

PREDIKSI PENYEBARAN MALARIA DI KABUPATEN ASAHAN MENGGUNAKAN  
METODE MULTILAYER PERCEPTRON (MLP) DAN WHALE OPTIMIZATION  
ALGORITHM (WOA)

SKRIPSI

LEVINA GUNAWAN

201402052



PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2025

PREDIKSI PENYEBARAN MALARIA DI KABUPATEN ASAHAN MENGGUNAKAN METODE  
MULTILAYER PERCEPTRON (MLP) DAN WHALE OPTIMIZATION ALGORITHM (WOA)

SKRIPSI

Diajukan untuk melengkapi tugas dan memenuhi syarat memperoleh ijazah Sarjana Teknologi Informasi

LEVINA GUNAWAN

201402052



PROGRAM STUDI S1 TEKNOLOGI INFORMASI  
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA  
MEDAN  
2025

**PERSETUJUAN**

Judul : PREDIKSI PENYEBARAN MALARIA DI  
KABUPATEN ASAHAN MENGGUNAKAN  
METODE MULTILAYER PERCEPTRON (MLP)  
DAN WHALE OPTIMIZATION ALGORITHM (WOA)

Kategori : SKRIPSI

Nama : LEVINA GUNAWAN

Nomor Induk Mahasiswa : 201402052

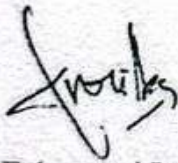
Program Studi : SARJANA (S-1) TEKNOLOGI INFORMASI

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI INFORMASI  
UNIVERSITAS SUMATERA UTARA

Medan, 6 Januari 2025

Komisi Pembimbing:

Pembimbing 2



Fahrurrozi Lubis, B.IT., M.Sc.IT.

NIP. 198610122018052001

Pembimbing 1



Prof. Dr. Romi Fadillah Rahmat, B.Comp.Sc.,  
M.Sc.

NIP. 198603032010121004

Diketahui/disetujui oleh

Program Studi S1 Teknologi Informasi



Dedy Arisandi, S.T., M.Kom.

NIP. 197908312009121002

**PERNYATAAN**

**PREDIKSI PENYEBARAN MALARIA DI KABUPATEN ASAHAN  
MENGUNAKAN METODE MULTILAYER PERCEPTRON  
(MLP) DAN WHALE OPTIMIZATION  
ALGORITHM (WOA)**

**SKRIPSI**

Saya mengakui bahwa skripsi ini adalah hasil karya saya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing telah disebutkan sumbernya.

Medan, 6 Januari 2025



Levina Gunawan

201402052



## UCAPAN TERIMA KASIH

Puji Syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas berkat dan rahmat-Nya, penulis dapat menyelesaikan skripsi dengan judul “PREDIKSI PENYEBARAN MALARIA DI KABUPATEN ASAHAN MENGGUNAKAN METODE MULTILAYER PERCEPTRON (MLP) DAN WHALE OPTIMIZATION ALGORITHM (WOA)”. Penulis membuat laporan ini dalam rangka memenuhi salah satu persyaratan akademis dalam rangka menyelesaikan kuliah di fakultas Ilmu Komputer dan Teknologi Informasi Program Studi S1 Teknologi Informasi Universitas Sumatera Utara. Penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak sangatlah sulit bagi penulis untuk menyelesaikan penelitian ini. Oleh karena itu, Penulis mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Romi Fadillah Rahmat, B.Comp.Sc., M.Sc. selaku dosen pembimbing 1 atas ilmu, waktu, dan kesabaran dalam penulisan skripsi ini.
2. Bapak Fahrurrozi Lubis, B.IT., M.Sc.IT. selaku dosen pembimbing 2 atas ilmu, waktu, dan kesabaran dalam penulisan skripsi ini.
3. Bapak Mohammad Fadly Syah Putra, M.Sc. selaku dosen pembimbing 1 dan Bapak Ainul Hizriadi S.Kom., M.Sc. selaku dosen pembimbing 2 atas saran dan masukan yang diberikan kepada penulis.
4. Bapak Dedy Arisandi, S.T., M.Kom. dan Bapak Ivan Jaya S.Si., M.Kom. selaku Ketua dan Sekretaris Program Studi S1 Teknologi Informasi Universitas Sumatera Utara.
5. Semua pihak yang telah membantu hingga terselesaikannya skripsi ini.

Akhir kata, penulis berharap semoga Tuhan Yang Maha Esa berkenan membalas segala kebaikan semua pihak yang telah membantu dan semoga skripsi ini membawa manfaat.

Medan, 6 Januari 2025

Penulis



Levia Gunawan

201402052



## ABSTRAK

Malaria merupakan penyakit yang dapat menular melalui gigitan nyamuk. Saat ini, malaria telah menjadi salah satu penyakit menular dengan tingkat kematian yang tinggi. Menurut data dari *World Health Organization*, diperkirakan sebanyak 249 juta kasus malaria telah tercatat secara global pada tahun 2022. Di Indonesia sendiri, kasus malaria mencapai total 1,2 juta kasus. Kabupaten Asahan menjadi kabupaten yang tergolong endemis sedang di Sumatera Utara. Selain itu, faktor iklim meliputi suhu, curah hujan, kelembapan, dan lain-lain juga mempengaruhi penyebaran malaria secara signifikan. Sehubungan dengan tujuan nasional pemerintah untuk membasmi malaria pada tahun 2030, diperlukan adanya model prediksi yang akurat untuk memprediksi angka penyebaran malaria sehingga penanganan dapat dilakukan secara efektif. Penelitian ini mengajukan model MLP-WOA yang merupakan *hybrid Multilayer Perceptron* (MLP) dan *Whale Optimization Algorithm* (WOA) untuk memprediksi angka penyebaran malaria di Kabupaten Asahan, Sumatera Utara secara mingguan. Peneliti juga akan melakukan sintesis dataset untuk tahun 2015-2018 dengan menggunakan TimeGAN. Model yang paling optimal menghasilkan nilai *test* MAE, MSE, RMSE, dan  $R^2$  sebesar 5.320, 40.910, 6.396, dan 0.808 berturut-turut. Parameter model yang digunakan yaitu *split data* 7:3, 2000 iterasi, 20 agen pencari, 2 *hidden layer*, serta 5 dan 4 *neuron* pada setiap *hidden layer*. *TimeGAN* membantu menambahkan data historikal dan meningkatkan performa model dengan cukup signifikan. Curah hujan, banjir dan genangan air, serta kepadatan penduduk menunjukkan pengaruh signifikan terhadap kasus positif malaria di Kabupaten Asahan

**Kata Kunci:** Prediksi, Penyebaran Malaria, *Multilayer Perceptron*, *Whale Optimization Algorithm*, *Generative Adversarial Network*, *TimeGAN*, *Hybrid*

**PREDICTION OF MALARIA ABUNDANCE IN ASAHAN REGENCY USING  
MULTILAYER PERCEPTRON (MLP) AND WHALE OPTIMIZATION  
ALGORITHM (WOA) METHOD**

**ABSTRACT**

*Malaria is a disease that can be transmitted through mosquito bites. Today, malaria has become an infectious disease with a high mortality rate. According to data from the World Health Organization, an estimated 249 million cases of malaria have been recorded globally by 2022. In Indonesia alone, malaria cases will reach a total of 1.2 million cases by 2022. Asahan Regency is one of the districts that is classified as moderately endemic in North Sumatra. In addition, climatic factors including temperature, rainfall, humidity, and others also significantly affect the spread of malaria. In connection with the government's national goal to eradicate malaria by 2030, a more accurate prediction model is needed to predict the number of malaria abundance so treatment can be performed effectively. This research proposes the MLP-WOA model which is a hybrid of Multilayer Perceptron (MLP) and Whale Optimization Algorithm (WOA) method to predict the number of malaria abundance in Asahan Regency, North Sumatra on a weekly basis. This research will also synthesize the dataset for 2015-2018 using TimeGAN. The most optimal model produced test MAE, MSE, RMSE, and  $R^2$  of 5.320, 40.910, 6.396, and 0.808, respectively. The model parameters used were a data split of 7:3, 2000 iterations, 20 search agents, 2 hidden layers, and 5 and 4 neurons in each hidden layer. TimeGAN helped to add historical data and improved the model performance significantly. Rainfall, floods and puddles, and population density showed a significant influence on positive malaria cases in Asahan Regency.*

**Keywords:** Prediction, Malaria Abundance, Multilayer Perceptron, Whale Optimization Algorithm, Generative Adversarial Network, TimeGAN, Hybrid

## DAFTAR ISI

PERSETUJUAN .....	ii
PERNYATAAN.....	iii
UCAPAN TERIMA KASIH .....	iv
ABSTRAK.....	v
<i>ABSTRACT</i> .....	vi
DAFTAR ISI .....	vii
DAFTAR GAMBAR .....	ix
DAFTAR TABEL .....	xi
BAB 1 PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	4
1.3. Tujuan Penelitian.....	4
1.4. Batasan Penelitian .....	4
1.5. Manfaat Penelitian.....	4
1.6. Metodologi Penelitian .....	5
1.7. Sistematika Penulisan.....	6
BAB 2 LANDASAN TEORI.....	8
2.1. Malaria .....	8
2.2. Jenis Malaria .....	8
2.3. Hubungan Faktor Meteorologi dengan Malaria .....	10
2.4. Iklim dan Topografi di Kabupaten Asahan.....	10
2.5. Generative Adversarial Networks .....	11
2.6. Multilayer Perceptron.....	12
2.7. Whale Optimization Algorithm .....	14
2.8. Penelitian Terdahulu.....	17
2.9. Perbedaan Penelitian .....	21
BAB 3 ANALISIS DAN PERANCANGAN SISTEM .....	24
3.1 Data .....	24
3.2 Arsitektur Umum.....	24
3.2.1 Data Collection .....	25
3.2.2 Data Pre-processing .....	25
3.2.3 Data Splitting .....	28



3.2.4	Model Training.....	28
3.2.5	Model Evaluation.....	33
3.2.6	Output .....	34
3.3	Rancangan Antarmuka Sistem .....	34
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM .....		40
4.1	Implementasi Sistem .....	40
4.1.1	Spesifikasi Perangkat .....	40
4.1.2	Implementasi Tampilan Antarmuka Sistem .....	40
4.2	Implementasi dan Pengujian Model Data Tahun 2019-2022 .....	46
4.2.1	Dataset Tahun 2019-2022 .....	46
4.2.2	Implementasi Model .....	47
4.2.3	Hasil Pengujian Model.....	51
4.2.4	Evaluasi.....	54
4.3	Implementasi dan Pengujian Model Data Tahun 2015-2022 .....	54
4.3.1	Dataset Tahun 2015-2022 .....	54
4.3.2	Implementasi Model .....	57
4.3.3	Hasil Pengujian Model.....	60
4.3.4	Evaluasi.....	65
4.4	Diskusi.....	66
BAB 5 KESIMPULAN DAN SARAN .....		68
5.1	Kesimpulan.....	68
5.2	Saran.....	68
DAFTAR PUSTAKA.....		69

## DAFTAR GAMBAR

Gambar 2.1	<i>Generative Adversarial Networks</i>	12
Gambar 2.2	<i>Arsitektur Multilayer Perceptron</i>	14
Gambar 2.3	<i>Shrinking Encircling Mechanism</i>	16
Gambar 2.4	<i>Spiral Updating Position</i>	16
Gambar 2.5	<i>Pseudocode Algoritma WOA</i>	17
Gambar 3.2.1	<i>Arsitektur Umum</i>	25
Gambar 3.2.2	<i>Pseudocode proses data cleaning</i>	26
Gambar 3.2.3	<i>Pseudocode proses synthetic data generation</i>	27
Gambar 3.2.4	<i>Pseudocode model training MLP-WOA</i>	29
Gambar 3.3.1	<i>Rancangan tampilan home page</i>	34
Gambar 3.3.2	<i>Rancangan tampilan training page</i>	35
Gambar 3.3.3	<i>Rancangan tampilan training data</i>	35
Gambar 3.3.4	<i>Rancangan tampilan training metrics</i>	36
Gambar 3.3.5	<i>Rancangan tampilan testing page</i>	36
Gambar 3.3.6	<i>Rancangan tampilan testing comparison</i>	37
Gambar 3.3.7	<i>Rancangan tampilan testing plot</i>	37
Gambar 3.3.8	<i>Rancangan tampilan predicting page</i>	38
Gambar 3.3.9	<i>Rancangan tampilan predicting result</i>	38
Gambar 3.3.10	<i>Rancangan tampilan predicting plot</i>	39
Gambar 4.1.1	<i>Implementasi tampilan home page</i>	41
Gambar 4.1.2	<i>Implementasi tampilan training page</i>	41

Gambar 4.1.3	Implementasi tampilan <i>training data</i>	42
Gambar 4.1.4	Implementasi tampilan <i>training metrics</i>	42
Gambar 4.1.5	Implementasi tampilan <i>testing page</i>	43
Gambar 4.1.6	Implementasi tampilan <i>testing comparison</i>	43
Gambar 4.1.7	Implementasi tampilan <i>testing plot</i>	44
Gambar 4.1.8	Implementasi tampilan <i>predicting page</i>	44
Gambar 4.1.9	Implementasi tampilan <i>predicting result</i>	45
Gambar 4.1.10	Implementasi tampilan <i>predicting plot</i>	45
Gambar 4.1.11	Implementasi tampilan <i>requirement page</i>	46
Gambar 4.2.1	<i>Correlation Heatmap</i> data kasus malaria 2019-2022	46
Gambar 4.2.2	Plot optimasi <i>training data</i> tiap iterasi tahun 2019-2022	50
Gambar 4.2.3	Plot perbandingan nilai aktual dan prediksi <i>test data</i> 2019-2022	54
Gambar 4.3.2	Plot perbandingan data asli dan data sintesis	55
Gambar 4.3.2	Plot perbandingan PCA dan TSNE data asli dan data sintesis	56
Gambar 4.3.3	<i>Correlation Heatmap</i> data kasus malaria tahun 2015-2022	56
Gambar 4.3.4	Plot optimasi <i>training data</i> tiap iterasi tahun 2015-2022	60
Gambar 4.3.5	Plot perbandingan nilai aktual dan prediksi <i>test data</i> 2015-2022	65

## DAFTAR TABEL

Tabel 3.2.1	Tabel variabel indeks tingkat banjir dan total kasus	28
Tabel 3.2.2	Tabel perbandingan nilai aktual dan prediksi untuk MAE	31
Tabel 3.2.3	Tabel perbandingan nilai aktual dan prediksi untuk RMSE	32
Tabel 3.2.4	Tabel perbandingan nilai aktual dan prediksi untuk MSE	33
Tabel 3.2.5	Tabel perbandingan nilai aktual dan prediksi untuk $R^2$	33
Tabel 4.2.1	Eksperimen parameter <i>split data</i> tahun 2019-2022	47
Tabel 4.2.2	Eksperimen jumlah <i>hidden layer</i> tahun 2019-2022	48
Tabel 4.2.3	Eksperimen jumlah <i>neuron</i> tahun 2019-2022	48
Tabel 4.2.4	Eksperimen jumlah agen pencari tahun 2019-2022	49
Tabel 4.2.5	Eksperimen jumlah iterasi tahun 2019-2022	50
Tabel 4.2.6	Perbandingan nilai aktual dan prediksi pada <i>test data</i> 2019-2022	51
Tabel 4.2.7	Evaluasi hasil prediksi tahun 2019-2022	54
Tabel 4.3.1	Eksperimen <i>hyperparameter TimeGAN</i>	55
Tabel 4.3.2	Eksperimen parameter <i>split data</i> tahun 2015-2022	57
Tabel 4.3.3	Eksperimen jumlah <i>hidden layer</i> tahun 2015-2022	57
Tabel 4.3.4	Eksperimen jumlah <i>neuron</i> tahun 2015-2022	58
Tabel 4.3.5	Eksperimen jumlah agen pencari tahun 2015-2022	59
Tabel 4.3.6	Eksperimen jumlah iterasi tahun 2015-2022	60
Tabel 4.3.7	Perbandingan nilai aktual dan prediksi pada <i>test data</i> 2015-2022	61
Tabel 4.3.8	Evaluasi hasil prediksi tahun 2015-2022	66



## **BAB 1**

### **PENDAHULUAN**

#### **1.1. Latar Belakang**

Malaria merupakan penyakit yang telah mempengaruhi populasi manusia sejak zaman dahulu. Saat ini, malaria masih menjadi penyakit dengan tingkat kematian yang tinggi (Garrido-Cardenas et al., 2019). Nyamuk *anopheles* betina yang terinfeksi parasit *Plasmodium* menjadi vektor penyebab penyakit ini. Data dari *World Health Organization* menunjukkan bahwa kasus malaria secara global mencapai 249 juta kasus yang tersebar di 85 negara endemis malaria pada tahun 2022. Ini menunjukkan adanya peningkatan kasus sebanyak 5 juta kasus dibandingkan dengan tahun 2021. Kasus malaria telah mengalami penurunan dari tahun 2000 sebanyak 243 juta kasus hingga tahun 2014 sebanyak 230 juta kasus (World Health Organization, 2023). Namun, pada tahun 2015, kasus malaria kembali mengalami peningkatan hingga mencapai puncaknya pada tahun 2020 (tahun pertama pandemi COVID) yang mengalami peningkatan hingga 11 juta kasus dengan total 244 juta kasus dibandingkan tahun sebelumnya. Sementara itu, angka kematian akibat malaria pada tahun 2022 mencapai 608000 kasus (World Health Organization, 2023).

Menurut *World Health Organization*, Kawasan Asia Tenggara berkontribusi sebanyak 5,2 juta kasus malaria pada tahun 2022, yakni 2% dari kasus malaria global (World Health Organization, 2023). Hal ini menunjukkan adanya penurunan hingga 7 juta kasus dibandingkan dengan tahun 2021. Adapun angka kematian malaria menurun 2000 kasus pada tahun 2022. Walaupun begitu, Indonesia tercatat sebagai negara dengan peningkatan kasus malaria tertinggi kedua setelah India, yakni mencapai 1,2 juta kasus (World Health Organization, 2023). Berdasarkan data dari Badan Pusat Statistik (BPS), kasus malaria per seribu penduduk atau *Annual Parasite Incidence* (API) di Sumatera Utara adalah sebesar 0.17 pada tahun 2021 yang mana telah mengalami kenaikan 0.1 dari tahun 2020 (Badan Pusat Statistik, 2024). 22 dari 33 kabupaten/kota di Sumatera Utara juga telah dinyatakan mencapai eliminasi malaria pada tahun 2022 (Badan Pusat Statistik, 2024). Walaupun begitu, kasus malaria masih rentan terjadi di area yang sulit dijangkau dengan kondisi geografis yang kurang mendukung dan mobilitas tinggi sehingga menjadi tantangan dalam upaya mengeliminasi malaria di Sumatera Utara (Fahmi et al., 2022). Kabupaten Asahan merupakan kabupaten di Sumatera Utara yang termasuk dalam endemis sedang dengan API sebesar 1-5 per 1000

penduduk (Sinaga, 2022). Penelitian yang dilakukan oleh Fahmi et al pada tahun 2022 menunjukkan bahwa Kabupaten Asahan merupakan salah satu lokasi dengan kasus malaria tertinggi di Sumatera Utara (Fahmi et al., 2022).

Dalam upaya mempercepat kemajuan menuju eliminasi malaria, *World Health Organization* (WHO) menyediakan sebuah kerangka kerja komprehensif dalam memandu negara-negara yang disebut dengan *Global Technical Strategy* (GTS). Strategi ini menetapkan target untuk mengurangi angka kejadian dan angka kematian. Secara umum, kawasan Asia Tenggara mengikuti jalur untuk mencapai target GTS 2025 dan GTS 2030. Walaupun begitu, Indonesia tidak mengalami penurunan kasus sejak tahun 2015, bahkan menjadi satu-satunya negara dengan angka kematian malaria yang meningkat di Asia Tenggara (World Health Organization, 2023). Sehubungan dengan target Indonesia untuk membasmi penyakit malaria pada tahun 2030, 5 regional ditetapkan oleh kementerian kesehatan Indonesia sebagai target eliminasi. Sumatera Utara menjadi salah satu kawasan yang menjadi target eliminasi malaria pada regional kedua (Rokom, 2022).

