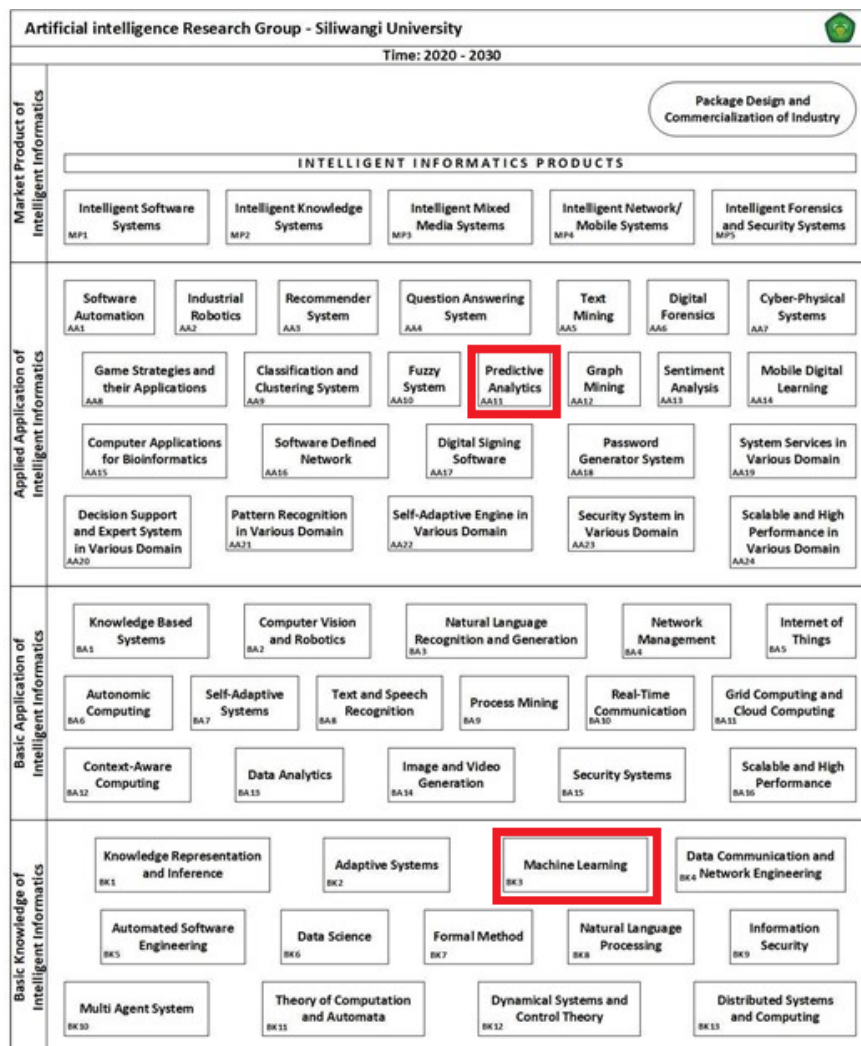


BAB III

METODOLOGI PENELITIAN

3.1 Peta Jalan (*Road Map*) Penelitian

penelitian ini mengikuti roadmap *Artificial Intelligence* Universitas Siliwangi dengan fokus penelitian ini pada pemodelan prediksi kasus MPXV menggunakan ANN dengan membandingkan tiga algoritma optimasi. Roadmap penelitian ini dapat dilihat pada Gambar 3.1



