mengisi Kuesioner: Aktor melakukan input data berupa jawaban atas pertanyaan yang terdapat dalam kuesioner. Tahap ini menjadi fondasi awal dari proses identifikasi atau analisis yang akan dilakukan sistem.
- 2) Menganalisis Gejala: Setelah kuesioner diisi, sistem akan memproses data tersebut untuk mengidentifikasi gejala atau kondisi yang relevan berdasarkan

logika atau algoritma yang telah ditentukan sebelumnya, seperti logika fuzzy atau machine learning.

- 3) Menyimpan: Hasil dari analisis dan data kuesioner akan disimpan ke dalam basis data untuk kepentingan rekam jejak, pelaporan, atau tindak lanjut analisis di kemudian hari.

Ketiga proses utama ini menunjukkan alur kerja sistem yang sederhana namun terstruktur, yang memudahkan pengguna dalam melakukan deteksi atau asesmen awal terhadap gejala tertentu melalui sistem berbasis komputer. Implementasi desain ini akan diterapkan pada tahap pengembangan sistem untuk memastikan setiap fungsi berjalan sesuai dengan peran dan alur yang telah ditentukan dalam *use case diagram*.