Sehubungan dengan target pemerintah untuk mengeliminasi malaria pada tahun 2030, diperlukan adanya model prediksi yang akurat dalam memprediksi angka kejadian malaria sehingga tindakan pencegahan dapat dilakukan secara dini dan efektif. Berbagai model prediksi telah diusulkan pada penelitian terdahulu untuk memprediksi malaria, baik dari resiko penyebaran, maupun angka penyebaran. Penelitian terkait juga telah dilakukan dengan berbagai metode, yaitu *traditional machine learning model* seperti yang dilakukan oleh Kim, Yoonhee et al. (2019); Nkiruka, Odu et al. (2021); Adamu & Singh (2021); Mbunge, E. et al. (2022); Chen, S. et al. (2019); Kalipe, G. et al. (2018); Aheto, J. M. K. et al. (2021), *deep learning model* seperti yang dilakukan oleh Mfisimana, Lionel Divin et al. (2022); Singh, A. et al. (2023); Thakur & Dharavath (2019), dan *time-series model* seperti yang dilakukan oleh Zinszer, K. et al. (2015); Santosh et al. (2020); Haddawy, P. et al. (2018); Kurnianingsih, K. et al. (2020); Wang, Mengyang et al. (2019); Fisher, T. M. et al. (2023); Adeola, A. M. et al. (2019); Mopuri, R. et al. (2020). Berdasarkan penelitian Mfisimana, Lionel Divin et al., peneliti membandingkan *machine learning model* dan *deep learning model* dalam memprediksi angka kejadian malaria pada 4 sub-grup. Mfisimana, Lionel Divin et al. membandingkan *Generalized Linear Model* (Bates et al., 2014) dan *Artificial Neural Network* (Gurney, 1997) dan memperoleh bahwa ANN unggul dalam hal performa dan

akurasi dibandingkan dengan GLM yang menunjukkan bahwa *deep learning model* memiliki keakuratan yang lebih tinggi dibandingkan dengan *machine learning model* (Mfisimana et al., 2022). Selain itu, banyak penelitian terdahulu yang menggunakan *time series model* dalam memprediksi penyebaran malaria, seperti *Long Short-term Memory*, *Auto-Regressive Integrated Moving Average*, dan *Seasonal Auto-Regressive Integrated Moving Average* dan menunjukkan bahwa metode tersebut cocok dipakai untuk *seasonal data* seperti data malaria. Penelitian terdahulu telah menunjukkan bahwa variabel iklim, seperti curah hujan, suhu, kelembapan, radiasi, dan lain-lain mempengaruhi penyebaran malaria secara signifikan (Nkiruka, Odu et al., 2021; Santosh, Thakur et al., 2020; Mfisimana, Lionel Divin et al., 2022; Kim, Yoohee et al., 2019; Adamu & Singh, 2021; Fahmi et al., 2022; Wang, Mengyang et al., 2019).

Peneliti Alameer, Zakaria et al. (2019) mengajukan model *hybrid Multilayer Perceptron-Whale Optimization Algorithm* (MLP-WOA) (Mirjalili & Lewis, 2016) dalam memprediksi harga emas. Peneliti kemudian membandingkan performa algoritma ini dengan algoritma optimasi lain, seperti *Particle Swarm Optimization* (Kennedy & Eberhart, 1995), *Genetic Algorithm* (Holland, 1992), *Grey Wolf Optimization* (Mirjalili & Lewis, 2014), serta model *time series* yakni ARIMA. Hasil penelitian menunjukkan bahwa model MLP-WOA dapat meningkatkan performa prediksi *time series data* dibandingkan dengan algoritma optimasi lainnya (Alameer et al., 2019). Penelitian-penelitian lain dengan tipe data *time series* juga telah dilakukan dengan menggunakan model MLP-WOA, seperti prediksi kecepatan angin (Samadianfard et al., 2020) dan prediksi curah hujan (Diop et al., 2020). Oleh karena model *hybrid* MLP-WOA belum pernah digunakan dalam memprediksi angka penyebaran malaria yang juga merupakan *time series data*, penulis mengusulkan untuk mengimplementasi metode MLP-WOA dalam memprediksi angka penyebaran malaria untuk memperoleh hasil prediksi yang lebih akurat. Penelitian terdahulu juga telah menunjukkan bahwa meningkatnya kasus malaria berhubungan erat dengan faktor iklim. Oleh karena itu, penulis mengusulkan untuk melakukan penelitian berjudul “Prediksi Penyebaran Malaria di Kabupaten Asahan Menggunakan Metode Multilayer Perceptron (MLP) dan Whale Optimization Algorithm (WOA)”.



## 1.2. Rumusan Masalah

Untuk mencapai tujuan nasional pemerintah dalam membasmi malaria pada tahun 2030, dibutuhkan model prediksi yang akurat untuk memprediksi tingkat penyebaran malaria. Penggunaan algoritma *Multilayer Perceptron* (MLP) dan *Whale Optimization Algorithm* (WOA) masih belum digunakan dalam memprediksi penyebaran malaria. Penggunaan kedua algoritma ini diharapkan dapat menghasilkan tingkat akurasi dan performa yang baik dalam memprediksi angka penyebaran malaria.

## 1.3. Tujuan Penelitian

Penelitian ini bertujuan untuk mengimplementasikan metode *hybrid Multilayer Perceptron* (MLP) dan *Whale Optimization Algorithm* (WOA) untuk memperoleh prediksi yang akurat dalam memprediksi angka penyebaran malaria di Kabupaten Asahan.

## 1.4. Batasan Penelitian

Penulis membatasi ruang lingkup penelitian agar tidak terjadi penyimpangan dalam ruang lingkup penelitian, yakni:

1. Objek wilayah penelitian ini hanya terbatas pada wilayah Kabupaten Asahan
2. Faktor penelitian yang dipertimbangkan hanya terbatas pada hubungan antara data iklim dan angka kejadian malaria.
3. Variabel pada penelitian ini mencakup suhu maksimum, durasi sinar matahari, indeks jumlah hujan, indeks curah hujan, indeks tingkat banjir dan genangan air, indeks kepadatan dan pertambahan penduduk, dan total kasus positif malaria.

## 1.5. Manfaat Penelitian

Manfaat dari penelitian yang dilakukan yaitu:

1. Memperoleh model prediksi yang akurat dengan menggunakan algoritma *hybrid*.
2. Mengetahui potensi kenaikan kasus positif malaria sehingga masyarakat dapat lebih waspada.
3. Membantu dinas kesehatan dalam melakukan tindakan pencegahan berupa antisipasi terhadap potensi penyebaran malaria yang tinggi.



## 1.6. Metodologi Penelitian

Adapun metodologi penelitian yang terdapat pada penelitian ini adalah:

### 1. Tinjauan Pustaka

Tinjauan pustaka merupakan tahapan pertama yang dilakukan penulis untuk mencari permasalahan dan menentukan judul penelitian. Penulis mengumpulkan informasi melalui berbagai sumber, seperti jurnal, *conference paper*, buku, artikel, dan lain-lain. Penulis mengumpulkan informasi yang membahas mengenai malaria, prediksi penyebaran malaria, *Multilayer Perceptron*, *Whale Optimization Algorithm*, *Generative Adversarial Networks*, dan lain-lain.

### 2. Pengkajian Masalah

Setelah literatur, penulis mulai melakukan analisis terhadap masalah terkait penyebaran kasus malaria serta faktor-faktor yang dapat mempengaruhi tinggi rendahnya penyebaran tersebut berdasarkan sumber-sumber literatur yang telah diperoleh sebelumnya. Selain itu, penulis juga melakukan pemahaman mengenai algoritma *Multilayer Perceptron* dan *Whale Optimization Algorithm* serta bagaimana kedua metode tersebut dapat dihubungkan untuk meningkatkan akurasi prediksi. Penulis juga melakukan pemahaman terhadap proses sintesis data dengan menggunakan *Time-series Generative Adversarial Networks* (TimeGAN).

### 3. Pengambilan Data

Penulis mengumpulkan dataset kasus penyebaran malaria tahun 2019-2022. Dataset tersebut akan diambil dari penelitian sebelumnya yang dilakukan oleh Fahmi et al. (2022)

### 4. Perancangan Sistem

Pada tahapan ini, model prediksi penyebaran malaria dan desain antarmuka sistem akan dirancang dan disesuaikan dengan masalah terkait penelitian ini.

### 5. Pengimplementasian Sistem

Penulis akan menerapkan rancangan sistem dan struktur model *training* agar memperoleh hasil seperti yang telah diharapkan. Penulis akan merancang model hibrida MLP-WOA dengan menggunakan bahasa pemrograman *Python* beserta *library* seperti *Tensorflow*, *pandas*, *numpy*, dan lain-lain.

## 6. Pengujian

Pada tahapan ini, model prediksi yang sudah diimplementasikan akan diuji dengan *data testing* sehingga dapat diuji kualitas dari model prediksi *Multilayer Perceptron-Whale Optimization Algorithm*.

## 7. Penulisan Laporan

Setelah sistem dibangun dan diuji, penulis akan menyusun laporan penelitian. Penulis akan menjelaskan seluruh tahapan dari awal hingga ketika model prediksi dapat diaplikasikan seperti yang telah diharapkan.

### 1.7. Sistematika Penulisan

Dalam penulisannya, skripsi ini mengandung lima bagian, yakni:

#### Bab 1: Pendahuluan

Bagian pendahuluan berisi penjelasan mengenai hal yang mendasari alasan dilakukannya penelitian ini. Bagian pendahuluan terdiri dari penjelasan yang dibagi menjadi beberapa sub bab, yaitu latar belakang permasalahan, rumusan masalah, tujuan dilakukannya penelitian, batasan masalah, manfaat dilakukannya penelitian, metodologi penelitian, dan sistematika penulisan.

#### Bab 2: Landasan Teori

Bagian landasan teori berisi penjelasan mengenai teori-teori yang mendukung permasalahan atas penelitian yang dilakukan. Teori-teori tersebut meliputi malaria, jenis malaria, hubungan faktor meteorologi dengan kasus malaria, *Generative Adversarial Networks* (GAN), *Multilayer Perceptron* (MLP), dan *Whale Optimization Algorithm* (WOA). Selain itu, bagian ini juga berisi beberapa penelitian terdahulu yang terkait serta perbedaan penelitian ini dengan penelitian terdahulu.

#### Bab 3: Analisis dan Perancangan Sistem

Bagian ini berisi penjelasan mengenai dataset yang digunakan, arsitektur umum, proses *data preprocessing*, *model training*, *model evaluation*, *model output*, dan rancangan *interface* yang akan dibangun.

#### Bab 4: Implementasi dan Pengujian Sistem

Bagian implementasi dan pengujian sistem berisi penjelasan mengenai proses penerapan arsitektur dan rancangan antarmuka sistem atau *interface*. Selain itu, bagian

ini juga memuat spesifikasi *hardware* dan *software* yang digunakan serta proses evaluasinya.

## Bab 5: Kesimpulan dan Saran

Bagian ini berisi kesimpulan yang ditarik berdasarkan hasil penelitian dan saran yang dapat diberikan untuk dijadikan referensi penelitian selanjutnya



## **BAB 2**

### **LANDASAN TEORI**

#### **2.1. Malaria**

Penyakit malaria sering terjadi terutama di negara tropis, seperti Indonesia. Infeksi ini disebabkan oleh parasit *Plasmodium* sehingga tidak akan menyebar antara sesama manusia, melainkan melalui perantara nyamuk *Anopheles*, transfusi darah, hingga jarum suntik. Bayi, balita, wanita hamil, wisatawan, dan penderita HIV atau AIDS memiliki resiko lebih tinggi terkena infeksi parah dari malaria, seperti kejang, kesulitan bernapas, hingga kematian jika tidak segera diobati. Adapun 5 jenis parasit *Plasmodium* penyebar penyakit malaria, terdiri dari *P. vivax*, *P. falciparum*, *P. ovale*, *P. knowlesi*, dan *P. malariae*. *P. falciparum* sebagai parasit mematikan dan banyak ditemukan di benua Afrika, sedangkan Parasit yang paling dominan di negara lainnya adalah *P. vivax* (World Health Organization, 2023). Pada tahun 2022, WHO telah melaporkan adanya 249 juta kasus malaria secara global dengan 1,2 juta kasus diantaranya terjadi di Indonesia (World Health Organization, 2023). Beberapa pencegahan yang dapat dilakukan diantaranya adalah kontrol vektor dengan insektisida atau penyemprotan residu, vaksin, kemoterapi pencegahan, dan lain-lain (World Health Organization, 2023).

Nyamuk *Anopheles* betina dapat menyebarkan malaria karena nyamuk betina perlu menggigit manusia untuk perkembangan telurnya. Telur nyamuk berkisar antara 100-300 butir dengan umur rata-rata 25,6 hari. Namun, hanya nyamuk *Anopheles* tertentu memiliki peran sebagai vektor malaria. Nyamuk vektor memiliki ciri-ciri kontak tinggi dengan manusia, umur lebih panjang (sekurang-kurangnya 9-16 hari untuk mendukung perkembangan sporozoit), spesies yang lebih dominan, dan telah dikonfirmasi sebagai vektor di lokasi lain. Indonesia sendiri memiliki 80 spesies *Anopheles* dengan 20 di antaranya telah dilaporkan dapat menularkan parasit *Plasmodium* dan berpotensi sebagai vektor malaria (Amrul Munif, 2009).

#### **2.2. Jenis Malaria**

Terdapat empat macam parasit *Plasmodium* penyebab penyakit malaria. Spesies *Plasmodium* yang berbeda akan memiliki cara penanganan dan gejala yang berbeda pula.



a. *Plasmodium falciparum*

*Plasmodium falciparum* adalah parasit yang dapat menyebabkan malaria tropikana. Malaria jenis ini merupakan jenis malaria ganas dengan gejala berat. *Plasmodium falciparum* umumnya muncul di daerah tropis dan sub tropis dan menjadi malaria yang paling membunuh bagi manusia yang tinggal di area tersebut. Parasit ini paling banyak ditemukan di benua Afrika (Biologijk, 2017). Terdapat lebih dari 400 spesies nyamuk tersebut dengan 67 di antaranya terbukti menjadi vektor malaria. Sarang nyamuk ini bisa berada di genangan air atau cabang pohon besar sehingga area dataran rendah lebih rentan terhadap jenis parasit *Plasmodium falciparum*. Gejala-gejala yang ditimbulkan oleh spesies ini, yakni anemia, naik turunnya suhu yang tidak beraturan, parasitemia, dan splenomegaly (Erlita, 2017).

b. *Plasmodium vivax*

Malaria tertiana disebabkan oleh *Plasmodium Vivax*. Malaria jenis ini paling sering terjadi di negara – negara selain Afrika, salah satunya Indonesia (Erlita, 2017). Indonesia sering terjangkit kasus malaria tertiana dengan hampir 40% disebabkan oleh *Plasmodium vivax* (Biologijk, 2017). Adapun gejala yang ditimbulkan oleh malaria jenis ini cukup unik, yakni timbulnya demam setiap 48 jam sekali dan berlangsung sekitar 1-8 jam. Setelah mereda, demam akan muncul kembali 48 jam selanjutnya. Penderita akan merasakan gejala baru setelah 1 minggu, seperti menggigil, tidak enak badan, dan masuk angin (Erlita, 2017).

c. *Plasmodium malariae*

Malaria quartana disebabkan oleh *Plasmodium malariae*. Malaria jenis ini banyak ditemukan di wilayah Eropa, wilayah Asia Tenggara seperti Indonesia, Malaysia, dan Myanmar, wilayah Asia Selatan, seperti Srilanka dan India, Afrika, dan New Guineea. Parasit jenis ini tergolong sebagai parasit tidak ganas tetapi sangat mudah menyerang manusia (Biologijk, 2017). Berbeda dengan *Plasmodium vivax*, *Plasmodium malariae* akan menyebabkan demam yang muncul setiap 72 jam sekali. Adapaun gejala lain juga dapat ditemukan, seperti berkeringat dingin, demam, sakit kepala, hilang kesadaran, hingga gangguan ginjal ketika mencapai titik kronis (Erlita, 2017).

d. *Plasmodium ovale*

Malaria ovale disebabkan oleh *Plasmodium ovale*. Termasuk penyakit yang sangat langka terjangkit oleh manusia, parasit ini sulit didiagnosa karena kemiripan

gejalanya dengan *P. vivax* sehingga perlu ditemukan terlebih dahulu parasit di dalam darah pasien (Erlita, 2017). Parasit ini dapat ditemukan di Afrika, Vietnam, Filipina, hingga Eropa dan Amerika Serikat. Malaria ovale tergolong ringan karena dapat sembuh dengan sendirinya tanpa melalui pengobatan. Adapun gejala yang ditimbulkan oleh parasit ini, yakni demam 48 jam sekali, sakit kepala, tidak enak badan, dan menggigil (Biologijk, 2017)

### **2.3. Hubungan Faktor Meteorologi dengan Malaria**

Variabel iklim mempengaruhi pertumbuhan dan perkembangan parasit secara signifikan. Proses morfologi nyamuk sangat bergantung pada suhu, lingkungan, dan genangan air. Banjir berpotensi untuk menyebarkan penyakit yang ditimbulkan oleh nyamuk vektor malaria (Zhang, Y., Bi & Hiller, 2010). Selain itu, curah hujan berpotensi untuk mendorong pertumbuhan vegetasi, sehingga meningkatkan jumlah nyamuk yang berkembang biak (Sadoine et al., 2018). Sementara itu, tekanan udara dan radiasi matahari yang tinggi dapat menghambat pertumbuhan dan perkembangan malaria (Nkiruka et al., 2021). Temperatur yang meningkat dapat menyebabkan gangguan pertumbuhan nyamuk sehingga suhu maksimum dapat memberikan efek negatif atau menghambat terjadinya malaria. Secara khusus, proporsi parasit yang dapat bertahan hidup akan menurun pada suhu 32-34°C (Hussien, 2019).

Suhu permukaan tanah siang hari yang sangat tinggi (di atas 29°C) dapat mengakibatkan penurunan kejadian malaria. Secara umum, Indeks vegetasi selisih ternormalisasi (NVDI) mempunyai korelasi yang positif terhadap kejadian malaria jangka pendek, tetapi menunjukkan sebaliknya jika berjangka panjang. Sementara itu, ketinggian dan jarak ke badan air berdampak secara negatif terhadap kejadian malaria (Ssempiira et al., 2018). Oleh karena itu, variabel iklim, seperti temperatur, curah hujan, kelembapan, ketinggian, dan tekanan udara memiliki dampak yang signifikan terhadap perkembangan dan penyebaran parasit malaria sehingga perlu dipertimbangkan ketika membangun model prediksi.

### **2.4. Iklim dan Topografi di Kabupaten Asahan**

Kabupaten Asahan terletak pada wilayah timur Sumatera Utara, tepatnya pada garis 2°30'00" LU – 3°10'00" LU dan 99°01' BT – 100°00' BT (BPS Kabupaten Asahan, 2018). Data dari BPS menunjukkan bahwa penduduk Kabupaten Asahan mencapai 802.563 penduduk (Badan Pusat Statistik, 2024). Kabupaten Asahan mencakup luas

area sebesar 3.732,97 km<sup>2</sup>. Kabupaten ini memiliki 25 Kecamatan serta 204 Desa/Kelurahan (BPS Kabupaten Asahan, 2018).

Kabupaten Asahan berada pada 0-1.000 m di atas permukaan laut. Berdasarkan topografi, Kabupaten ini terletak pada bagian Timur Sumatera Utara yang merupakan dataran rendah dengan kelembapan dan curah hujan yang tinggi (Sianturi, 2023). Kabupaten Asahan beriklim tropis dengan dua musim, yakni musim kemarau (Juni – September), musim hujan (November – Maret), dan musim pancaroba terjadi pada peralihan kedua musim. Tercatat hingga tahun 2016, curah hujan bulanan di Kabupaten Asahan berkisar antara 172,5 – 2.070 mm/bulan (BPS Kabupaten Asahan, 2018).

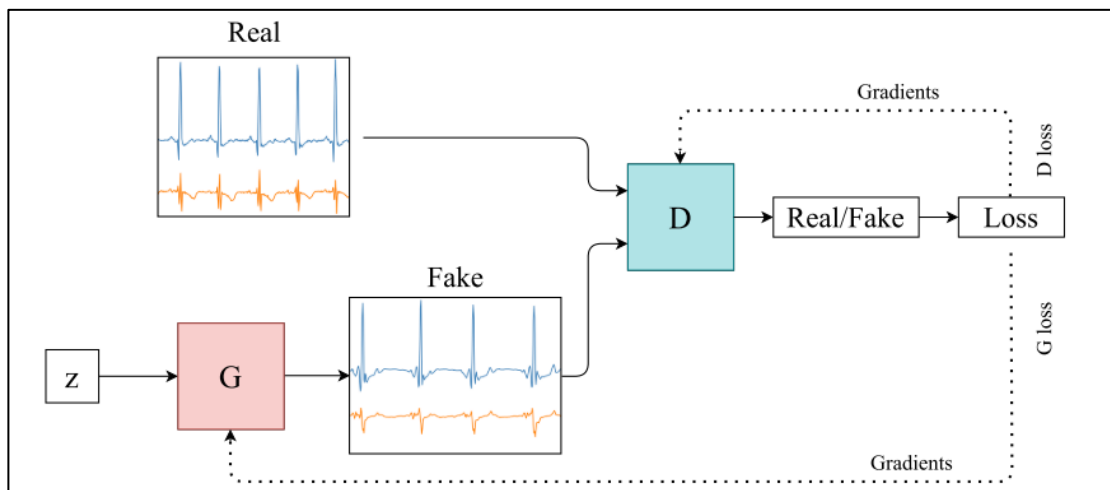
## 2.5. Generative Adversarial Networks

*Generative Adversarial Networks* (GAN) merupakan model generatif yang terdiri dari *generator* dan *discriminator*. GAN digunakan untuk menghasilkan dan memanipulasi data sintesis dengan menggunakan *Neural Network*. GAN telah digunakan untuk menghasilkan data dengan kualitas tinggi pada domain *computer vision* (De Souza et al., 2023), *natural language processing* (Thirumagal & Saruladha, 2021), hingga *time series* (Yu, L. et al., 2017; Wiese, M. et al., 2020). *Generator* dan *discriminator* umumnya menggunakan dua *neural network* yang berbeda. *Generator* digunakan untuk menghasilkan data sintesis yang mirip dengan data *training* dengan mengambil *noise* acak  $z \in \mathbb{R}^r$ . *Discriminator* mencoba untuk menentukan apakah data yang dihasilkan oleh *generator* asli atau palsu. *Generator* akan berusaha meningkatkan tingkat kesalahan *discriminator*, sedangkan *discriminator* akan berusaha untuk meminimalkannya sehingga dapat memperoleh data sintesis yang memiliki tingkat kemiripan yang tinggi dengan data asli, baik dari struktur, distribusi, maupun karakteristiknya (Brophy et al., 2021). Struktur umum dari *Generative Adversarial Networks* seperti yang ditunjukkan oleh gambar 2.1.