Gambar 3.1 Roadmap Penelitian AIS Siliwangi

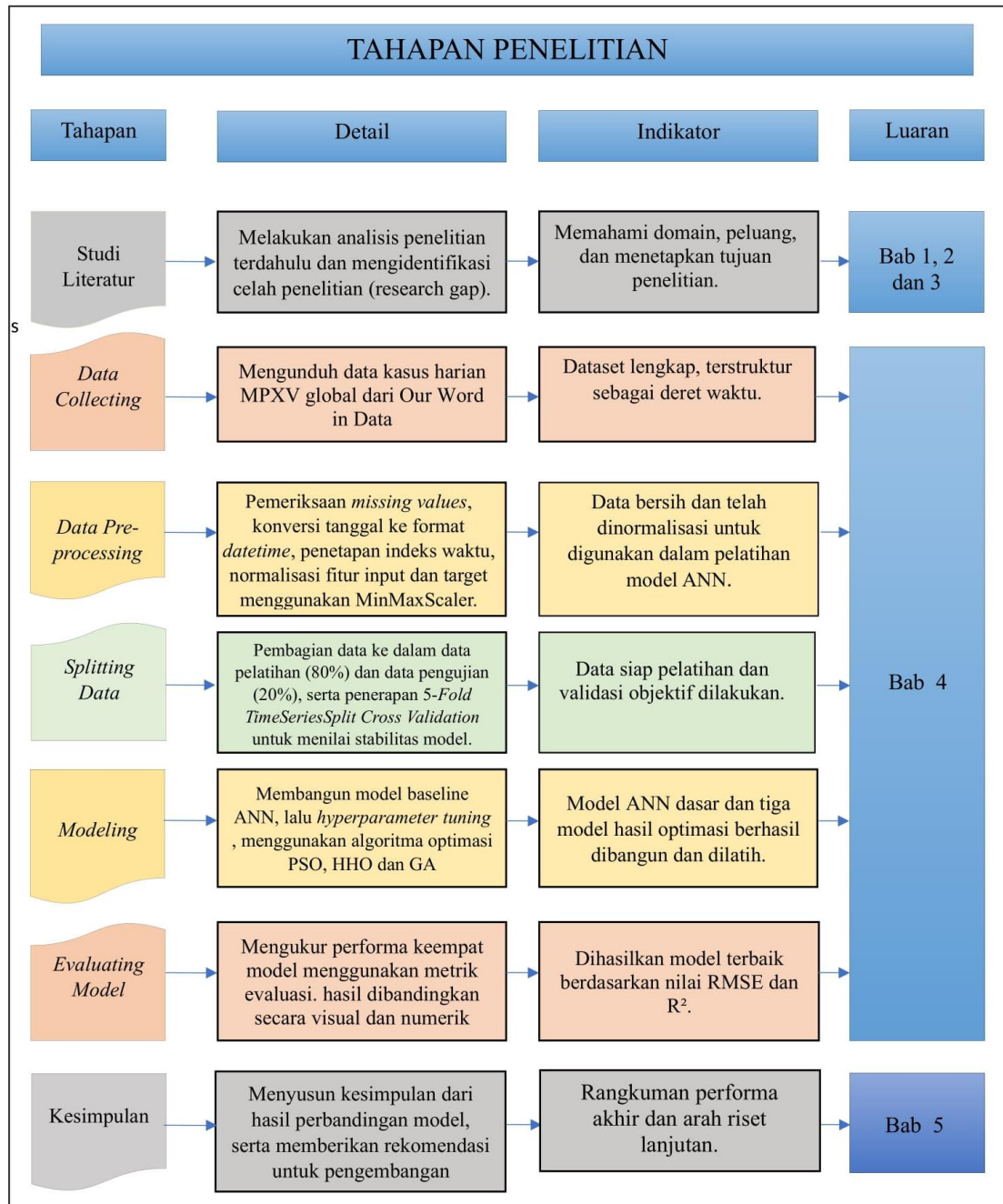
Penelitian ini bertujuan untuk menganalisis dan membandingkan performa masing-masing algoritma dalam mengoptimalkan model ANN untuk meningkatkan akurasi prediksi kasus MPXV. Evaluasi dilakukan dengan menggunakan matrik RMSE dan R-Squared (R^2) untuk membandingkan kinerja model. Hasil penelitian ini diharapkan dapat memberikan wawasan tentang penerapan *machine learning* dalam epidemiologi dan mendukung upaya mitigasi wabah.

3.2 Metode Penelitian

Metode penelitian ini dirancang untuk membangun dan mengevaluasi model ANN dalam memprediksi kasus MPXV dengan cakupan data global dan rentang waktu yang lebih panjang. Selain itu, penelitian ini membandingkan beberapa algoritma optimasi untuk meningkatkan akurasi prediksi model. Gambar 3.3 mengilustrasikan tahapan utama dalam proses penelitian.

3.2.1 Studi Literatur

Studi literatur mencakup penelaahan berbagai sumber informasi, seperti jurnal ilmiah, buku, situs web resmi, dan laporan penelitian yang berkaitan dengan prediksi epidemiologi, ANN, metode optimasi, serta pemodelan penyebaran *Monkeypox Virus* (MPXV). Analisis terhadap literatur ini bertujuan untuk memperoleh pemahaman yang komprehensif mengenai kerangka konseptual, teknologi, dan metodologi yang relevan dengan penelitian ini.