GAN umumnya digunakan untuk menghasilkan data non-sekuensial, sehingga pemodelan data *time series* menjadi tantangan bagi GAN karena perlu memperhatikan dan mempertahankan dinamika temporal yang unik pada deret waktu. *Time-series Generative Adversarial Network* (TimeGAN) merupakan GAN yang dapat memodelkan data *time series*. Metode ini memperkenalkan beberapa teknik baru seperti *stepwise supervised loss* menggunakan data asli sebagai *supervision* untuk menangkap distribusi bersyarat bertahap dalam data, *embedding network* yang menyediakan



pemetaan untuk mengurangi dimensi tinggi, dan menggeneralisasi *framework* sehingga dapat digunakan untuk menangani data campuran statis dan deret waktu. Pada dataset yang beragam mulai dari data *continuous* hingga data *discrete*, TimeGAN mampu mengungguli model generative GAN lainnya berdasarkan pengukuran *similarity* dan kemampuan prediktif (Yoon et al., 2019).



Gambar 2.1 *Generative Adversarial Networks* (Brophy, E. et al., 2021)

## 2.6. Multilayer Perceptron

*Multilayer Perceptron* (MLP), sering disebut dengan istilah *Vanilla Neural Network* merupakan jenis *Artificial Neural Network* yang pertama kali dibuat sebagai cabang dari jaringan *feed-forward* (Zhou et al., 2020). *Feed-forward Neural Network* (Svozil et al., 1997) adalah bentuk spesifik dari *Supervised Neural Network* yang mana setiap neuron didistribusikan pada layer yang bertumpuk dan terhubung sepenuhnya dalam satu arah (Basheer & Hajmeer, 2000). MLP dibangun berdasarkan sistem saraf otak manusia dan mempunyai kemampuan yang tinggi dalam memodelkan perilaku non-linear dan sistem yang kompleks (Samadianfard et al., 2020). *Multilayer Perceptron* terdiri dari sekumpulan *layer* dan setiap *layer* terdiri dari sekumpulan neuron yang setiap neuron pada setiap *layer* akan saling berhubungan dalam satu arah. Secara umum, arsitektur MLP mencakup 3 lapisan utama, yaitu *input layer*, *hidden layer*, dan *output layer*. Setiap koneksi neuron antar *layer* akan diberikan bobot untuk penyesuaian model. Selain itu, setiap neuron pada layer selain *input layer* juga memiliki *bias* yang akan ikut dikalkulasi (Alameer et al., 2019).

Data *training* akan dimasukkan pada *input layer* sebagai variabel input dengan tiap neuron melambangkan tiap fitur yang ada. Selanjutnya, setiap *input neuron* akan dihitung bersama dengan bobot yang menghubungkan dengan *hidden layer* dan



*bias* yang ada pada neuron di *hidden layer*. Hasil perhitungan akan disebar pada tiap neuron pada *hidden layer* kemudian akan diaktivasi ( $\varphi$ ) dengan fungsi hiperbolik tangen nonlinier. Selanjutnya, tiap neuron dari *hidden layer* pertama akan kembali dikalkulasi dengan bobot dan *bias* pada neuron di *hidden layer* dan diaktivasi. Proses ini akan berulang untuk setiap *hidden layer* hingga mencapai *output layer* (Alameer et al., 2019). Pada *output layer*, hanya terdiri dari satu neuron saja tanpa adanya fungsi aktivasi (aktivasi linear). Hal ini karena model yang akan dibangun adalah model prediksi regresi. Arsitektur umum model *Multilayer Perceptron* seperti yang ditunjukkan pada gambar 3. Adapun rumus *output* atau hasil perhitungan dari setiap neuron dapat dituliskan sebagai berikut:

$$O_i^{(\ell)} = \varphi(u_i^{(\ell)}) = \varphi\left(\sum_{j=1}^{n_{\ell-1}} O_j^{(\ell-1)} w_{j,i}^{(\ell)} + w_{0,i}^{(\ell)}\right), 1 \leq \ell \leq L$$

Keterangan:

$O_i^{(\ell)}$  = *output* dari neuron  $i$  *layer* ke -  $\ell$

$\ell$  = *layer* selain *input layer*

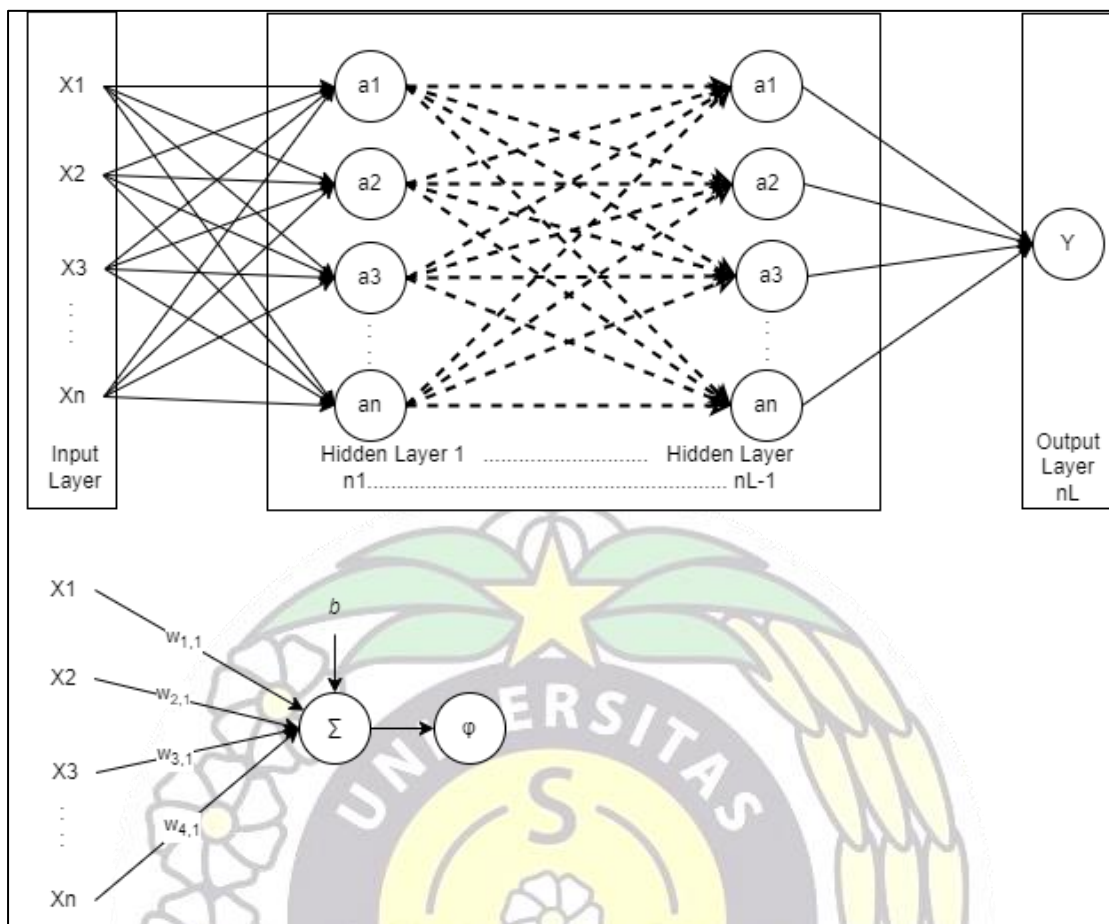
$\varphi$  = fungsi aktivasi

$n_{\ell-1}$  = jumlah neuron pada *layer* sebelumnya

$w_{j,i}^{(\ell)}$  = bobot yang menghubungkan dari neuron  $i$  *layer* ke -  $\ell$  dengan neuron dari *layer* sebelumnya

$w_{0,i}^{(\ell)}$  = *bias* dari neuron  $i$  *layer* ke -  $\ell$

$L$  = jumlah *layer* selain *input layer*



Gambar 2.2 Arsitektur *Multilayer Perceptron*

## 2.7. Whale Optimization Algorithm

*Whale Optimization Algorithm* merupakan algoritma optimasi meta-heuristik yang terinspirasi dari serangan paus *humpback* saat berburu mangsa, yakni *bubble net hunting strategy*. Perilaku paus *humpback* dalam mencari mangsa, mengepung mangsa, dan memakan mangsa disimulasikan untuk mencegah masalah optimasi. Saat berburu mangsa, paus *humpback* akan menyelam sedalam 12 meter kemudian bergerak ke permukaan air sambil membuat gelembung-gelembung air di sepanjang jalur yang berbentuk spiral yang semakin mengecil untuk mengepung dan menyerang mangsa. Dalam algoritma optimasi WOA, jaring gelembung disimulasikan menggunakan gerakan spiral, sedangkan gerakan mengecil disimulasikan dengan menggunakan gerakan melingkar mengecil. Perilaku paus *humpback* diformulasikan ke dalam tiga model matematika WOA, yaitu *encircling prey*, *bubble-net attacking method*, dan *search for prey* (Mirjalili & Lewis, 2016).

*Encircling prey* merupakan gerakan paus *humpback* dalam mencari, mengenali, dan mengitari mangsanya. Dalam menggunakan algoritma WOA, sekumpulan agen

pencari atau populasi paus akan diinisialisasi. Selanjutnya, algoritma WOA akan menghitung fungsi objektif dari tiap agen pencari untuk mencari solusi optimal sementara. Solusi optimal ini dapat berupa fungsi objektif maksimum maupun minimum tergantung dengan model prediksi. Setelah menemukan agen pencari optimal, agen pencari lainnya akan bergerak mendekati agen pencari optimal tersebut layaknya gerakan paus *humpback* dalam mengelilingi mangsa (Mirjalili & Lewis, 2016).

$$\vec{D} = |\vec{C} \cdot \vec{X}^*(t) - \vec{X}(t)| \quad (1)$$

$$\vec{X}(t+1) = \vec{X}^*(t) - \vec{A} \cdot \vec{D} \quad (2)$$

$$\vec{A} = 2\vec{a} \cdot \vec{r} - \vec{a} \quad (3)$$

$$\vec{C} = 2 \cdot \vec{r} \quad (4)$$

Keterangan:

$\vec{D}$  = Jarak antara posisi vektor saat ini dengan posisi vektor solusi terbaik

$\vec{C}$  = Koefisien vektor

$\vec{A}$  = Koefisien vektor dengan nilai *random* dari  $[-a, a]$

$\vec{X}^*(t)$  = posisi vektor terbaik saat berada pada iterasi  $t$

$\vec{X}(t)$  = posisi vektor pada iterasi  $t$

$\vec{X}(t+1)$  = posisi vektor pada iterasi selanjutnya yang akan di-*update* secara spiral

$\vec{a}$  = nilai yang berkurang secara linear dari 2 sampai 0 pada setiap iterasi

$\vec{r}$  = vektor *random* dari  $[0,1]$

$t$  = iterasi

*Bubble-net attacking method* memodelkan perilaku paus *humpback* dalam membuat gelembung untuk mengepung mangsa. Secara matematis, metode ini dapat dirumuskan dengan dua pendekatan, yaitu *shrinking encircling mechanism* dan *spiral updating position*. *Shrinking encircling mechanism* merumuskan perilaku paus *humpback* dalam membentuk lingkaran yang semakin mengecil untuk mengepung mangsa. Koefisien vektor  $\vec{a}$  akan berkurang dari 2 sampai 0 selama iterasi yang menunjukkan posisi vektor agen pencari yang akan diperbarui menuju posisi baru antara vektor terbaik dengan vektor saat ini secara linear mengecil. *Spiral updating position* menghitung jarak antara vektor terbaik dan vektor saat ini, kemudian vektor agen pencari akan diperbarui posisinya secara spiral mendekati vektor solusi optimal (Mirjalili & Lewis, 2016). Kedua model matematis ini disajikan pada gambar 4 dan 5.

$$\vec{X}(t+1) = \begin{cases} \vec{X}^*(t) - \vec{A} \cdot \vec{D} & \text{if } p < 0.5 \\ \vec{D}' \cdot e^{bl} \cdot \cos(2\pi l) + \vec{X}^*(t) & \text{if } p \geq 0.5 \end{cases} \quad (5)$$

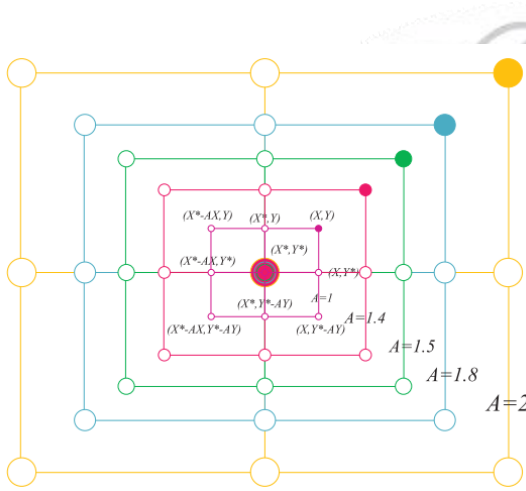
Keterangan:

$\vec{D}'$  = Jarak antara posisi vektor saat ini dengan posisi vektor solusi terbaik ( $\vec{D} = |\vec{X}^*(t) - \vec{X}(t)|$ )

$b$  = konstanta yang mendefinisikan bentuk spiral logaritmik

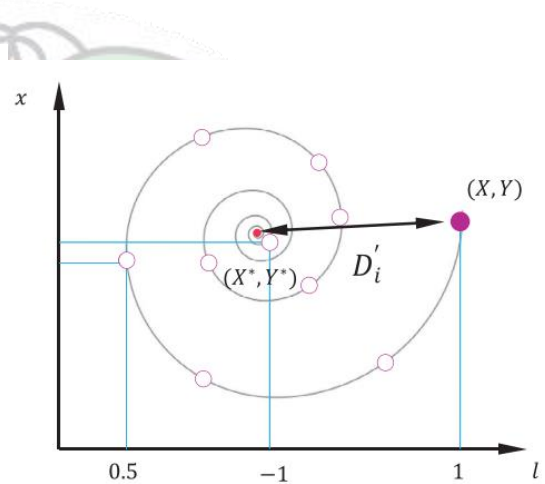
$l$  = bilangan acak dalam  $[-1,1]$

$p$  = bilangan acak dalam  $[0,1]$  yang menentukan pendekatan apa yang akan digunakan



Gambar 2.3 Shrinking Encircling Mechanism

(Mirjalili & Lewis, 2016)



Gambar 2.4 Spiral Updating Position

(Mirjalili & Lewis, 2016)

*Search for prey* memodelkan paus *humpback* dalam mencari mangsanya. Algoritma WOA menggunakan koefisien vektor  $\vec{A}$  untuk mengubah fase agen pencari antara eksplorasi dan eksploitasi, serta nilai acak  $p$  untuk menentukan pendekatan yang dilakukan. Jika  $p$  lebih kecil dari 0.5, agen pencari akan melakukan *shrinking encircling mechanism*, dengan fase bergantung pada koefisien vektor  $\vec{A}$ . Jika nilai mutlak  $\vec{A}$  lebih kecil dari 1, agen pencari akan mendekati posisi vektor solusi terbaik. Sebaliknya, jika nilai mutlak dari koefisien vektor  $\vec{A}$  lebih besar sama dengan 1, agen pencari akan menjauhi vektor solusi terbaik dan mendekati vektor agen pencari lain secara acak. Hal ini dilakukan agar algoritma dapat melakukan eksplorasi secara global sehingga dapat menghindari solusi optimal lokal. Jika  $p$  lebih besar dari 0.5, agen pencari akan melakukan *spiral updating position* (Mirjalili & Lewis, 2016). Algoritma cara kerja *Whale Optimization Algorithm* tercantum secara lengkap pada gambar 2.5.

$$\vec{D} = |\vec{C} \cdot \vec{X}_{rand} - \vec{X}| \quad (7)$$



$$\vec{X}(t + 1) = \overrightarrow{X_{rand}} - \vec{A} \cdot \vec{D} \quad (8)$$

Keterangan:

$\vec{D}$  = Jarak antara posisi vektor saat ini dengan posisi vektor acak

$\overrightarrow{X_{rand}}$  = posisi vektor acak

```

Inisialisasi populasi paus atau agen pencari  $X_i$  ( $i = 1, 2, \dots, n$ )
Hitung fungsi objektif setiap agen pencari
Inisialisasi  $X^*$  dengan agen pencari solusi terbaik
while ( $t < \text{iterasi maksimal}$ )
    for setiap agen pencari
        Perbarui nilai  $a$ ,  $A$ ,  $C$ ,  $l$ , dan  $p$ 
        if ( $p < 0.5$ ) then
            if ( $|A| < 1$ ) then
                Perbarui posisi agen pencari dengan rumus 2
            else if ( $|A| \geq 1$ ) then
                Pilih sebuah agen pencari secara acak
                Perbarui posisi agen pencari dengan rumus 8
            end if
        else if ( $p \geq 0.5$ ) then
            Perbarui posisi agen pencari dengan rumus 6
        end if
    end for
    Periksa apakah ada agen pencari yang melewati ruang pencarian dan ubahlah
    Hitung fungsi objektif dari setiap agen pencari
    Perbarui  $X^*$  jika ada agen pencari dengan solusi yang lebih baik
end while

```

Gambar 2.5 Pseudocode Algoritma WOA (Mirjalili & Lewis, 2016)

## 2.8. Penelitian Terdahulu

Penelitian – penelitian terdahulu telah memanfaatkan data iklim dalam memprediksi kasus malaria, baik angka penyebaran maupun potensi penyebaran. Pada tahun 2021, terdapat penelitian untuk memprediksi kasus malaria di kawasan Afrika Sub-Sahara dengan menggunakan data klinis dari tahun 1990 - 2017 yang diambil dari

6 negara yang terpilih (Burkina, Faso, Mali, Republik Niger, Nigeria, Kamerun, dan Republik Demokratik Kongo) dari dataset yang disediakan oleh WHO. Data iklim tahunan diambil dari *National Centre for Atmospheric Research* (NCAR) meliputi curah hujan, suhu, kelembapan, tekanan, dan radiasi permukaan dengan angka penyebaran malaria sebagai target prediksi. Angka penyebaran malaria ini kemudian dinormalisasi dan dihitung *threshold*-nya, sehingga prediksi berupa klasifikasi biner. Jika hasil prediksi berada di atas *threshold*, maka negara tersebut memiliki potensi penyebaran malaria yang tinggi, begitupun sebaliknya. Peneliti melakukan feature engineering untuk mengidentifikasi faktor iklim yang mempengaruhi kejadian malaria dan diperoleh bahwa curah hujan, suhu, dan radiasi permukaan berpengaruh secara signifikan. Selanjutnya peneliti menggunakan *K-means Clustering* untuk mendeteksi *outlier* dan algoritma XGBoost untuk melakukan klasifikasi yang diperoleh akurasi maksimal mencapai 98% (Nkiruka et al., 2021).

Pada tahun 2022, peneliti melakukan penelitian untuk membangun kerangka kerja prediktif untuk memprediksi kasus malaria di Burundi pada empat sub-kelompok, yaitu anak-anak dibawah 11 bulan, anak-anak di antara 12-59 bulan, wanita hamil di bawah 3 bulan, dan wanita hamil di atas 3 bulan. Selain itu, data iklim seperti suhu, kelembapan, dan curah hujan serta data geografis seperti wilayah, garis bujur, garis lintang, ketinggian, jumlah penduduk, dan pendidikan juga digunakan untuk meningkatkan akurasi prediksi. Data lainnya yang digunakan seperti ITN yang diberikan pada anak, ITN yang diberikan pada ibu hamil, kasus malaria berat, dan kasus malaria lainnya. Peneliti membandingkan dua algoritma, yaitu *Generalized Linear Model* (GLM) dan *Artificial Neural Network* (ANN) untuk melakukan prediksi dan diperoleh bahwa tingkat pendidikan dan penggunaan ITN berpengaruh terhadap penurunan kasus dan beberapa variabel lain berpengaruh terhadap peningkatan kasus, seperti curah hujan, kelembapan, jumlah wanita hamil, dan jumlah anak di bawah lima tahun. ANN memiliki performa yang lebih baik secara umum dibandingkan dengan GLM (Mfisimana et al., 2022).

Pada tahun 2020, dilakukan penelitian untuk memprediksi kejadian malaria di empat lokasi di Telangana, India (Aswaraopeta, Khammam, Aswapuram, dan Venkatapuram) dengan menggunakan *Big Data* dan algoritma *Long Short-Term Memory* (LSTM). Peneliti menggunakan data klinis *Rural Health Centers* (RHC) dari tahun 1995 – 2018 dari keempat lokasi geografis tersebut. Peneliti juga menggunakan

data lingkungan berupa curah hujan, temperatur, kelembapan relatif, dan indeks vegetatif dari satelit MODIS dan TRMM. Analisis menunjukkan bahwa variabel lingkungan memainkan peran penting dalam meningkatkan akurasi prediksi. Model LSTM menunjukkan akurasi 96,11% pada lokasi Venkatapuram, 94,42% pada lokasi Aswapuram, 88,72% pada lokasi Kammam, dan 93,26% pada lokasi Aswaraopeta (Santosh et al., 2020).

Pada tahun 2021, terdapat sebuah penelitian untuk membandingkan kinerja enam model prediksi yang berbeda, yaitu *Support Vector Machine*, *Random Forest*, *K-Nearest Neighbor*, *Logistic Regression*, *Decision Tree*, dan *Naïve Bayes* untuk memprediksi kasus positif dan negatif malaria. Peneliti menggunakan data kejadian malaria (kasus malaria setiap tahun, populasi beresiko per 1000) yang diambil dari WHO serta data iklim seperti atmosfer, curah hujan, dan suhu dari portal iklim dunia. Peneliti memprediksi kasus malaria dengan algoritma klasifikasi biner, yakni 1 untuk kasus positif dan 0 untuk kasus negatif. Berdasarkan hasil penelitian, *Random Forest* memiliki akurasi yang terbaik yakni sebesar 97.72% (Adamu & Singh, 2021).

Pada tahun 2019, dilakukan penelitian untuk membandingkan kinerja model prediksi *traditional single time series* dan metode *stacking* dalam memprediksi angka penyebaran malaria di provinsi Yunnan. Peneliti menggunakan data klinis bulanan dari provinsi Yunnan sejak tahun 2011 – 2017 yang diambil dari *National Scientific Data Sharing Platform for Population and Health* dan data meteorologi seperti suhu, kelembapan relatif, tekanan udara, tekanan uap, tingkat kelembapan, kecepatan angin, curah hujan, dan durasi sinar matahari yang diambil dari *National Meteorological Information Center*. Peneliti menggunakan model jaringan ARIMA, STL+ARIMA, BP-ANN, dan LSTM secara terpisah dan kemudian menggunakan metode *stacking* untuk menggabungkan beberapa model tersebut. Hasil RMSE dari empat sub-model secara terpisah adalah 13,176, 14,543, 9,571 dan 7,208; hasil MASE adalah 0,469, 0,472, 0,296 dan 0,266, serta hasil MAD adalah 6,403, 7,658, 5,871, dan 5,691. Setelah menggunakan arsitektur *stacking* yang dikombinasikan dengan empat model di atas, nilai RMSE, MASE dan MAD menurun menjadi 6.810, 0.224 dan 4.625 (Wang et al., 2019).

Pada tahun 2019, peneliti melakukan penelitian untuk menyelidiki hubungan non-linear antara kejadian malaria dan faktor cuaca. Peneliti menggunakan data klinis pasien di Vhembe, Limpopo, Afrika Selatan dari tahun 1998 – 2015 dan data



meteorologi berupa curah hujan dan suhu. Peneliti menggunakan model *Distributed Lag Nonlinear Model* (DLNM) untuk menyelidiki hubungan dan memprediksi kejadian malaria. Walaupun secara umum model ini menunjukkan akurasi yang tinggi, namun terjadi akurasi yang rendah untuk musim 2014-2015 yang terutama disebabkan oleh wabah besar yang tidak terduga. Curah hujan tampaknya tidak berkontribusi besar terhadap wabah besar ini. Di Vhembe, kasus lokal merupakan penyebab utama beban malaria. Namun, ada peningkatan mendadak dalam proporsi kasus malaria impor selama dua musim sebelumnya, 2012-2013 dan 2013-2014 (Kim et al., 2019).

Pada tahun 2022, terdapat penelitian untuk memahami faktor risiko malaria di Sumatera Utara. Peneliti menggunakan data malaria tahun 2019-2020 yang diperoleh dari data dinas kesehatan Indonesia berupa usia, jenis kelamin, bulan kunjungan, pekerjaan, alamat, jenis malaria, jenis pemeriksaan malaria, hasil pemeriksaan dan jenis penularan (lokal/impor). Selain itu, peneliti juga menggunakan data meteorologi yang diperoleh dari BMKG dan diambil dari 10 stasiun monitor cuaca tahun 2019-2020, terdiri dari curah hujan bulanan, tingkat kelembaban relatif, suhu maksimum dan minimum. *Multivariate Logistic Regression* dilakukan untuk memahami faktor risiko malaria impor dan *Zero-inflated Poisson Multivariable Regression* digunakan untuk mempelajari faktor penyebab iklim dari malaria lokal. Hasil penelitian menunjukkan bahwa seperlima dari kasus malaria di Sumatera Utara disebabkan oleh malaria impor dengan kota Tanjung Balai dan Kabupaten Asahan sebagai titik panas. Pekerjaan juga mempengaruhi penyebaran malaria, yang mana personil militer, pekerja hutan, dan penambang memiliki resiko lebih tinggi terhadap kasus impor malaria. Selain itu, kelembaban relatif, ketinggian, dan suhu menjadi faktor penting yang mempengaruhi penyebaran malaria (Fahmi et al., 2022).

Pada tahun 2019, dilakukan penelitian untuk memprediksi harga emas dengan menggunakan metode *Multilayer Perceptron-Whale Optimization Algorithm* (MLP-WOA). Peneliti menggunakan data observasi bulanan harga emas dari September 1987 – Agustus 2017, harga perak, tembaga, besi, minyak, nilai tukar, nilai inflasi China dan Amerika Serikat. Peneliti membandingkan metode MLP-WOA dengan algoritma lain seperti *classic Neural Network*, PSO-NN, GA-NN, GWO-NN, dan ARIMA. Hasil penelitian menunjukkan bahwa metode MLP-WOA mengalami peningkatan akurasi prediksi *time series data* dibandingkan dengan metode lainnya (Alameer et al., 2019).



Pada tahun 2020, peneliti melakukan penelitian untuk memprediksi kecepatan angin untuk meningkatkan dan mengendalikan tenaga angin terbarukan di Iran Utara dengan menggunakan metode *Multilayer Perceptron-Whale Optimization Algorithm* (MLP-WOA). Peneliti menggunakan data kecepatan angin yang diperoleh dari 10 stasiun di Iran. Peneliti kemudian membandingkan metode MLP-WOA dengan metode MLP-GA dan MLP. Hasil penelitian menunjukkan bahwa metode MLP-WOA mampu mengurangi nilai RMSE dari 0,570~2,995 menjadi 0,523~2,751, sehingga disimpulkan bahwa algoritma WOA mampu meningkatkan hasil prediksi algoritma MLP (Samadianfard et al., 2020).

Pada tahun 2020, terdapat penelitian untuk memprediksi curah hujan tahunan dengan metode *Multilayer Perceptron-Whale Optimization Algorithm* (MLP-WOA). Peneliti menggunakan 3 variabel input curah hujan yang diambil dari 2 stasiun sinoptik di Sinegal. Peneliti kemudian membandingkan metode hibrid tersebut dengan metode MLP biasa. Hasil penelitian menunjukkan bahwa metode MLP-WOA mampu sedikit meningkatkan nilai akurasi prediksi curah hujan tahunan dibandingkan dengan metode MLP biasa (Diop et al., 2020).

## 2.9. Perbedaan Penelitian

Penelitian-penelitian terdahulu mengenai prediksi penyebaran malaria telah banyak dilakukan dengan keberagaman metode, data, dan variabel input. Output dari model prediksi dapat berupa angka penyebaran atau potensi penyebaran. Nkiruka, Odu et al. (2021), melakukan penelitian dalam memprediksi kasus malaria dengan menggunakan data klinis dari 6 negara di Kawasan Afrika Sub-Sahara dari tahun 1990-2017. Peneliti juga menggunakan variabel iklim berupa curah hujan, tekanan udara, radiasi sinar matahari, suhu, dan kelembapan. Penelitian yang dilakukan bersifat klasifikasi biner yang menentukan apakah suatu negara memiliki potensi penyebaran yang tinggi atau sebaliknya dengan menggunakan metode *XGBoost*. Peneliti juga memanfaatkan *K-means clustering* untuk menghapus *outlier* (Nkiruka et al., 2021).

Perbedaan penelitian yang dilakukan oleh Mfisimana, Lionel Divin et al. (2022) terdapat pada wilayah, metode, variabel *input* dan *output* dari model prediksi. Peneliti melakukan penelitian untuk memprediksi angka penyebaran malaria yang terbagi menjadi 4 sub-kelompok (anak-anak di bawah 11 bulan, anak-anak antara 12-59 bulan, wanita hamil di bawah 3 bulan, dan wanita hamil di atas 3 bulan) dengan menggunakan data total kasus di provinsi-provinsi berbeda di Burundi dari tahun 2010-2017. Peneliti

menggunakan variabel *input* berupa densitas populasi, jumlah wanita hamil, jumlah anak-anak di bawah 5 tahun, jumlah ITN, pendidikan, dan variabel iklim (kelembapan, curah hujan, ketinggian, suhu). Peneliti menggunakan *Generalized Linear Model* dan *Artificial Neural Network* sebagai metode prediksi (Mfisimana et al., 2022).

Perbedaan penelitian yang dilakukan oleh Santosh, Thakur et al. (2020) terletak pada wilayah, metode, dan variabel *input*. Peneliti melakukan penelitian untuk memprediksi angka penyebaran malaria di 4 lokasi di India dengan menggunakan data klinis dari tahun 1995-2018. Variabel iklim yang digunakan peneliti berupa curah hujan, temperatur, kelembapan, dan indeks vegetatif. Peneliti menggunakan *Long Short-Term Memory* sebagai algoritma dalam membangun model prediksi (Santosh et al., 2020).

Peneliti Adamu & Singh (2021) melakukan penelitian untuk memprediksi potensi malaria di Afrika dengan menggunakan data kejadian malaria serta variabel iklim berupa atmosfer, curah hujan, dan suhu. Perbedaan penelitian yang dilakukan peneliti ini terdapat pada wilayah penelitian, metode, dan variabel *output*. Peneliti melakukan penelitian yang bersifat klasifikasi biner yakni ada atau tidaknya wabah di suatu wilayah sehingga tidak secara spesifik menyatakan jumlah angka malaria pada wilayah tersebut. Pada penelitian ini, peneliti menggunakan 6 model *machine learning* dan membandingkan hasil prediksinya (Adamu & Singh, 2021).

Penelitian yang dilakukan oleh Wang, Mengyang et al. (2019) memiliki perbedaan yang terletak pada wilayah penelitian, metode, dan variabel *input*. Peneliti melakukan penelitian untuk memprediksi angka kejadian malaria di provinsi Yunnan dengan menggunakan data klinis bulanan dari tahun 2011-2017. Variabel iklim yang digunakan berupa suhu, kelembapan relatif, tekanan udara, tekanan uap, kecepatan angin, curah hujan, dan durasi sinar matahari. Metode yang digunakan peneliti antara lain ARIMA, SARIMA, BP-ANN dan LSTM yang digunakan sebagai pembanding serta metode *stacking* sebagai metode yang diajukan dalam penelitian tersebut (Wang et al., 2019).

Peneliti Kim, Yoonhee et al. (2019) melakukan penelitian untuk memprediksi angka kejadian malaria di Vhembe, Limpopo, Afrika Selatan dengan menggunakan data klinis dari tahun 1998-2015 serta variabel iklim seperti curah hujan dan suhu. Perbedaan penelitian yang dilakukan oleh peneliti ini terletak pada wilayah penelitian dan metode. Metode yang digunakan peneliti ini adalah *Distributed Lag Nonlinear Model* (DLNM)

untuk menyelidiki hubungan variabel dan memprediksi kasus kejadian malaria (Kim et al., 2019).

Berdasarkan perbandingan penelitian yang telah dipaparkan sebelumnya, penelitian ini memiliki perbedaan dibandingkan dengan penelitian terdahulu dari segi wilayah penelitian, metode yang digunakan, data yang digunakan, dan variabel *input*. Penelitian ini menggunakan data angka kejadian malaria di Kabupaten Asahan yang diambil dari penelitian terdahulu (Fahmi et al., 2022). Data tersebut berisi variabel suhu maksimum, durasi sinar matahari, indeks jumlah hujan, indeks curah hujan, indeks tingkat banjir dan genangan air, indeks kepadatan dan pertambahan penduduk sebagai variabel *input* dan total kasus positif malaria sebagai variabel *output*. Data yang tersedia adalah dataset tahun 2019-2022, sehingga data tersebut akan digabung dengan data sintesis yang dihasilkan menggunakan *Time-series Generative Adversarial Network* (TimeGAN) dari tahun 2015-2018 untuk meningkatkan akurasi model. Metode hibrida *Multilayer Perceptron-Whale Optimization Algorithm* (MLP-WOA) akan digunakan pada penelitian ini.



## BAB 3

### ANALISIS DAN PERANCANGAN SISTEM

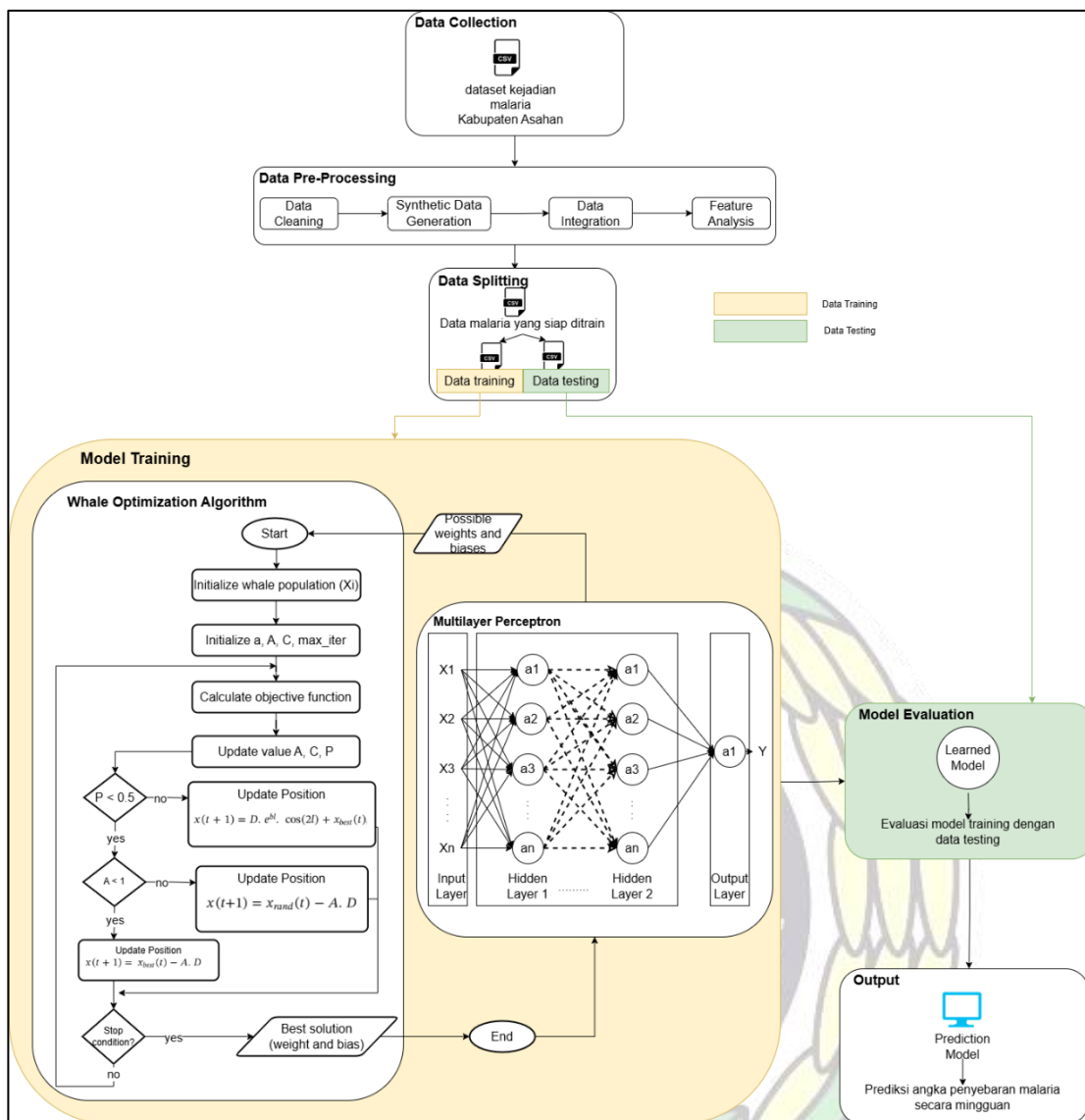
#### 3.1 Data

Pada tahap ini, penulis mengumpulkan data kasus positif malaria yang ada di Kabupaten Asahan. Data ini diperoleh dari data penelitian sebelumnya yang berjudul “Spatial analysis to evaluate risk of malaria in Northern Sumatera, Indonesia” (Fahmi et al., 2022). Adapun data yang diperoleh berisikan variabel minggu, suhu maksimum, durasi sinar matahari, indeks jumlah hujan, indeks curah hujan, indeks tingkat banjir dan genangan air, indeks kepadatan dan pertambahan penduduk dan total kasus positif malaria. Data yang diperoleh adalah data tahun 2019-2022.

#### 3.2 Arsitektur Umum

Penyusunan model prediksi melibatkan beberapa tahapan. Yang pertama adalah pengumpulan data, yakni data iklim dengan kasus positif malaria mingguan. Dataset tersebut akan melalui proses *preprocessing*, yang dimulai dari *data cleaning*, *synthetic data generation*, *data integration*, dan *feature analysis*. Pada tahapan *synthetic data generation*, dataset tahun 2019-2022 akan digunakan untuk memperoleh data sintesis tahun 2015-2018 dengan menggunakan TimeGAN. Setelah melalui tahapan *preprocessing*, dataset yang telah bersih akan di-*split* menjadi *train data* dan *test data*. Selanjutnya, dilakukan proses pembentukan model hibrida MLP-WOA dengan menggunakan *data training*. Model *training* tersebut nantinya akan dievaluasi dengan menggunakan *data testing*. Proses dari setiap tahapan ditunjukkan pada Gambar 3.2.1.





Gambar 3.2.1 Arsitektur Umum

### 3.2.1 Data Collection

Pada tahap ini, penulis mengambil data kasus positif malaria yang ada di Kabupaten Asahan dari tahun 2019-2022 yang diperoleh dari penelitian Fahmi et al. (2022) dengan total 210 data. Penelitian ini akan menggunakan model generatif TimeGAN untuk menghasilkan data sintesis tahun 2015-2018 dengan mempelajari pola/tren data klinis yang telah ada sehingga akan menambah data hingga 415 data.

### 3.2.2 Data Pre-processing

Pada tahapan ini, dataset yang telah diperoleh akan dibersihkan, disintensis, digabungkan, dan dilakukan analisis variabel yang mempengaruhi hasil dari model prediksi. Adapun tahapan dari *data preprocessing* meliputi:

### 3.2.2.1 Data Cleaning

Untuk menjaga kualitas data, data yang telah dikumpulkan harus dibersihkan. *Missing values* serta *outlier* perlu diperbaiki. *Missing values* diisi dengan nilai rata-rata dari variabel tersebut. Sedangkan, penulisan tidak konsisten diperbaiki agar tidak mempengaruhi hasil prediksi. Dataset juga perlu diperiksa ada tidaknya *outlier* yang dapat mempengaruhi penyebaran data. Jika ada, baris tersebut perlu di-drop. *Pseudocode* proses *data cleaning* tercantum pada gambar 3.2.2.

**Function** *clean\_data*:

*Inisialisasi dataframe\_bersih sebagai salinan data asli*

**For** setiap kolom dalam *dataframe\_bersih*:

**If** kolom memiliki *missing value* **Then**:

*Ganti missing value dengan nilai rata-rata dari kolom tersebut*

**End If**

**End For**

**For** setiap kolom dalam *dataframe\_bersih*:

*Hitung interquartile\_range*

*Hitung lower\_bound dan upper\_bound*

**If** kolom memiliki *outlier* **Then**:

*Hapus baris tersebut*

**End If**

**End For**

**Return** *dataframe\_bersih*

Gambar 3.2.2 *Pseudocode* proses *data cleaning*

### 3.2.2.2 Synthetic Data Generation

Dataset yang diperoleh dari penelitian sebelumnya (Fahmi et al., 2022) memiliki rentang dari tahun 2019-2022. Untuk memperoleh dataset angka penyebaran malaria dari tahun 2015-2018, diperlukan data buatan atau data sintesis yang dapat mensimulasikan struktur, distribusi, dan karakteristik dari data asli. Metode yang akan digunakan adalah *Time-series GAN* (Yoon et al., 2019) karena mampu mempertahankan dinamika temporal deret waktu. *Library ydata\_synthetic* akan digunakan untuk mengimplementasikan *TimeGAN*. Data sintesis yang telah dihasilkan akan dikembalikan ke rentang skala aslinya dengan *inverse transform* karena *library* tersebut

akan secara otomatis menormalisasikan data menggunakan *MinMaxScaler*. *Pseudocode* proses *synthetic data generation* tertera pada gambar 3.2.3.

**Function** *generate\_synthetic\_data*:

*Impor library ydata\_synthetic*

*Drop kolom weeks dari original\_dataset*

*Normalisasi dengan MinMaxScaler dan bagi dataset sesuai seq\_len dengan library*

*Train model TimeGAN dengan dataset*

*Generate data\_sintesis dari model TimeGAN yang sudah ditrain*

*Tambahkan kolom weeks ke data\_sintesis*

**Return** *data\_sintesis*

Gambar 3.2.3 *Pseudocode* proses *synthetic data generation*

### 3.2.2.3 Data Integration

Pada tahapan ini, data asli dan data sintesis yang telah dihasilkan akan digabungkan dan diintegrasikan kedalam satu dataset sehingga dapat digunakan untuk *training* dan *testing*.