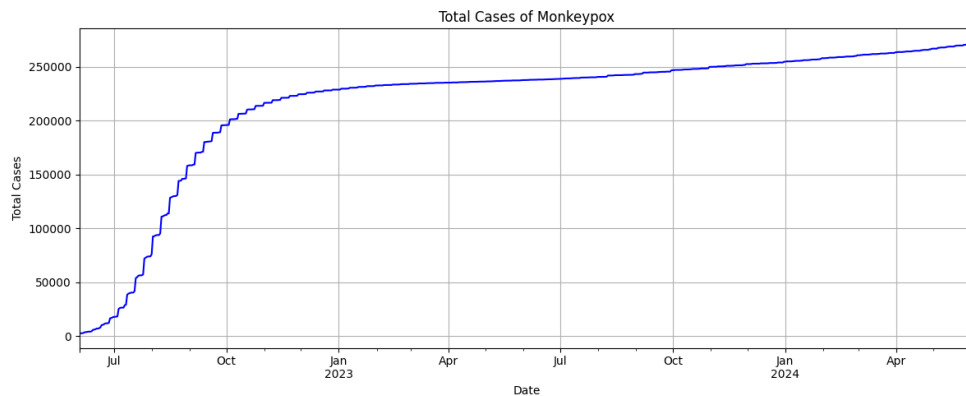


Gambar 3.2 Tahap Penelitian

3.2.2 Data Collecting

Dataset yang digunakan dalam penelitian ini merupakan kumpulan data harian mengenai kasus MPXV global yang diunduh secara langsung dari *Our World*

in Data. Data yang diperoleh terdiri atas 732 data *entry* yang menyajikan informasi komprehensif yang terdiri dari *location* (lokasi pelaporan), *date* (tanggal pengambilan data), *iso_code* (kode negara), *total_cases* (jumlah total kasus), *total_deaths* (jumlah total kematian), *new_cases* (kasus baru), *new_deaths* (kematian baru), *new_cases_smoothed* (rata-rata kasus baru yang dihaluskan), *new_deaths_smoothed* (rata-rata kematian baru yang dihaluskan), *new_cases_per_million* (kasus baru per juta penduduk), *total_cases_per_million* (total kasus per juta penduduk), *new_cases_smoothed_per_million* (kasus baru yang dihaluskan per juta penduduk), *new_deaths_per_million* (kematian baru per juta penduduk), *total_deaths_per_million* (total kematian per juta penduduk), *new_deaths_smoothed_per_million* (kematian baru yang dihaluskan per juta penduduk), *suspected_cases_cumulative* (jumlah kumulatif kasus yang dicurigai), dan *annotation* (keterangan). Adapun data yang dipakai mencakup informasi tanggal dan total kasus, kemudian disusun dalam format *time series*. Data diunduh dari repository resmi sehingga keakuratan dan konsistensinya dapat dipertanggungjawabkan, untuk grafik tren total kasus global MPXV ditunjukan pada Gambar 3.4. Dataset tersebut selanjutnya digunakan sebagai basis untuk analisis mendalam serta pengembangan model prediktif berbasis ANN yang bertujuan untuk memproyeksikan kasus cacar monyet secara global.



Gambar 3.3 Grafik tren total kasus

3.2.3 Data Preprocessing

Tahap pra-pemrosesan data dalam penelitian ini dilakukan untuk memastikan bahwa dataset yang digunakan memiliki kualitas yang optimal sebelum diterapkan pada model prediktif. Proses ini meliputi beberapa tahapan sebagai berikut:

a. Pemeriksaan dan penanganan *missing values*

Data awal diperiksa untuk mendeteksi adanya nilai yang hilang, apabila ditemukan *missing values*, metode *backward fill* (bfill) diterapkan untuk mengisi nilai yang kosong dengan data dari *entry* berikutnya. Pemeriksaan terhadap nilai hilang dilakukan menggunakan fungsi `.isnull().sum()` untuk mengetahui apakah terdapat *entry* kosong dalam kolom terutama kolom *total_cases*.

b. Konversi Kolom Tanggal dan Penetapan Indeks Waktu

Data memiliki kolom *date* yang merepresentasikan informasi waktu. Kolom ini dikonversi ke format *datetime* dan kemudian dijadikan sebagai indeks *DataFrame*. Penetapan indeks waktu penting agar struktur data memenuhi

karakteristik deret waktu (*time series*), sehingga memungkinkan analisis temporal dilakukan secara akurat.