### 3.2.2.4 Feature Analysis

*Feature Analysis* merupakan tahapan dimana variabel input pada dataset akan dianalisis korelasinya dengan variabel target. Metode yang digunakan adalah *Pearson Correlation* yang dapat digunakan untuk melihat sejauh mana hubungan linear antar 2 variabel. Nilai *pearson correlation* berkisar antara -1 sampai 1. Nilai 1 menunjukkan bahwa kedua variabel tersebut memiliki hubungan positif sempurna dan nilai -1 berarti variabel menunjukkan hubungan negative sempurna. Nilai 0 berarti tidak ada korelasi antara kedua variabel tersebut.

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2 \sum_{i=1}^n (Y_i - \bar{Y})^2}}$$

Keterangan:

$r$  = pearson correlation antara variabel X dan Y

$n$  = jumlah pasangan data yang diobservasi

$X_i$  dan  $Y_i$  = nilai – nilai dari variabel X dan Y dari data ke-i

$\bar{X}$  dan  $\bar{Y}$  = nilai rata-rata dari variabel X dan Y

Penerapan perhitungan *Pearson Correlation* terhadap tabel 3.2.1 yang merupakan sampel data untuk variabel input indeks tingkat banjir dan variabel output total kasus malaria adalah sebagai berikut:

Tabel 3.2.1 Tabel variabel indeks tingkat banjir dan total kasus

Flood Index	Total Case
6.04	24
7.40	36
7.68	40

$$\bar{x} = \frac{6.04 + 7.40 + 7.68}{3} = 7.04$$

$$\bar{y} = \frac{24 + 36 + 40}{3} = 33.33$$

$$r = \frac{(6.04 - 7.04)(24 - 33.33) + (7.40 - 7.04)(36 - 33.33) + (7.68 - 7.04)(40 - 33.33)}{\sqrt{(6.04 - 7.04)^2 + (7.40 - 7.04)^2 + (7.68 - 7.04)^2} \sqrt{(24 - 33.33)^2 + (36 - 33.33)^2 + (40 - 33.33)^2}}$$

$$\approx 0.997$$

### 3.2.3 Data Splitting

Tahapan selanjutnya adalah membagi data yang telah di-*preprocessed* menjadi *train data* dan *test data*. Penulis membagi data tersebut dengan perbandingan *train data* dan *test data* 9:1, 8:2, dan 7:3 untuk dibandingkan hasil prediksinya. *Train data* dipakai dalam proses pelatihan model prediksi, sedangkan *testing data* akan digunakan untuk mengevaluasi model prediksi.

### 3.2.4 Model Training

Pada tahapan ini, model prediksi akan dibuat dengan menggunakan metode hibrida *Multilayer Perceptron-Whale Optimization Algorithm* (MLP-WOA) dengan menggunakan data *training* yang telah di-*split* sebelumnya. Model *Multilayer Perceptron* menggunakan teknik pembelajaran *back propagation* untuk mengurangi kesalahan dalam model prediksi (Hecht-Nielsen, 1992). *Back propagation* akan menghitung gradien dari *cost function* MLP, seperti MAE dan RMSE. Selanjutnya, teknik ini akan memperbarui bobot dan *bias* dengan suatu metode optimasi yang disebut *gradient-descent* untuk mengurangi fungsi kesalahan hingga mencapai solusi optimal (Zhou et al., 2020). Salah satu kekurangan dari algoritma tradisional adalah optima lokal dan kecepatan konvergensi yang lambat (Alameer et al., 2019). Oleh karena itu, *Whale*



*Optimization Algorithm* akan bertindak sebagai algoritma yang mengoptimalkan hasil prediksi model *Multilayer Perceptron* sebagai pengganti *back propagation*.

*Whale Optimization Algorithm* menginisialisasi semua kemungkinan bobot dan *bias* dari metode MLP sebagai agen pencari. Selanjutnya, fungsi *cost* dari model MLP digunakan sebagai fungsi objektif WOA untuk menemukan bobot dan *bias* yang paling optimal selama proses training, yakni MAE pada penelitian ini. Sebelum iterasi, WOA akan meneruskan solusi vektor bobot dan *bias* saat ini dari sekumpulan agen pencari ke model MLP untuk di-*train*. Vektor solusi dengan nilai MAE paling optimal akan ditetapkan sebagai vektor solusi terbaik. Selanjutnya, pada setiap iterasi, WOA memperbarui posisi vektor agen pencari lainnya untuk mendekati atau menemukan solusi yang lebih baik bergantung pada koefisien acaknya. Pada setiap akhir iterasi, setiap vektor agen pencari akan diteruskan ke MLP untuk dihitung nilai MAE. Jika ditemukan nilai MAE yang lebih optimal, vektor solusi terbaik akan diperbarui (Mirjalili & Lewis, 2016). Proses ini akan diulang hingga mencapai iterasi maksimal dan solusi terbaik akan diteruskan ke model MLP sebagai bobot dan *bias* yang akan digunakan pada tahapan *testing*. Dalam proses perhitungan nilai MAE pada MLP, fungsi aktivasi ReLU pada *hidden layer*. *Pseudocode* penerapan model MLP-WOA pada data kasus positif malaria dapat dilihat pada gambar 3.2.4.

```

inisialisasi jumlah_agen_pencari
inisialisasi jumlah_iterasi
inisialisasi jumlah_layers_dan_nodes
inisialisasi weights dan biases MLP
meratakan array weights_dan_biases menjadi one dimensional
inisialisasi nilai lowerbound (lb), upperbound (ub) untuk setiap weights_dan_biases
for setiap agen in jumlah_agen_pencari
    hasilkan nilai acak untuk setiap weights_dan_biases dalam rentang lb dan ub
    tetapkan weights_dan_biases setiap agen dengan nilai acak
    masukkan weights_dan_biases ke MLP
    hitung nilai MAE pada dataset malaria dan simpan ke array scores
    simpan weights_dan_biases agen pada array positions
end for
inisialisasi nilai MAE dari array scores yang paling optimal sebagai prey_score
inisialisasi weights_dan_biases prey_score sebagai prey_position

```

```

while ( $t < \text{jumlah\_iterasi}$ )
    hitung nilai  $a$ 
    for setiap agen in  $\text{jumlah\_agen\_pencari}$ 
        hitung nilai acak  $A$ ,  $C$ , dan  $p$ 
        if ( $p < 0.5$ ) then
            for setiap dimension in  $\text{weights\_dan\_biases}$  agen
                if ( $|A| < 1$ ) then
                    hitung  $D$  untuk jarak antara dimension prey
                    dan agen
                    perbarui dimension agen menuju prey dengan
                     $A$  dan  $D$ 
                else if ( $|A| \geq 1$ ) then
                    pilih sebuah agen secara acak
                    hitung  $D$  untuk jarak antara dimension agen
                    dan agen acak
                    perbarui dimension agen menuju agen acak
                    dengan  $A$  dan  $D$ 
                end if
            end for
        else if ( $p \geq 0.5$ ) then
            hitung nilai acak  $l$ 
            for setiap dimension in  $\text{weights\_dan\_biases}$  agen
                hitung  $D$  untuk jarak antara dimension agen dan prey
                perbarui dimensi agen dengan rumus spiral
                menggunakan  $D$  dan  $l$ 
            end for
        end if
        periksa apakah nilai  $\text{weights\_dan\_biases}$  setiap agen berada dalam
        rentang  $lb$  dan  $ub$ 
        masukkan  $\text{weights\_dan\_biases}$  agen ke MLP
        hitung nilai MAE pada dataset malaria sebagai  $\text{agen\_score}$ 
        if  $\text{agen\_score} < \text{prey\_score}$  then
            perbarui  $\text{prey\_score}$  dan  $\text{prey\_position}$  dengan agen saat ini

```

```

        end if
    end for
end while
perbarui weights_dan_biases MLP dengan prey_position
uji model dengan menggunakan test dataset malaria
tampilkan hasil evaluasi

```

Gambar 3.2.4 Pseudocode model training MLP-WOA

### 3.2.4.1 Evaluation Metrics (Cost Function)

*Cost function* akan digunakan untuk menghitung perbedaan antara nilai prediksi dengan nilai asli. *Cost function* pada MLP akan digunakan oleh WOA sebagai fungsi objektif untuk mencari agen pencari terbaik (*weight* dan *bias* terbaik). *Cost function* yang akan digunakan dalam penelitian ini adalah MAE, MSE, RMSE, dan  $R^2$ .

a. Mean Absolute Error (MAE)

*Mean Absolute Error* adalah metrik evaluasi yang menghitung rata-rata selisih mutlak dari hasil prediksi ( $\hat{y}$ ) dengan nilai sebenarnya ( $y$ ). MAE memberikan bobot yang sama pada setiap kesalahan sehingga tidak akan berdampak terlalu besar pada kesalahan yang besar. MAE tidak terlalu rentan terhadap *outlier* dan memiliki satuan yang sama. Rumus MAE adalah sebagai berikut:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Penerapan perhitungan metrik MAE terhadap tabel 3.2.2 yang merupakan sampel perbandingan nilai aktual dengan nilai prediksi adalah sebagai berikut:

Tabel 3.2.2 Tabel perbandingan nilai aktual dan prediksi untuk MAE

Weeks	Actual Total Case	Predicted Total Case
2021-02-08 00:00:00	17	16
2021-02-15 00:00:00	14	10
2021-02-22 00:00:00	13	13

$$MAE = \frac{|17 - 16| + |14 - 10| + |13 - 13|}{3} \approx 1.67$$

b. Root Mean Squared Error (RMSE)

*Root Mean Squared Error* adalah metrik evaluasi yang digunakan untuk menghitung akar dari rata-rata kuadrat selisih dari hasil prediksi ( $\hat{y}$ ) dengan nilai sebenarnya ( $y$ ). RMSE akan memberikan bobot lebih pada kesalahan yang lebih besar karena adanya kuadrat yang memperkuat dampak kesalahan. RMSE lebih rentan terhadap *outlier* dan memiliki satuan yang sama dengan nilai hasilnya. RMSE dapat dirumuskan sebagai berikut:

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

Penerapan perhitungan metrik RMSE terhadap tabel 3.2.3 yang merupakan sampel perbandingan nilai aktual dengan nilai prediksi adalah sebagai berikut:

Tabel 3.2.3 Tabel perbandingan nilai aktual dan prediksi untuk RMSE

Weeks	Actual Total Case	Predicted Total Case
2021-02-08 00:00:00	17	16
2021-02-15 00:00:00	14	10
2021-02-22 00:00:00	13	13

$$RMSE = \sqrt{\frac{(17 - 16)^2 + (14 - 10)^2 + (13 - 13)^2}{3}} \approx 2.38$$

c. Mean Squared Error (MSE)

*Mean Squared Error* adalah metrik evaluasi yang digunakan untuk menghitung seberapa jauh rata-rata kuadrat selisih antara nilai prediksi ( $\hat{y}$ ) dengan nilai sebenarnya ( $y$ ). MSE dapat memberikan bobot kesalahan yang lebih besar dari RMSE karena merupakan hasil kuadrat dari RMSE. Selain itu, metrik evaluasi ini memiliki komputasi yang lebih efisien dibandingkan dengan RMSE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

Penerapan perhitungan metrik MSE terhadap tabel 3.2.4 yang merupakan sampel perbandingan nilai aktual dengan nilai prediksi adalah sebagai berikut:



Tabel 3.2.4 Tabel perbandingan nilai aktual dan prediksi untuk MSE

Weeks	Actual Total Case	Predicted Total Case
2021-02-08 00:00:00	17	16
2021-02-15 00:00:00	14	10
2021-02-22 00:00:00	13	13

$$MSE = \frac{(17 - 16)^2 + (14 - 10)^2 + (13 - 13)^2}{3} \approx 5.67$$

d. R-squared ( $R^2$ )

*R-squared* merupakan metrics evaluasi yang digunakan untuk mengukur seberapa besar variabilitas dari variabel *output* yang dapat dijelaskan oleh variabel *input*. Nilai dari  $R^2$  berkisar antara 0 sampai 1, dengan 0 berarti model tidak mampu menjelaskan variabilitas dan 1 berarti model sangat cocok dengan data yang ada.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Penerapan perhitungan metrik  $R^2$  terhadap tabel 3.2.5 yang merupakan sampel perbandingan nilai aktual dengan nilai prediksi adalah sebagai berikut:

Tabel 3.2.5 Tabel perbandingan nilai aktual dan prediksi untuk  $R^2$ 

Weeks	Actual Total Case	Predicted Total Case
2021-02-08 00:00:00	17	16
2021-02-15 00:00:00	14	10
2021-02-22 00:00:00	13	13
2022-05-02 00:00:00	38	33
2022-05-09 00:00:00	38	32
2022-05-16 00:00:00	43	39

$$\bar{y}_i = \frac{17+14+13+38+38+43}{6} \approx 27.17$$

$$R^2 = 1 - \frac{(17 - 16)^2 + (14 - 10)^2 + (13 - 13)^2 + (38 - 33)^2 + (38 - 32)^2 + (43 - 39)^2}{(17 - 27.17)^2 + (14 - 27.17)^2 + (13 - 27.17)^2 + (38 - 27.17)^2 + (38 - 27.17)^2 + (43 - 27.17)^2} \approx 0.9024$$

### 3.2.5 Model Evaluation

Pada tahapan ini, model *training* yang telah dihasilkan akan diuji dengan menggunakan dataset *testing* sebagai pembandingnya untuk mengevaluasi apakah

model tersebut telah menghasilkan *output* sesuai dengan yang diharapkan. Performa dari model *testing* akan diuji menggunakan *evaluation metrics* yang telah tercantum pada bagian 3.2.4.1.

### 3.2.6 Output

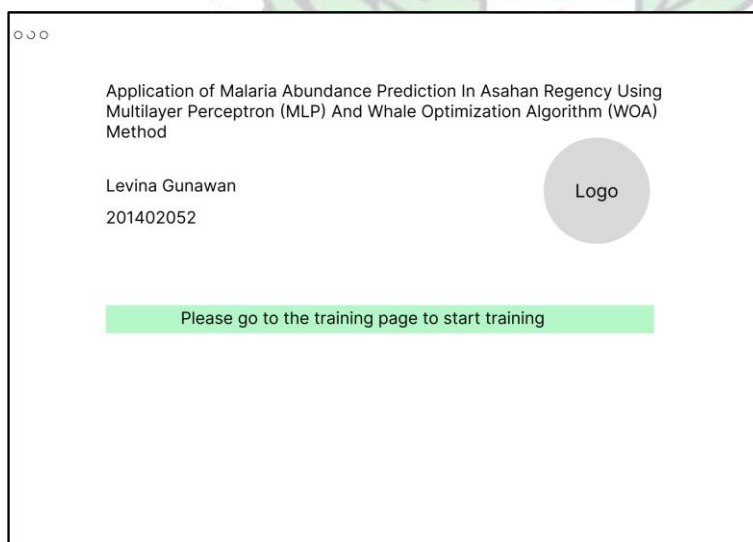
Pada tahapan ini, model prediksi hibrida MLP-WOA dengan *weight* dan *bias* yang paling optimal akan dihasilkan. Hasil prediksi tersebut berupa prediksi angka penyebaran malaria di Kabupaten Asahan secara mingguan. Model tersebut diaplikasikan dalam bentuk sistem *website* sehingga dapat dimanfaatkan di kemudian hari.

## 3.3 Rancangan Antarmuka Sistem

Sebelum mengaplikasikan model menjadi aplikasi *website*, perancangan antarmuka sistem akan dilakukan terlebih dahulu. Pada *website* tersebut, terdapat 3 menu utama pada bagian *sidebar*, yakni menu *training*, *testing*, dan *predicting*. User akan menuju ke menu-menu tersebut secara berurutan untuk melakukan proses *training* hingga *predicting*.

### 1. Rancangan tampilan *home page*

Halaman *home page* sebagai tampilan awal *website* berisi keterangan berupa judul penelitian, nama peneliti, NIM peneliti, dan logo universitas peneliti. Tombol *sidebar* digunakan untuk mengakses halaman *training*, *testing*, dan *predicting*. Tampilan *home page* dapat dilihat pada gambar 3.3.1.



Gambar 3.3.1 Rancangan tampilan *home page*

## 2. Rancangan tampilan *training page*

Ketika user mengakses menu *training* pada bagian *sidebar*, tampilan ini akan muncul. Bagian ini berisi informasi awal yang mirip dengan *home page* serta pemberitahuan untuk meng-*upload* data *training* pada bagian *sidebar*. Pada bagian ini, akan muncul sebuah *form* untuk mengunggah data yang akan di-*train*. Tampilan *training page* dapat dilihat pada gambar 3.3.2.

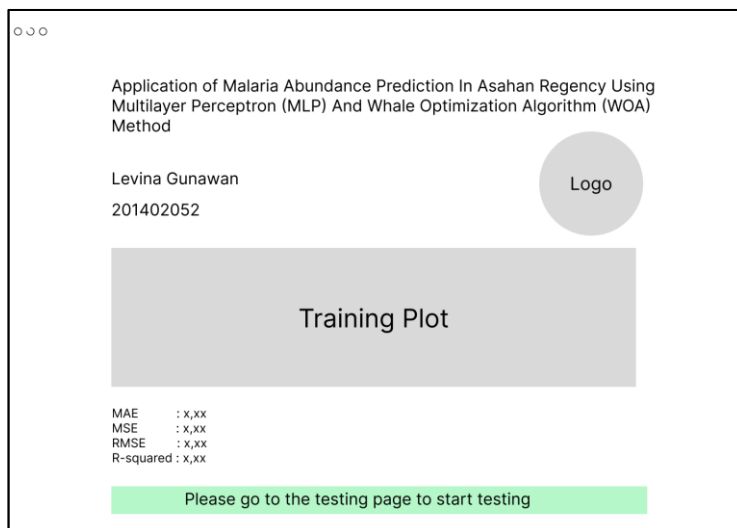
Gambar 3.3.2 Rancangan tampilan *training page*

Setelah mengklik tombol *Start Training*, tampilan utama akan menampilkan yang akan di-*train* serta sebuah tombol untuk melihat hasil *training*. Tampilan *training data* ini dapat dilihat pada gambar 3.3.3.

X1	X2	X3	X4	X5	weeks	Y
Val1	Val2	Val3	Val4	Val5	Val6	Val7
Val8	Val9	Val10	Val11	Val12	Val13	Val14

Gambar 3.3.3 Rancangan tampilan *training data*

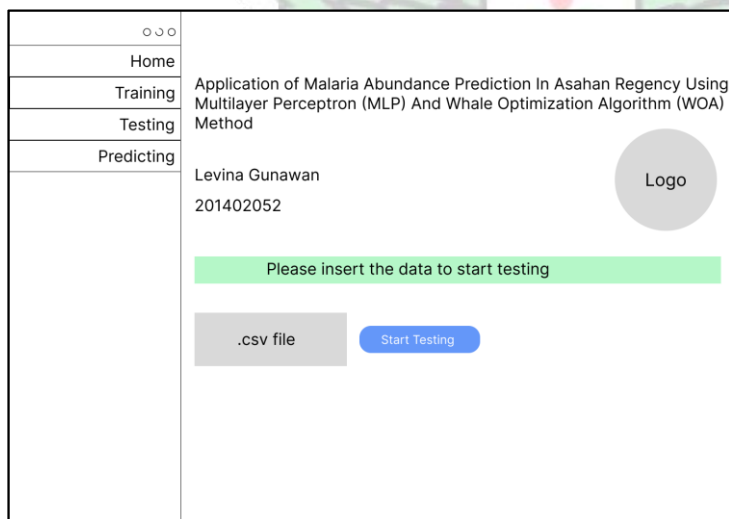
Setelah mengklik tombol *See Result*, pengguna akan dialihkan ke halaman yang berisi plot *training* dan hasil *evaluation metrics* menggunakan MAE, RMSE, MSE, dan  $R^2$ . Sebuah pesan untuk melanjutkan proses ke menu *testing* akan muncul. Tampilan *training metrics* dapat dilihat pada gambar 3.3.4.



Gambar 3.3.4 Rancangan tampilan *training metrics*

### 3. Rancangan tampilan *testing page*

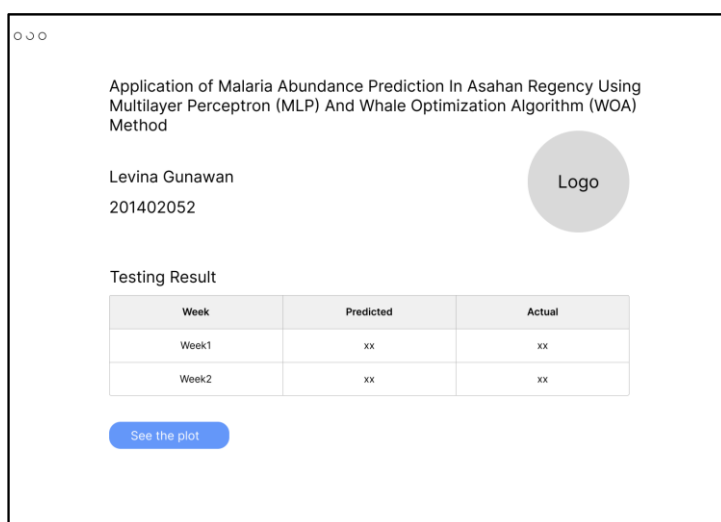
Ketika mengakses menu *testing* pada bagian *sidebar*, halaman ini akan muncul. Tampilan ini juga berisi informasi awal yang mirip dengan *home page* serta pemberitahuan untuk meng-*upload* data *testing*. Pada bagian ini, akan muncul sebuah *form* untuk mengunggah data untuk proses *testing model*. Tampilan *testing page* tertera pada gambar 3.3.5.



Gambar 3.3.5 Rancangan tampilan *testing page*

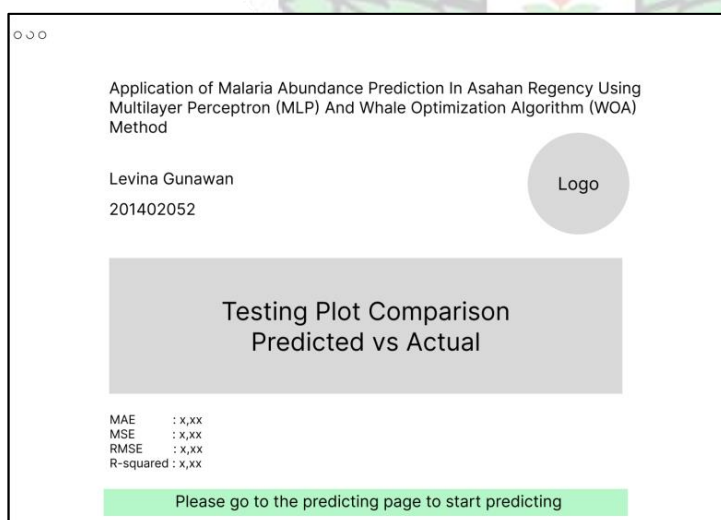


Setelah mengklik tombol *Start Testing*, tampilan utama akan menampilkan tabel perbandingan antara nilai *predicted* dan *actual* dari data *data testing* menggunakan model yang telah di-*train*. Bagian bawah tampilan terdapat sebuah tombol untuk melihat plot perbandingan hasil *testing*. Tampilan *testing comparison* ini ditunjukkan oleh gambar 3.3.6.



Gambar 3.3.6 Rancangan tampilan *testing comparison*

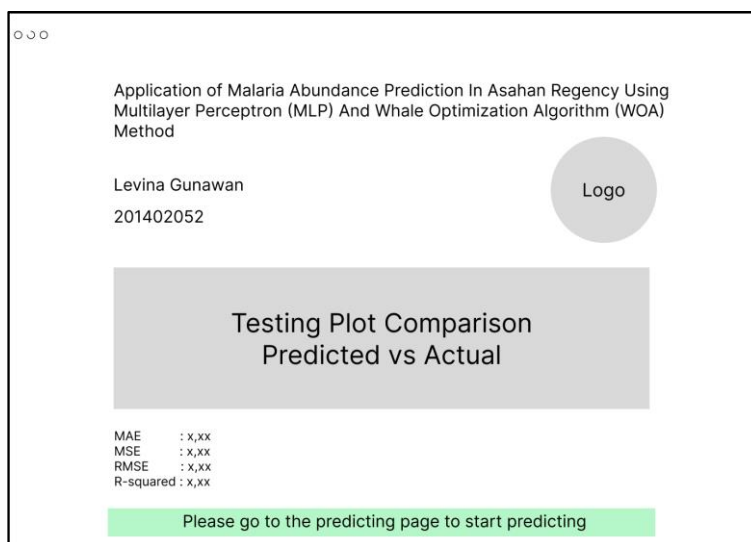
Setelah mengklik tombol *See the plot*, pengguna akan dialihkan ke halaman yang berisi plot perbandingan nilai *predicted* dan *actual* dari proses *testing*. Selanjutnya, hasil *evaluation metrics* menggunakan MAE, RMSE, MSE, dan  $R^2$  juga akan ditampilkan. Sebuah pesan untuk melanjutkan proses ke menu *predicting* akan muncul. Tampilan *testing plot* terlihat pada gambar 3.3.7.



Gambar 3.3.7 Rancangan tampilan *testing plot*

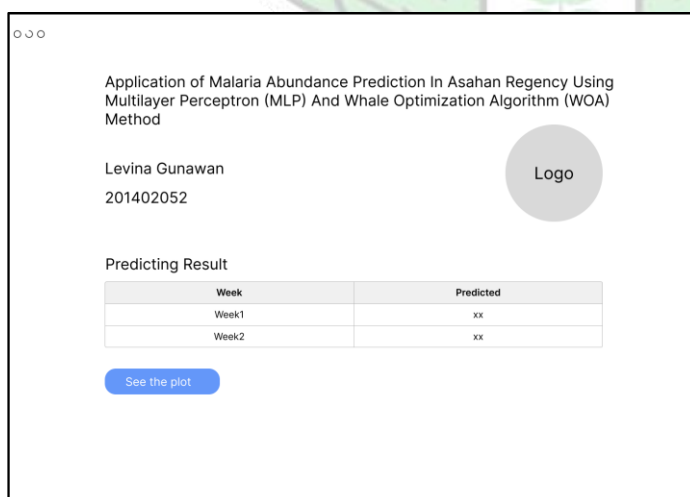
#### 4. Rancangan tampilan *predicting page*

Tampilan ini muncul saat mengakses menu *predicting* pada bagian *sidebar*. Tampilan ini juga berisi informasi awal yang mirip dengan *home page* serta pemberitahuan untuk meng-*upload* data yang akan diprediksi. Selanjutnya, sebuah *form* untuk mengunggah data prediksi akan muncul. Tampilan *predicting page* ditunjukkan pada gambar 3.3.8.



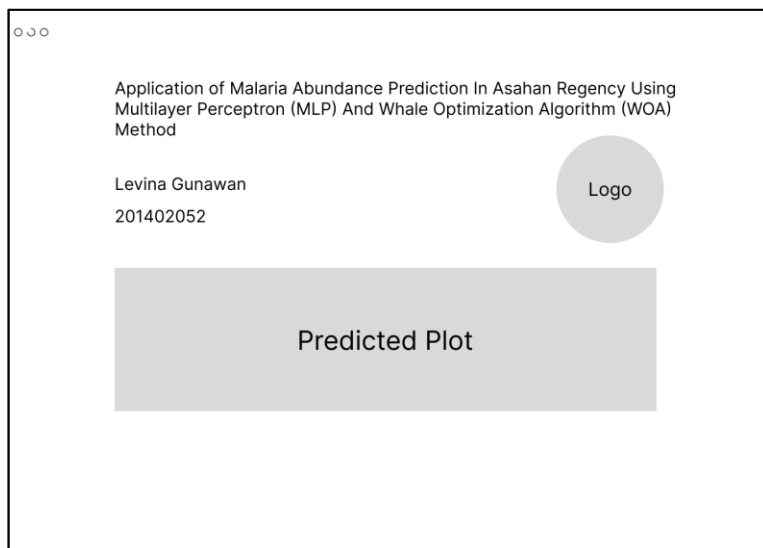
Gambar 3.3.8 Rancangan tampilan *predicting page*

Setelah mengklik tombol *Start Predicting*, model yang telah di-*testing* sebelumnya akan melakukan proses prediksi data. Halaman utama akan menampilkan tabel hasil prediksi total kasus malaria secara mingguan. Bagian bawah tampilan terdapat sebuah tombol untuk melihat plot perbandingan hasil *predicting*. Tampilan *predicting result* terlihat pada gambar 3.3.9.



Gambar 3.3.9 Rancangan tampilan *predicting result*

Setelah mengklik tombol *See the plot*, pengguna akan dialihkan ke halaman yang berisi plot nilai *predicted* dari proses *predicting*. Tampilan *predicting plot* tertera pada gambar 3.3.10.



Gambar 3.3.10 Rancangan tampilan *predicting plot*

## BAB 4

### IMPLEMENTASI DAN PENGUJIAN SISTEM

#### 4.1 Implementasi Sistem

Dalam mengimplementasikan sistem prediksi kasus kejadian malaria di kabupaten Asahan dengan menggunakan metode MLP-WOA, perangkat keras dan perangkat lunak perlu dipertimbangkan.

##### 4.1.1 Spesifikasi Perangkat

Ketentuan *hardware* dalam perancangan sistem prediksi kejadian malaria meliputi:

1. Laptop ASUS VivoBook dengan *processor* Intel Core i7 2.30 GHz
2. 8GB RAM

Persyaratan *software* untuk merancang sistem prediksi kejadian malaria meliputi:

1. Windows 10 64-bit OS
2. Bahasa pemrograman *python* versi 3.10.12 dengan beberapa *library* mencakup *tensorflow* versi 2.17.0, *pandas* versi 2.0.3, *numpy* versi 1.25.2, *matplotlib* versi 3.7.1, *seaborn* versi 0.13.1, dan *ydata\_synthetic*

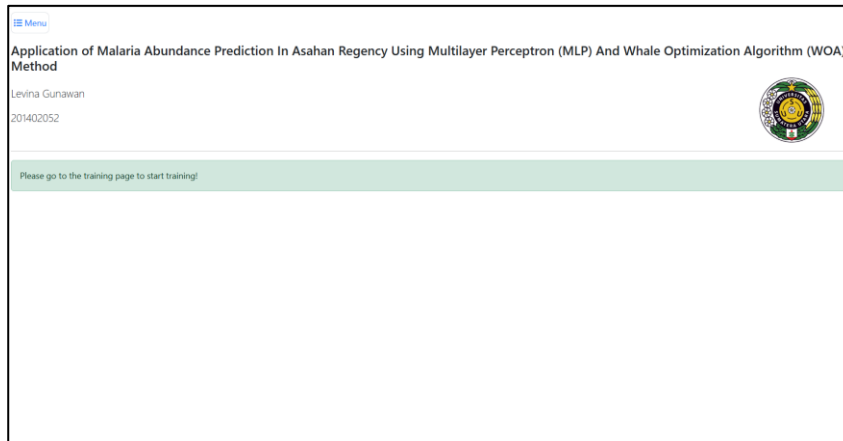
##### 4.1.2 Implementasi Tampilan Antarmuka Sistem

Setelah merancang *interface* sistem, rancangan tersebut akan diimplementasikan menjadi sebuah halaman *website*. Adapun hasil implementasi rancangan *interface* sebelumnya dari Bab 3 yaitu:

1. Tampilan *Home Page*

Tampilan yang muncul pertama kali ketika *user* mengakses *website*. Tampilan ini akan berisi informasi *website* beserta tombol untuk mengakses bagian *sidebar*. Tampilan *Home Page* ditunjukkan oleh gambar 4.1.1.

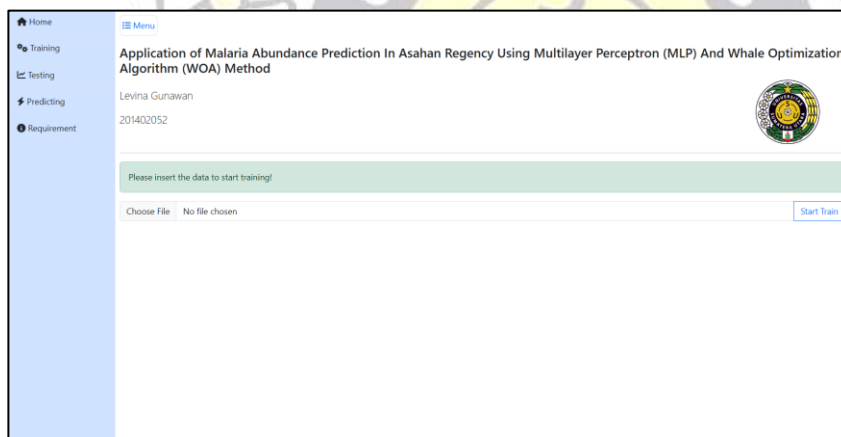




Gambar 4.1.1 Implementasi tampilan *home page*

## 2. Tampilan *Training Page*

Tampilan *training page* merupakan tampilan yang muncul ketika mengklik menu *training* pada bagian *sidebar*. Bagian ini terdapat *form* untuk meng-*upload* data *training* dalam bentuk .csv. Tampilan *training page* dapat dilihat pada gambar 4.1.2.



Gambar 4.1.2 Implementasi tampilan *training page*

Setelah mengklik tombol *Start Train*, data *training* yang di-*upload* akan ditampilkan. Terdapat sebuah tombol *See Result* untuk melihat evaluasi metrik dari *training data*, yakni MAE, MSE, RMSE, dan R-squared. Tampilan *training data* dan *training metrics* dapat dilihat pada gambar 4.1.3 dan 4.1.4.

Menu

Application of Malaria Abundance Prediction In Asahan Regency Using Multilayer Perceptron (MLP) And Whale Optimization Algorithm (WOA) Method

Levina Gunawan  
201402052

Training Data

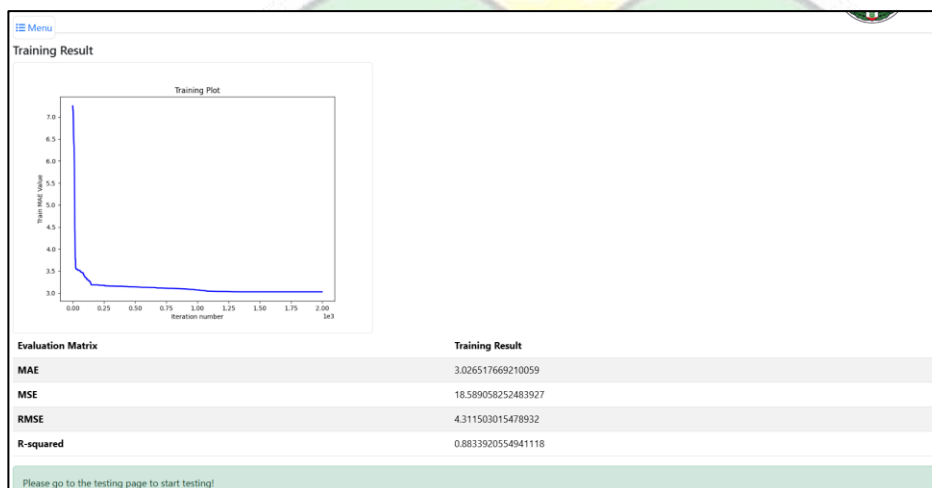
10 entries per page

Search:

#	TR_max	DR_M	CR	CH	FJ	PL_Den	total_case	weeks
1	2.99958635	5.893878904	0.282123526	2.441845672	2.823130592	3.038376025	2	2015-01-05 00:00:00
2	2.94242939	5.415418063	0.405121298	2.179284346	2.757094076	2.88528849	2	2015-01-12 00:00:00
3	3.023463357	6.091875319	0.461759948	2.510543616	2.86325008	2.783962238	2	2015-01-19 00:00:00
4	3.012156038	5.17954401	0.574479305	2.823254384	2.714843368	2.826130907	3	2015-01-26 00:00:00
5	3.014798767	4.473998374	0.514238072	2.01996375	2.820372042	3.000055793	5	2015-02-02 00:00:00
6	2.917496303	2.090989022	0.196117945	1.04349351	3.019078107	3.145554659	1	2015-02-09 00:00:00
7	2.9001349	3.439948927	0.087744479	1.453719428	3.311864281	3.063884084	1	2015-02-16 00:00:00
8	2.950600102	4.013070791	0.189232209	1.814549691	3.189489505	2.946672407	2	2015-02-23 00:00:00
9	2.916770875	1.641686143	0.182891426	1.026292455	2.520758863	3.104450004	1	2015-03-02 00:00:00
10	2.805052718	4.877257496	0.088277347	1.844605812	2.450473645	2.950799923	1	2015-03-09 00:00:00

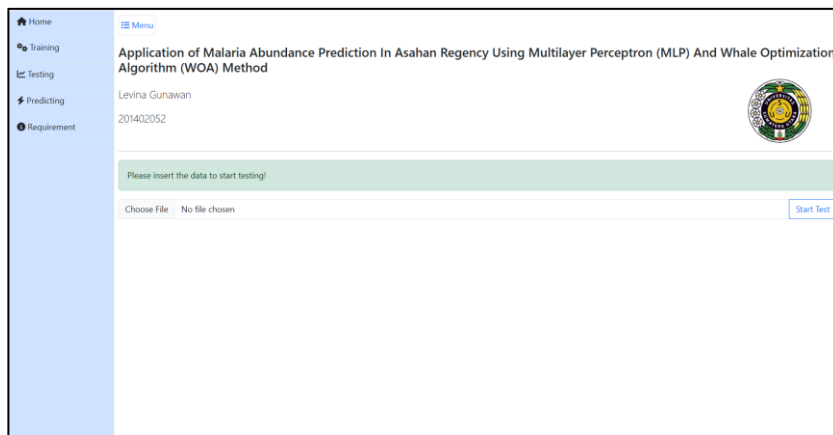
Showing 1 to 10 of 294 entries

See Result

Gambar 4.1.3 Implementasi tampilan *training data*Gambar 4.1.4 Implementasi tampilan *training metrics*

### 3. Tampilan *Testing Page*

Tampilan *testing page* merupakan tampilan yang muncul ketika mengklik menu *testing* pada bagian *sidebar*. Bagian ini terdapat *form* untuk meng-upload data *testing* dalam bentuk .csv. Tampilan *testing page* ditunjukkan pada gambar 4.1.5.

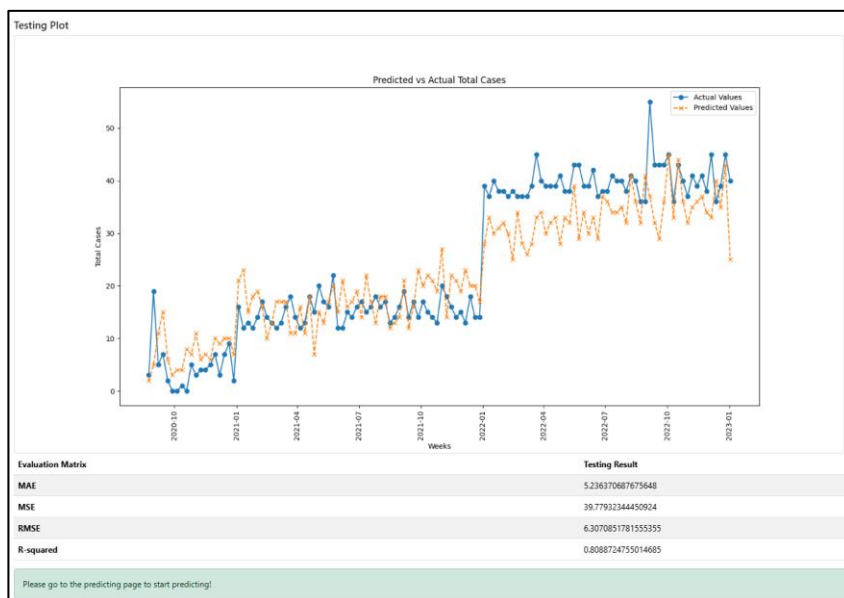


Gambar 4.1.5 Implementasi tampilan *testing page*

Setelah mengklik tombol *Start Test*, perbandingan hasil prediksi dengan nilai aktual akan ditampilkan dalam bentuk tabel. Terdapat sebuah tombol *See Plot* untuk melihat plot perbandingan nilai aktual dan prediksi serta tabel evaluasi metrik dari *testing data*, yakni MAE, MSE, RMSE, dan R-squared. Tampilan *testing result* dan *training plot* terlihat pada gambar 4.1.6 dan 4.1.7.

#	Weeks	Actual	Predicted
1	2020-08-24 00:00:00	3	2
2	2020-08-31 00:00:00	19	5
3	2020-09-07 00:00:00	5	11
4	2020-09-14 00:00:00	7	15
5	2020-09-21 00:00:00	2	6
6	2020-09-28 00:00:00	0	3
7	2020-10-05 00:00:00	0	4
8	2020-10-12 00:00:00	1	4
9	2020-10-19 00:00:00	0	8
10	2020-10-26 00:00:00	5	7

Gambar 4.1.6 Implementasi tampilan *testing comparison*



Gambar 4.1.7 Implementasi tampilan *testing plot*

#### 4. Tampilan *Predicting Page*

Tampilan *predicting page* merupakan tampilan yang muncul ketika mengklik menu *predicting* pada bagian *sidebar*. Bagian ini terdapat *form* untuk meng-*upload* data *predicting* dalam bentuk .csv. Tampilan *predicting page* tertera pada gambar 4.1.8.

Application of Malaria Abundance Prediction In Asahan Regency Using Multilayer Perceptron (MLP) And Whale Optimization Algorithm (WOA) Method

Levina Gunawan

201402052

Please insert the data to start predicting!

Choose File No file chosen Start Predict

Gambar 4.1.8 Implementasi tampilan *predicting page*

Setelah mengklik tombol *Start Predict*, hasil prediksi akan ditampilkan dalam bentuk tabel. Terdapat sebuah tombol *See Plot* untuk melihat hasil prediksi. Tampilan *predicting result* dan *predicting plot* ditunjukkan oleh gambar 4.1.9 dan 4.1.10.