c. Normalisasi Data Input dan Target

Seluruh fitur input (x) dan target output (y) dinormalisasi menggunakan teknik *Min-Max Scaling*. Normalisasi diperlukan agar model ANN dapat belajar secara efisien, terutama karena model ini sensitif terhadap skala data input. metode penskalaan linear yang mengubah nilai asli menjadi rentang $[0,1]$ berdasarkan nilai minimum dan maksimum dalam dataset dengan menggunakan persamaan sebagai berikut:

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (5)$$

dimana :

X : nilai asli data

X_{min} : nilai terkecil dalam kolom

X_{max} : nilai terbesar dalam kolom

Proses normalisasi dilakukan menggunakan pustaka *MinMaxScaler* dari library *Scikit-Learn*. Hasil prediksi akan dikembalikan ke skala aslinya melalui proses denormalisasi agar dapat dibandingkan secara langsung dengan nilai aktual setelah proses prediksi selesai.

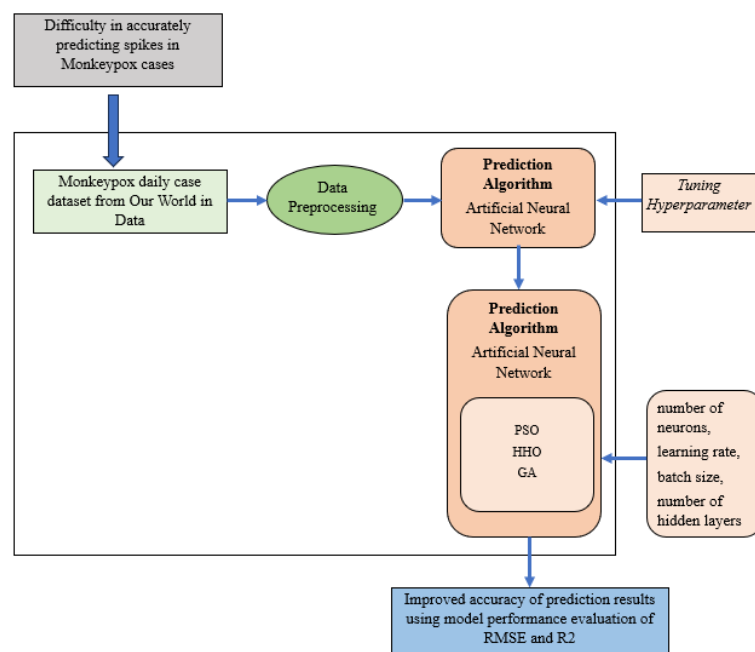
3.2.4 *Splitting data*

Data yang telah dibentuk dalam format input-output dan dinormalisasi, kemudian dibagi menjadi data pelatihan (80%) dan data pengujian (20%) menggunakan fungsi *train_test_split*. Skema pembagian ini dipilih dengan merujuk pada penelitian oleh (Sakti dkk., 2024) yang menunjukkan bahwa proporsi 80:20

menghasilkan nilai *Mean Squared Error* (MSE) terendah dibandingkan dengan proporsi lainnya, seperti 90:10 dan 70:30, dalam konteks peramalan deret waktu menggunakan model ANN. Berdasarkan temuan tersebut, pembagian 80:20 dipandang sebagai skenario yang optimal untuk menghasilkan model dengan akurasi prediksi yang lebih baik.

Selanjutnya, untuk memperoleh evaluasi performa model yang lebih menyeluruh dan tidak bergantung pada satu subset data tertentu, dilakukan proses validasi menggunakan *5-Fold Time Series Cross Validation* pada data pelatihan. Teknik ini membagi data latih menjadi sepuluh bagian yang digunakan secara bergantian sebagai data validasi, sementara sisanya digunakan untuk pelatihan. Nilai *Root Mean Squared Error* (RMSE) dihitung pada setiap fold, kemudian dirata-ratakan sebagai representasi kinerja model yang lebih stabil dan objektif.