Menu

Application of Malaria Abundance Prediction In Asahan Regency Using Multilayer Perceptron (MLP) And Whale Optimization Algorithm (WOA) Method

Levina Gunawan  
201402052

Predicting Result

10 entries per page

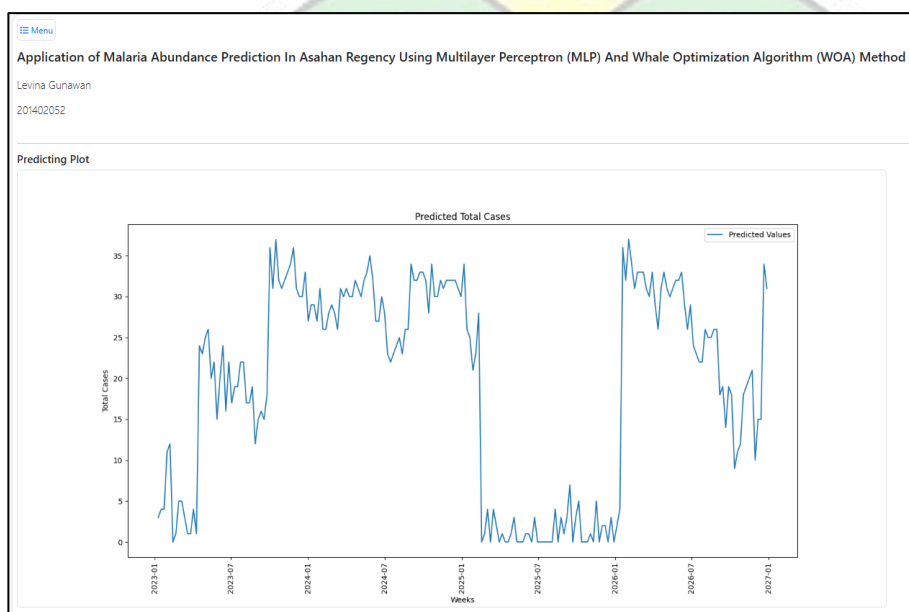
Search:

Weeks	Predicted
2023-01-09 00:00:00	3
2023-01-16 00:00:00	4
2023-01-23 00:00:00	4
2023-01-30 00:00:00	11
2023-02-06 00:00:00	12
2023-02-13 00:00:00	0
2023-02-20 00:00:00	1
2023-02-27 00:00:00	5
2023-03-06 00:00:00	5
2023-03-13 00:00:00	3

Showing 1 to 10 of 208 entries

See Plot

Gambar 4.1.9 Implementasi tampilan *predicting result*



Gambar 4.1.10 Implementasi tampilan *predicting plot*

## 5. Tampilan *Requirement Page*

Tampilan *requirement page* merupakan tampilan yang berisi penjelasan mengenai kolom-kolom pada data yang akan di-*upload*. Bagian ini juga berisi persyaratan data yang akan diunggah pada bagian *training*, *testing*, dan *predicting*. Tampilan *requirement page* terlihat pada gambar 4.1.11.

Home	Menu
Training	
Testing	
Predicting	
Requirement	

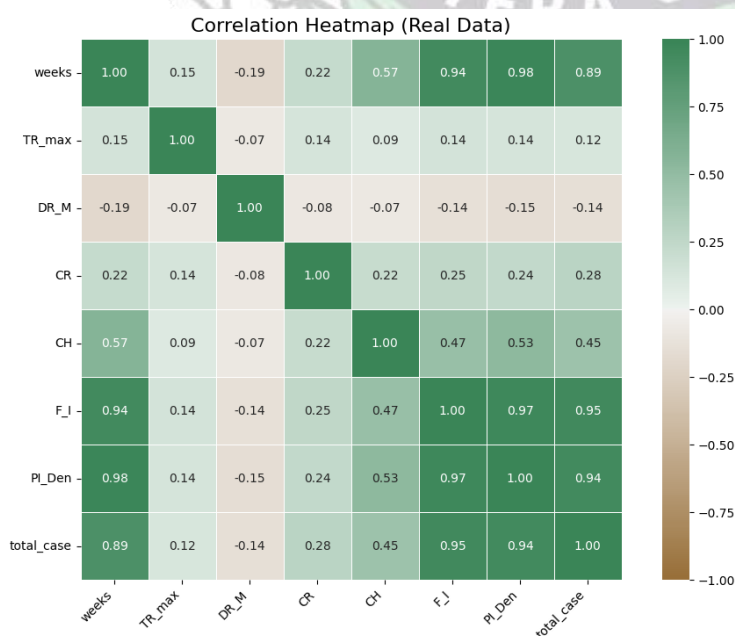
Data Requirement	
When uploading the training, testing, or prediction datasets, please ensure the following requirements are met	
#	Process Requirement
1	General
	Explanation of columns used in malaria predicting process are:
	1. TR_avg - average temperature
	2. DR_M - duration of sunlight
	3. CR - rainfall index
	4. CH - precipitation index
	5. F_I - Flood and waterlogging index
	6. PI_Den - population density and growth index
	7. total_case - total positive malaria cases
	8. weeks - datetime format per weeks
2	Training
	1. While uploading the training dataset, the following columns must be included and in the specified order: [TR_avg, DR_M, CR, CH, F_I, PI_Den, total_case, weeks]
	2. The column headers should match exactly as mentioned above (case-sensitive)
	3. The dataset must not contain any null values
	4. The data should be clean and free from duplicate entries
	5. The dataset must be uploaded in CSV format

Gambar 4.1.11 Implementasi tampilan *requirement page*

## 4.2 Implementasi dan Pengujian Model Data Tahun 2019-2022

### 4.2.1 Dataset Tahun 2019-2022

Dataset asli yang akan di-*train* merupakan data kasus malaria tahun 2019-2022 yang diperoleh dari penelitian yang dilakukan oleh Fahmi et al. pada tahun 2022. Total data yang tersedia adalah sebanyak 210 data. Dataset tersebut berisi 6 variabel input meliputi suhu maksimum (TR\_max), durasi sinar matahari (DR\_M), indeks jumlah hujan (CR), indeks curah hujan (CH), indeks tingkat banjir dan genangan air (F\_I), serta indeks kepadatan dan pertumbuhan penduduk (PI\_Den), dan 1 variabel output yakni total kasus malaria (total\_case). Adapun *Correlation Heatmap* pada data kasus malaria tahun 2019-2022 dapat dilihat pada gambar 4.2.1.



Gambar 4.2.1 *Correlation Heatmap* data kasus malaria 2019-2022

Berdasarkan gambar 4.2.1, indeks tingkat banjir dan genangan air, serta indeks kepadatan dan penambahan penduduk menunjukkan korelasi positif dengan nilai *Pearson Correlation* yang tinggi terhadap total kasus positif malaria, yakni 0.95 dan 0.94 secara berturut-turut. Hal ini menunjukkan adanya keterhubungan yang erat antara variabel-variabel tersebut dengan variabel target sehingga dapat mempengaruhi hasil prediksi secara signifikan. Sementara itu, indeks curah hujan, indeks jumlah hujan, suhu maksimum, dan durasi sinar matahari menunjukkan korelasi yang lebih rendah, yakni 0.45, 0.28, 0.12, dan -0.14 secara berturut-turut.

#### 4.2.2 Implementasi Model

Data yang akan di-*train* pada bagian ini adalah data kasus malaria tahun 2019-2022, yaitu data asli yang belum digabungkan dengan data sintesis. Dalam meningkatkan akurasi, parameter-parameter pada model MLP-WOA perlu dioptimalkan. *Whale Optimization Algorithm* akan bertindak sebagai algoritma optimasi dari *Multilayer Perceptron* untuk menemukan bobot dan bias yang paling optimal. Pada proses ini, peneliti melakukan beberapa eksperimen terhadap parameter-parameter, seperti jumlah iterasi, jumlah agen pencari, jumlah *hidden layer* dan *neuron*, dan *split data*. *Mean Absolute Error* digunakan sebagai metrik evaluasi pada eksperimen ini. Pada *setting* awal, parameter yang ditetapkan adalah 2000 iterasi, 20 agen pencari, dan 2 *hidden layer* masing-masing sebanyak 4 *neuron*. Detail eksperimen *split data* tertera pada tabel 4.2.1.

Tabel 4.2.1 Eksperimen parameter *split data* tahun 2019-2022

Split	Hidden Layer 1	Hidden Layer 2	Iteration	Search Agent	Train MAE	Test MAE	Test R <sup>2</sup>
9:1	4	4	2000	20	6.891	21.153	-23.588
8:2	4	4	2000	20	4.060	18.932	-30.366
7:3	4	4	2000	20	2.586	18.103	-3.363

Berdasarkan hasil eksperimen tabel 4.2.1, dapat dilihat bahwa *split data* 7:3 memperoleh *train* MAE yang paling rendah yaitu sebesar 2.586. Namun eksperimen menunjukkan bahwa perbandingan *split* 8:2 dan *split* 7:3 menghasilkan nilai *test* MAE yang tidak berbeda jauh, yakni sebesar 18.932 dan 18.103 berturut-turut. Oleh karena itu, nilai *test R-squared* ditambahkan untuk menentukan *split* yang paling optimal. Dapat dilihat pada hasil eksperimen bahwa *split* 7:3 memiliki nilai *test R-squared* yang

jauh lebih tinggi, yakni -3.363 dibandingkan dengan *split* 9:1 dan *split* 8:2 yang memperoleh nilai *R-squared* -23.588 dan -30.366 secara berturut-turut. Sehingga, *split* 7:3 menjadi pilihan terbaik yang akan digunakan dalam model karena lebih mampu menunjukkan variabilitas data.

Selanjutnya, eksperimen akan dilanjutkan dengan pemilihan jumlah *hidden layer* yang paling optimal. Model akan diimplementasikan dengan *split data* 7:3, 4 *neuron* pada tiap *hidden layer*, 20 agen pencari, dan 2000 iterasi. Detail eksperimen jumlah *hidden layer* disajikan pada tabel 4.2.2.

Tabel 4.2.2 Eksperimen jumlah *hidden layer* tahun 2019-2022

Split	Num of HL	Iteration	Search Agent	Train MAE	Test MAE
7:3	1	2000	20	2.404	20.455
7:3	2	2000	20	2.586	18.103
7:3	3	2000	20	2.431	13.800

Berdasarkan eksperimen pada tabel 4.2.2, eksperimen ketiga dengan 3 *hidden layer* menunjukkan *test* MAE yang paling rendah, yakni 13.800. Eksperimen pertama menunjukkan *train* MAE yang paling rendah yaitu sebesar 2.404. Oleh karena itu, eksperimen ketiga dengan jumlah *hidden layer* 3 akan digunakan dalam model prediksi karena memiliki akurasi yang paling baik pada *test data* sehingga menunjukkan bahwa kinerja model tersebut adalah yang paling unggul.

Setelah bereksperimen pada jumlah *hidden layer*, langkah selanjutnya adalah eksperimen pada jumlah *neuron* pada *layer* tersebut. Karena jumlah variabel input yang dimasukkan adalah 6 variabel, jumlah *neuron* yang digunakan dalam eksperimen hanya 2 - 5 *neuron*. Model akan diimplementasikan dengan *split data* 7:3, 3 *hidden layer*, 2000 iterasi, dan 20 agen pencari. Detail eksperimen jumlah *neuron* tercantum pada tabel 4.2.3.

Tabel 4.2.3 Eksperimen jumlah *neuron* tahun 2019-2022

Split	Neuron HL 1	Neuron HL 2	Neuron HL 3	Iteration	Search Agent	Train MAE	Test MAE
7:3	4	4	4	2000	20	2.431	13.800
7:3	4	4	2	2000	20	2.512	19.013



Split	Neuron HL 1	Neuron HL 2	Neuron HL 3	Iteration	Search Agent	Train MAE	Test MAE
7:3	4	2	2	2000	20	2.255	18.862
7:3	5	4	4	2000	20	2.420	14.554
7:3	5	4	2	2000	20	2.440	14.956
7:3	5	2	2	2000	20	2.797	13.350

Berdasarkan eksperimen pada tabel 4.2.3, eksperimen keenam menunjukkan bahwa *hidden layer* dengan jumlah *neuron* 5, 2, dan 2 memperoleh nilai *test* MAE yang paling rendah, yakni 13.350 dengan *train* MAE sebesar 2.797. Sementara itu, eksperimen ketiga dengan jumlah *neuron* 4, 2, dan 2 pada masing-masing *hidden layer* menunjukkan nilai *train* MAE yang paling optimal, yaitu 2.255 dengan *test* MAE sebesar 18.862. Oleh karena itu, jumlah *neuron* 5, 2, dan 2 akan digunakan dalam model prediksi karena memiliki kinerja yang lebih baik pada *test data*.

Setelah melakukan eksperimen untuk jumlah *neuron*, langkah selanjutnya adalah melakukan eksperimen untuk parameter jumlah agen pencari. Agen pencari merupakan vektor bobot dan bias dari *Multilayer Perceptron* yang mungkin dan akan di-update posisinya di setiap iterasi untuk meningkatkan akurasi. Model akan diimplementasikan dengan *split data* 7:3, 3 *hidden layer*, 5, 2, dan 2 *neuron* masing-masing *layer*, dan 2000 iterasi, Detail eksperimen jumlah agen pencari disajikan pada tabel 4.2.4.

Tabel 4.2.4 Eksperimen jumlah agen pencari tahun 2019-2022

Split	HL 1	HL 2	HL 3	Iteration	Search Agent	Train MAE	Test MAE
7:3	5	2	2	2000	20	2.797	13.350
7:3	5	2	2	2000	40	1.934	12.417
7:3	5	2	2	2000	60	1.875	12.362
7:3	5	2	2	2000	80	2.270	12.610

Berdasarkan eksperimen pada tabel 4.2.4, eksperimen ketiga dengan jumlah agen pencari sebanyak 60 menunjukkan *train* MAE dan *test* MAE yang paling rendah, yakni 1.875 dan 12.362. Sementara itu, eksperimen keempat dengan 80 agen pencari memperoleh *train* MAE dan *test* MAE sebesar 2.270 dan 12.610. Hal ini menunjukkan bahwa jumlah agen pencari yang semakin banyak tidak membuat akurasi model

menjadi lebih tinggi. Untuk itu, eksperimen ketiga dengan 60 agen pencari akan digunakan dalam model prediksi.

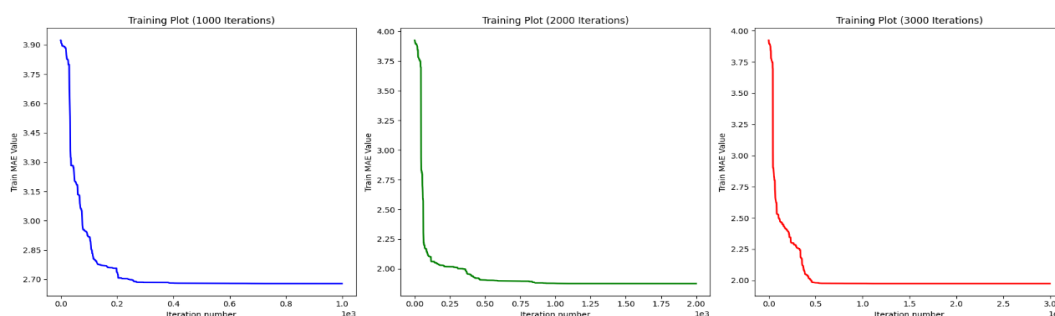
Setelah melakukan eksperimen jumlah agen pencari, langkah selanjutnya adalah melakukan eksperimen parameter jumlah iterasi. Model akan diimplementasikan dengan *split data 7:3*; 3 *hidden layer*; 5, 2, dan 2 *neuron* masing-masing *layer*; dan 60 agen pencari. Detail eksperimen jumlah iterasi dapat tertera pada tabel 4.2.5.

Tabel 4.2.5 Eksperimen jumlah iterasi tahun 2019-2022

Split	HL 1	HL 2	HL 3	Iteration	Search Agent	Train MAE	Test MAE
7:3	5	2	2	1000	60	2.679	17.100
7:3	5	2	2	2000	60	1.876	12.362
7:3	5	2	2	3000	60	1.974	12.990

Berdasarkan eksperimen pada tabel 4.2.5, dapat dilihat bahwa semakin banyaknya iterasi tidak menunjukkan bahwa bobot dan bias akan semakin optimal. Eksperimen kedua dengan 2000 iterasi menunjukkan *train MAE* dan *test MAE* yang paling rendah, yakni 1.876 dan 12.362. Oleh karena itu, pada eksperimen ini, eksperimen kedua dengan 2000 iterasi akan dipilih untuk digunakan dalam model prediksi karena memiliki nilai *train MAE* dan *test MAE* yang paling rendah, sehingga memiliki kinerja yang lebih baik.

Setelah eksperimen terhadap parameter model MLP-WOA telah dilakukan, diperoleh model yang paling optimal dengan parameter-parameter, yaitu *split data 7:3*; 2000 iterasi; 60 agen pencari; 3 *hidden layer*; serta 5, 2, dan 2 *neuron* pada setiap *hidden layer*. Plot hasil optimasi bobot dan bias dengan menggunakan *training data* pada model MLP-WOA dengan parameter tersebut ditunjukkan pada gambar 4.2.2.



Gambar 4.2.2 Plot optimasi *training data* tiap iterasi tahun 2019-2022

#### 4.2.3 Hasil Pengujian Model

Dataset yang digunakan sebagai *test data* adalah data kejadian malaria mingguan dari tanggal 25 Oktober 2021 – 2 Januari 2023. Hasil evaluasi dengan metrik MAE pada data *testing* pada bagian 4.2 yakni sebesar 12.362. Selanjutnya, hasil prediksi ini akan digabungkan dengan total kasus aktual sesuai dengan periode waktunya sehingga dapat dibandingkan dengan mana tercantum pada tabel 4.2.6.

Tabel 4.2.6 Perbandingan nilai aktual dan prediksi pada *test data* 2019-2022

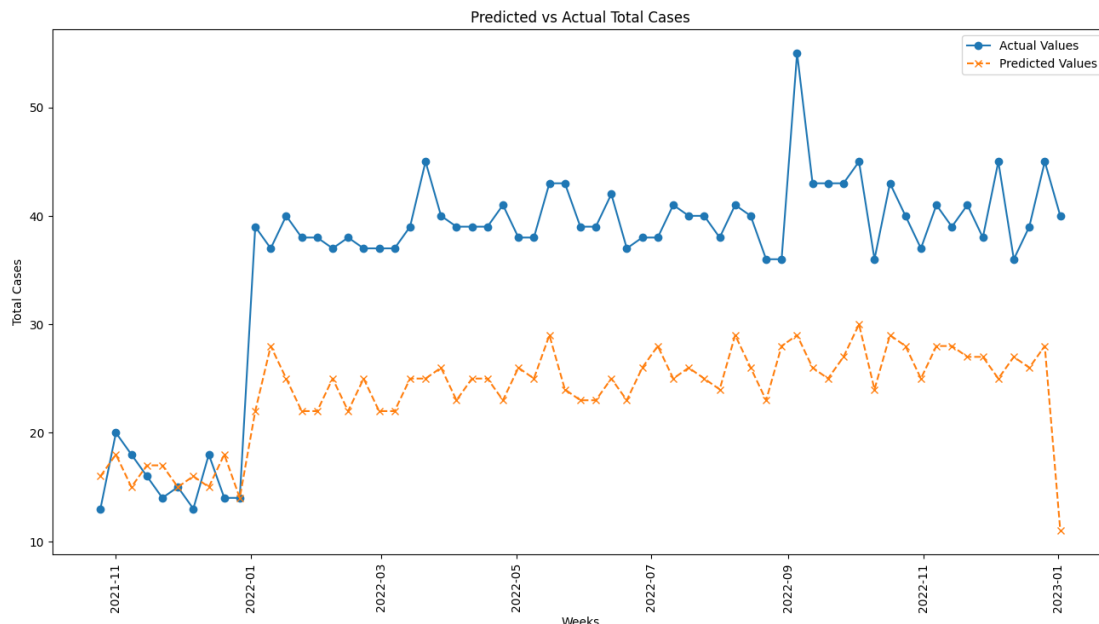
Weeks	Actual	Predicted
2021-10-25 00:00:00	13	16
2021-11-01 00:00:00	20	18
2021-11-08 00:00:00	18	15
2021-11-15 00:00:00	16	17
2021-11-22 00:00:00	14	17
2021-11-29 00:00:00	15	15
2021-12-06 00:00:00	13	16
2021-12-13 00:00:00	18	15
2021-12-20 00:00:00	14	18
2021-12-27 00:00:00	14	14
2022-01-03 00:00:00	39	22
2022-01-10 00:00:00	37	28
2022-01-17 00:00:00	40	25
2022-01-24 00:00:00	38	22
2022-01-31 00:00:00	38	22
2022-02-07 00:00:00	37	25
2022-02-14 00:00:00	38	22
2022-02-21 00:00:00	37	25
2022-02-28 00:00:00	37	22
2022-03-07 00:00:00	37	22
2022-03-14 00:00:00	39	25
2022-03-21 00:00:00	45	25
2022-03-28 00:00:00	40	26

Weeks	Actual	Predicted
2022-04-04 00:00:00	39	23
2022-04-11 00:00:00	39	25
2022-04-18 00:00:00	39	25
2022-04-25 00:00:00	41	23
2022-05-02 00:00:00	38	26
2022-05-09 00:00:00	38	25
2022-05-16 00:00:00	43	29
2022-05-23 00:00:00	43	24
2022-05-30 00:00:00	39	23
2022-06-06 00:00:00	39	23
2022-06-13 00:00:00	42	25
2022-06-20 00:00:00	37	23
2022-06-27 00:00:00	38	26
2022-07-04 00:00:00	38	28
2022-07-11 00:00:00	41	25
2022-07-18 00:00:00	40	26
2022-07-25 00:00:00	40	25
2022-08-01 00:00:00	38	24
2022-08-08 00:00:00	41	29
2022-08-15 00:00:00	40	26
2022-08-22 00:00:00	36	23
2022-08-29 00:00:00	36	28
2022-09-05 00:00:00	55	29
2022-09-12 00:00:00	43	26
2022-09-19 00:00:00	43	25
2022-09-26 00:00:00	43	27
2022-10-03 00:00:00	45	30
2022-10-10 00:00:00	36	24
2022-10-17 00:00:00	43	29
2022-10-24 00:00:00	40	28
2022-10-31 00:00:00	37	25



Weeks	Actual	Predicted
2022-11-07 00:00:00	41	28
2022-11-14 00:00:00	39	28
2022-11-21 00:00:00	41	27
2022-11-28 00:00:00	38	27
2022-12-05 00:00:00	45	25
2022-12-12 00:00:00	36	27
2022-12-19 00:00:00	39	26
2022-12-26 00:00:00	45	28
2023-01-02 00:00:00	40	11

Berdasarkan tabel 4.2.6, nilai aktual dengan jumlah total kasus di bawah 25 menunjukkan hasil prediksi yang cukup memuaskan dengan selisih yang minim. Hal ini ditunjukkan seperti pada 1 November 2021, total kasus malaria mingguan sebanyak 20 kasus dan diprediksi di angka 18. Sementara itu, nilai aktual dengan jumlah total kasus di atas 35 menunjukkan performa hasil prediksi yang kurang memuaskan. Hal ini ditunjukkan seperti pada 17 Januari 2022, total kasus malaria mingguan sebanyak 40 kasus tetapi diprediksi hanya 25 kasus. Performa tersebut disebabkan oleh masih sangat minimnya data yang di-*train* pada model. Perbandingan tersebut kemudian dibuat menjadi plot perbandingan dengan menggunakan *line plot*. Adapun plot perbandingan hasil prediksi dan nilai aktual terlihat pada gambar 4.2.3.