3.2.5 Modelling



Gambar 3.4 Kerangka penelitian

Gambar 3.4 menunjukkan kerangka penelitian dengan model prediksi yang digunakan berbasis ANN yang terdiri atas tiga lapisan utama, yaitu *input layer*, *hidden layer* serta *output layer*. Fungsi aktivasi *ReLU* diterapkan pada setiap layer guna mengatasi kompleksitas pola non-linear dalam data. Model dilatih dengan *Mean Squared Error* (MSE) sebagai fungsi *loss*. Dilakukan optimasi parameter model melalui penerapan tiga algoritma algoritma optimasi untuk meningkatkan kinerja prediksi sehingga penelitian ini membandingkan empat skema, sebagai berikut pada tabel 3.1

Tabel 3.1 Skema Model

Skema	Algoritma	Parameter	Hyperparameter yang Dioptimasi
1	ANN Baseline	– 1 <i>hidden layer</i>, 64 <i>neurons</i> (<i>ReLU</i>) – <i>Output layer</i>: 1 <i>neuron</i> (<i>linear</i>) – <i>Optimizer</i>: Adam (default) – <i>Loss</i>: MSE – <i>Epoch</i>: 100 – <i>Batch Size</i>: 32 – <i>Learning Rate</i>: 0.001 – <i>Callback</i>: <i>EarlyStopping</i> (<i>patience</i>=10)	<i>Custom parameter baseline</i>
2	ANN-PSO	– Partikel: 10 – Iterasi: 10 – <i>Inertia weight</i>: 0.5 – <i>Learning factors</i>: $c_1 = 1.5$, $c_2 = 1.5$ – <i>Fitness</i>: RMSE	– <i>num_neurons</i>: 8 – 128 – <i>learning_rate</i>: 0.0001 - 0.01 – <i>batch_size</i>: 16 – 64 – <i>num_layers</i>: 1 - 3
3	ANN-HH0	– Hawks: 10 – Iterasi: 10 – Strategi: <i>Soft/Hard besiege</i> + <i>Rapid dives</i> – Energi awal $E_0 \in [-1, 1]$ – <i>Fitness</i>: RMSE	– <i>num_neurons</i>: 8 – 128 – <i>learning_rate</i>: 0.0001 - 0.01 – <i>batch_size</i>: 16 – 64 – <i>num_layers</i>: 1 - 3
4	ANN-GA	– Populasi: 10 – Generasi: 10	– <i>num_neurons</i>: 8 – 128

		– <i>Parent selection</i>: 4 terbaik – <i>Crossover</i>: 1 titik acak – <i>Mutation rate</i>: 0.1 – <i>Fitness</i>: RMSE	– <i>learning_rate</i>: 0.0001 - 0.01 – <i>batch_size</i>: 16 – 64 – <i>num_layers</i>: 1 - 3
--	--	---	---

Skema 1 merupakan model dasar tanpa penerapan algoritma optimasi khusus. Pada Skema 2, optimasi dilakukan dengan menggunakan HHO untuk mengeksplorasi ruang solusi secara efektif. Skema 3 memanfaatkan PSO yang bekerja berdasarkan prinsip *swarm intelligence*, sedangkan Skema 4 menerapkan GA untuk optimasi secara *evolusioner*. Pendekatan ini diharapkan dapat meningkatkan kinerja prediksi ANN dalam memproyeksikan kasus cacar monyet secara global.

3.2.7 Evaluasi

Evaluasi dilakukan untuk menilai seberapa baik model dalam memprediksi kasus MPXV berdasarkan data aktual. Penelitian ini menggunakan dua matrik evaluasi utama, yaitu *Root Mean Squared Error* (RMSE) dan *R-Squared* (R^2). RMSE digunakan untuk mengukur rata-rata besar kesalahan antara nilai prediksi dan nilai aktual, dalam satuan yang sama dengan target aslinya. Semakin kecil nilai RMSE, semakin baik model dalam menghasilkan prediksi yang akurat. Sementara itu, R^2 mengukur proporsi varian dari data target yang berhasil dijelaskan oleh model. Nilai R^2 yang mendekati 1 menunjukkan bahwa model memiliki kemampuan prediksi yang sangat baik. Adapun rumus matematis dari kedua matrik tersebut tercantum pada persamaan (3) dan persamaan (4).

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y})^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum (y_i - \hat{y})^2}{\sum (y_i - \bar{y})^2} \quad (4)$$

dimana :

y_i : nilai aktual

$\hat{y}$: nilai prediksi

$\bar{y}$: nilai rata-rata dari nilai aktual

n : jumlah data

3.2.8 Penarikan Kesimpulan

Setelah melakukan evaluasi terhadap kinerja model, tahap selanjutnya adalah penarikan kesimpulan. Kesimpulan yang diambil dari tahap ini memberikan informasi tentang kinerja model dan memberikan saran perbaikan untuk penelitian selanjutnya.