Gambar 4.2.3 Plot perbandingan nilai aktual dan prediksi *test data* 2019-2022

#### 4.2.4 Evaluasi

Setelah melakukan prediksi dengan menggunakan data *training* dan *testing* dengan *Mean Absolute Error* sebagai *cost function*, selanjutnya juga akan dilakukan evaluasi dengan menggunakan metrik evaluasi lainnya, yaitu MSE, RMSE, dan  $R^2$ . Rumus perhitungan setiap metrik evaluasi dapat dilihat pada bagian 3.2.4.1. Evaluasi hasil prediksi tercantum pada tabel 4.2.7. Hasil *test*  $R^2$  pada model *training* ini adalah sebesar -1.045. Hal ini menunjukkan bahwa model ini belum mampu menunjukkan variabilitas dari data.

Tabel 4.2.7 Evaluasi hasil prediksi tahun 2019-2022

Metrics	Train Data	Test Data
MAE	1.876	12.362
MSE	6.348	182.787
RMSE	2.519	13.520
$R^2$	0.834	-1.045

### 4.3 Implementasi dan Pengujian Model Data Tahun 2015-2022

#### 4.3.1 Dataset Tahun 2015-2022

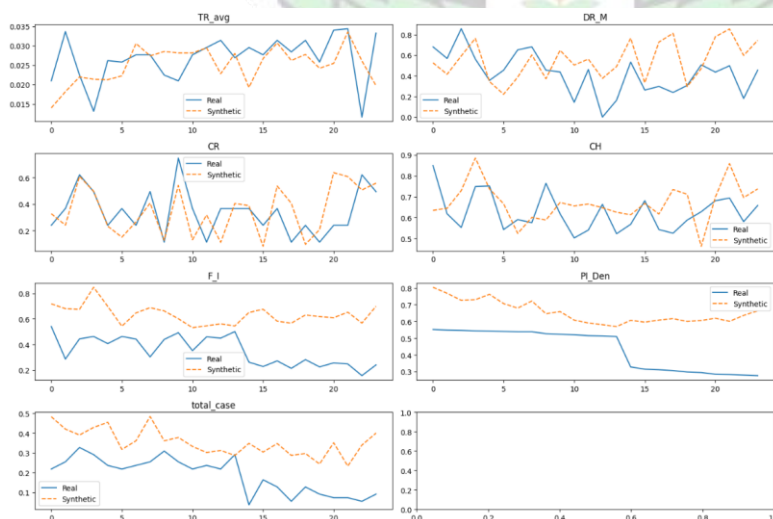
Dataset asli kasus malaria tahun 2019-2022 akan digunakan untuk menghasilkan data sintesis kasus malaria tahun 2015-2018 dengan

mengimplementasikan *TimeGAN* menggunakan *library ydata\_synthetic*. Peneliti melakukan eksperimen pada *hyperparameter TimeGAN* yang meliputi *batch size* dan *epoch*. Detail eksperimen *hyperparameter TimeGAN* dapat dilihat pada tabel 4.3.1.

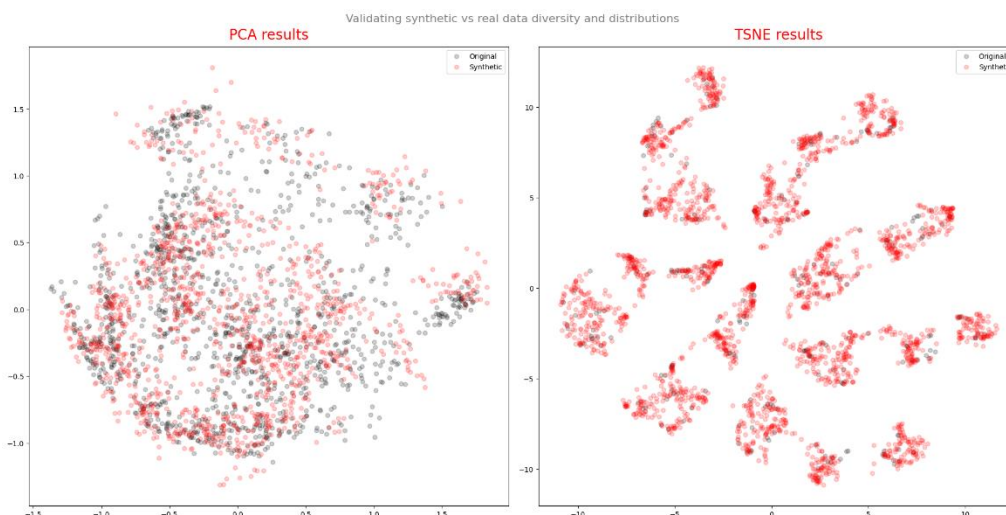
Tabel 4.3.1 Eksperimen *hyperparameter TimeGAN*

	Epoch	Batch Size	$R^2$	MAE
Real	300	64	0.271	0.105
Synthetic			0.055	0.100
Real	300	32	0.365	0.090
Synthetic			0.345	0.097
Real	400	16	0.371	0.083
Synthetic			0.404	0.086

Berdasarkan tabel 4.3.1, dapat dilihat bahwa eksperimen dengan *batch size* 32 dan *epoch* 300 menunjukkan selisih nilai  $R^2$  dan MAE yang paling rendah, yakni 0.365 dan 0.090 untuk data asli serta 0.345 dan 0.097 untuk data sintesis. Oleh karena itu, data sintesis dari eksperimen kedua dipilih karena lebih menyerupai distribusi data asli. Total data sintesis yang ditambahkan adalah sebanyak 205 data sehingga jumlah data gabungan sebanyak 415 data. Gambar plot perbandingan data asli dan data sintesis dapat dilihat pada gambar 4.3.1. Sementara itu, plot perbandingan PCA dan TSNE pada data asli dan data sintesis dapat dilihat pada gambar 4.3.2 yang menunjukkan bahwa data sintesis yang dihasilkan telah mampu mempelajari distribusi umum dari data asli karena memiliki penyebaran yang serupa.

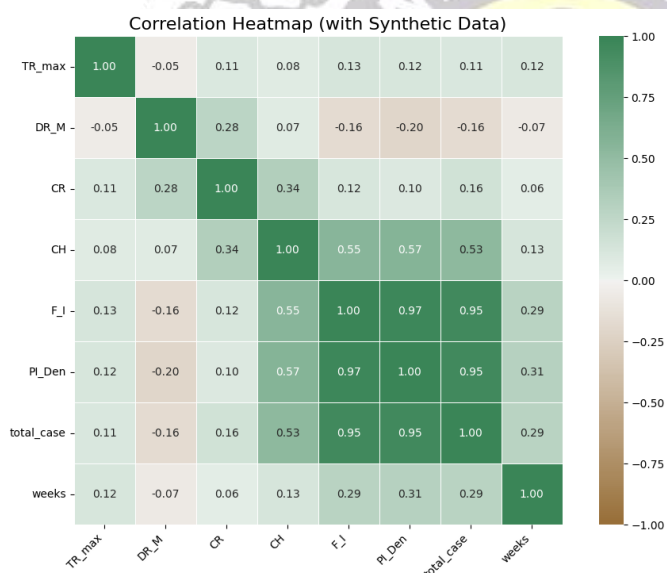


Gambar 4.3.1 Plot perbandingan data asli dan data sintesis



Gambar 4.3.2 Plot perbandingan PCA dan TSNE data asli dan data sintesis

Adapun *Correlation Heatmap* pada data kasus malaria tahun 2015-2022 dapat dilihat pada gambar 4.3.3.



Gambar 4.3.3 *Correlation Heatmap* data kasus malaria 2015-2022

Berdasarkan gambar 4.3.3, indeks tingkat banjir dan genangan air, serta indeks kepadatan dan pertambahan penduduk menunjukkan korelasi positif dengan nilai *Pearson Correlation* yang tinggi terhadap total kasus positif malaria, yakni 0.95 dan 0.95 secara berturut-turut. Hal ini menunjukkan adanya keterhubungan yang erat antara variabel-variabel tersebut dengan variabel target sehingga dapat mempengaruhi hasil prediksi secara signifikan. Sementara itu, indeks curah hujan, indeks jumlah hujan, suhu



maksimum, dan durasi sinar matahari menunjukkan korelasi yang lebih rendah, yakni 0.53, 0.16, 0.11, dan -0.16 secara berturut-turut.

#### 4.3.2 Implementasi Model

Data yang akan di-*train* pada bagian ini adalah data kasus malaria tahun 2015-2022, yaitu data yang telah digabungkan dengan data sintesis. Pada proses ini, peneliti melakukan beberapa eksperimen terhadap parameter-parameter, seperti jumlah iterasi, jumlah agen pencari, jumlah *hidden layer* dan *neuron*, dan *split data*. *Mean Absolute Error* juga digunakan sebagai metrik evaluasi pada eksperimen ini. Pada *setting* awal, parameter yang ditetapkan adalah 2000 iterasi, 20 agen pencari, dan 2 *hidden layer* masing masing sebanyak 4 *neuron*. Detail eksperimen *split data* tertera pada tabel 4.3.2.

Tabel 4.3.2 Eksperimen parameter *split data* tahun 2015-2022

Split	HL 1	HL 2	Iteration	Search Agent	Train MAE	Test MAE
9:1	4	4	2000	20	3.804	14.772
8:2	4	4	2000	20	3.272	8.311
7:3	4	4	2000	20	3.091	6.135

Berdasarkan hasil eksperimen tabel 4.3.2, dapat dilihat bahwa *split data* 7:3 memperoleh *train* MAE dan *test* MAE yang paling rendah yakni sebesar 3.091 dan 6.135 berturut-turut. Eksperimen ketiga memperoleh nilai *test* MAE yang jauh lebih rendah dibandingkan dengan kedua eksperimen lainnya dengan *split data* 9:1 dan 8:2 yang sebesar 14.772 dan 8.311. Hal ini menunjukkan selisih yang cukup jauh dengan *train* MAE sehingga model mungkin mengalami *overfitting*. Oleh karena itu, *split* 7:3 menjadi pilihan terbaik yang akan digunakan dalam model karena memiliki kinerja yang lebih baik pada *test data*.

Selanjutnya, eksperimen akan dilanjutkan dengan pemilihan jumlah *hidden layer* yang paling optimal. Model akan diimplementasikan dengan *split data* 7:3, 4 *neuron* pada tiap *hidden layer*, 20 agen pencari, dan 2000 iterasi. Detail eksperimen jumlah *hidden layer* terlihat pada tabel 4.3.3.

Tabel 4.3.3 Eksperimen jumlah *hidden layer* tahun 2015-2022

Split	Num of HL	Iteration	Search Agent	Train MAE	Test MAE
7:3	1	2000	20	3.620	7.378

Split	Num of HL	Iteration	Search Agent	Train MAE	Test MAE
7:3	2	2000	20	3.091	6.135
7:3	3	2000	20	3.137	5.993

Berdasarkan eksperimen pada tabel 4.3.3, eksperimen kedua dengan 2 *hidden layer* menunjukkan *train* MAE yang paling rendah, yakni 3.091 dengan *test* MAE 6.135. Sementara itu, eksperimen ketiga dengan 3 *hidden layer* menunjukkan *test* MAE yang paling rendah, yakni 5.993 dengan *train* MAE sebesar 3.137. Karena memiliki selisih yang minim, eksperimen kedua dengan jumlah *hidden layer* 2 dan eksperimen ketiga dengan jumlah *hidden layer* 3 akan digunakan dalam model prediksi dan ditelaah lebih lanjut.

Setelah bereksperimen pada jumlah *hidden layer*, langkah selanjutnya adalah eksperimen pada jumlah *neuron* pada *layer* tersebut. Model akan diimplementasikan dengan *split data* 7:3, 2 *hidden layer* atau 3 *hidden layer*, 2000 iterasi, dan 20 agen pencari. Detail eksperimen jumlah *neuron* disajikan pada tabel 4.3.4.

Tabel 4.3.4 Eksperimen jumlah *neuron* tahun 2015-2022

Split	Jumlah HL	Neuron	Iteration	Search Agent	Train MAE	Test MAE	Test R <sup>2</sup>
7:3	2	4,4	2000	20	3.091	6.642	0.746
7:3	2	5,4	2000	20	2.686	5.320	0.808
7:3	2	4,2	2000	20	3.246	6.642	0.705
7:3	2	5,2	2000	20	2.717	5.164	0.771
7:3	3	4,4,4	2000	20	3.137	5.993	0.751
7:3	3	4,4,2	2000	20	3.023	6.803	0.664
7:3	3	4,2,2	2000	20	4.590	9.254	0.293
7:3	3	5,4,4	2000	20	3.472	6.525	0.728
7:3	3	5,4,2	2000	20	2.517	5.555	0.756
7:3	3	5,2,2	2000	20	2.979	7.066	0.491

Berdasarkan eksperimen pada tabel 4.3.4, eksperimen keempat menunjukkan bahwa *hidden layer* dengan jumlah *neuron* 5 dan 2 memperoleh nilai *test* MAE yang paling rendah sebesar 5.164 dengan *train* MAE sebesar 2.717. Sementara itu, eksperimen kesembilan dengan jumlah *neuron* 5, 4, dan 2 menunjukkan nilai *train* MAE

yang paling rendah, yaitu 2.517 dengan nilai *test* MAE sebesar 5.555. Eksperimen kedua dengan jumlah *neuron* 5 dan 4 juga menunjukkan nilai *train* MAE dan *test* MAE yang rendah, yakni sebesar 2.686 dan 5.320. Oleh karena selisih yang minim, nilai *test R-squared* ditambahkan untuk menentukan jumlah *neuron* yang paling optimal. Terlihat bahwa jumlah *neuron* 5 dan 4 memiliki nilai *test R-squared* yang paling tinggi, yakni 0.808 dibandingkan dengan kedua eksperimen lainnya sebesar 0.771 dan 0.756 secara berturut-turut. Sehingga, jumlah *neuron* 5 dan 4 menjadi pilihan terbaik yang akan digunakan dalam model karena lebih mampu menunjukkan variabilitas data.

Setelah melakukan eksperimen untuk jumlah *neuron*, langkah selanjutnya adalah melakukan eksperimen untuk parameter jumlah agen pencari. Agen pencari merupakan vektor bobot dan bias dari *Multilayer Perceptron* yang mungkin dan akan di-*update* posisinya di setiap iterasi untuk meningkatkan akurasi. Model akan diimplementasikan dengan *split data* 7:3, 2 *hidden layer*, 5 dan 4 *neuron* masing-masing *layer*, dan 2000 iterasi, Detail eksperimen jumlah agen pencari tercantum pada tabel 4.3.5.

Tabel 4.3.5 Eksperimen jumlah agen pencari tahun 2015-2022

Split	HL 1	HL 2	Iteration	Search Agent	Train MAE	Test MAE	Test R <sup>2</sup>
7:3	5	4	2000	20	2.686	5.320	0.808
7:3	5	4	2000	40	2.264	5.236	0.712
7:3	5	4	2000	60	2.687	5.708	0.732

Berdasarkan eksperimen pada tabel 4.3.5, ketiga eksperimen tersebut menunjukkan nilai *train* MAE dan *test* MAE yang tidak berbeda jauh. Namun, eksperimen pertama dengan jumlah agen pencari 20 memiliki nilai R<sup>2</sup> yang paling tinggi, yakni sebesar 0.808 dibandingkan dengan eksperimen lain yang sebesar 0.712 dan 0.732. Untuk itu, eksperimen pertama dengan 20 agen pencari akan digunakan dalam model prediksi karena memiliki nilai *test R-squared* yang paling tinggi sehingga lebih mampu menunjukkan variabilitas data.

Setelah melakukan eksperimen jumlah agen pencari, langkah selanjutnya adalah melakukan eksperimen parameter jumlah iterasi. Model akan diimplementasikan

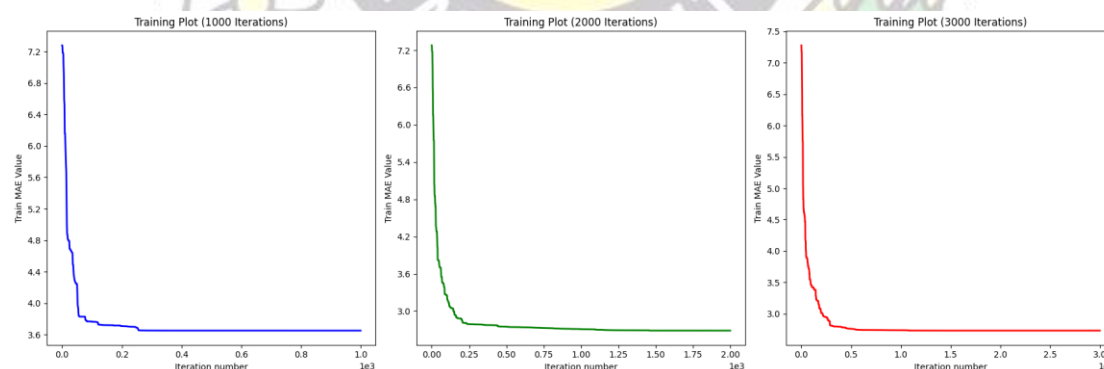
dengan *split data* 7:3, 2 *hidden layer*, 5 dan 4 *neuron* masing-masing *layer*, dan 20 agen pencari. Detail eksperimen jumlah iterasi disajikan pada tabel 4.3.6.

Tabel 4.3.6 Eksperimen jumlah iterasi tahun 2015-2022

Split	HL 1	HL 2	Iteration	Search Agent	Train MAE	Test MAE
7:3	5	4	1000	20	3.653	8.212
7:3	5	4	2000	20	2.686	5.320
7:3	5	4	3000	20	2.732	5.879

Berdasarkan eksperimen pada tabel 4.3.6, terlihat bahwa semakin banyaknya iterasi tidak menunjukkan bahwa bobot dan bias akan semakin optimal. Eksperimen kedua dengan 2000 iterasi menunjukkan *train* MAE dan *test* MAE yang paling rendah, yakni 2.686 dan 5.320 berturut-turut. Oleh karena itu, eksperimen kedua dengan jumlah iterasi 2000 akan digunakan dalam model prediksi karena memiliki kinerja yang lebih baik.

Setelah eksperimen terhadap parameter model MLP-WOA telah dilakukan, diperoleh model yang paling optimal dengan parameter-parameter, yaitu *split data* 7:3, 2000 iterasi, 20 agen pencari, 2 *hidden layer*, serta 5 dan 4 *neuron* pada setiap *hidden layer*. Plot hasil optimasi bobot dan bias dengan menggunakan *training data* pada model MLP-WOA dengan parameter tersebut ditunjukkan pada gambar 4.3.4.



Gambar 4.3.4 Plot optimasi *training data* tiap iterasi tahun 2015-2022

#### 4.3.3 Hasil Pengujian Model

Dataset yang digunakan sebagai *test data* adalah data kejadian malaria mingguan dari tanggal 10 Agustus 2020 – 2 Januari 2023. Hasil evaluasi dengan metrik MAE pada data *testing* pada bagian 4.3 yakni sebesar 5.320. Selanjutnya, hasil prediksi



ini akan digabungkan dengan total kasus aktual sesuai dengan periode waktunya sehingga dapat dibandingkan yang mana tertera pada tabel 4.3.7.

Tabel 4.3.7 Perbandingan nilai aktual dan prediksi pada *test data* 2015-2022

Weeks	Actual	Predicted
2020-08-10 00:00:00	0	9
2020-08-17 00:00:00	4	15
2020-08-24 00:00:00	3	2
2020-08-31 00:00:00	19	5
2020-09-07 00:00:00	5	11
2020-09-14 00:00:00	7	15
2020-09-21 00:00:00	2	6
2020-09-28 00:00:00	0	3
2020-10-05 00:00:00	0	4
2020-10-12 00:00:00	1	4
2020-10-19 00:00:00	0	8
2020-10-26 00:00:00	5	7
2020-11-02 00:00:00	3	11
2020-11-09 00:00:00	4	6
2020-11-16 00:00:00	4	7
2020-11-23 00:00:00	5	6
2020-11-30 00:00:00	7	10
2020-12-07 00:00:00	3	9
2020-12-14 00:00:00	7	10
2020-12-21 00:00:00	9	10
2020-12-28 00:00:00	2	7
2021-01-04 00:00:00	16	21
2021-01-11 00:00:00	12	23
2021-01-18 00:00:00	13	15
2021-01-25 00:00:00	12	18
2021-02-01 00:00:00	14	19
2021-02-08 00:00:00	17	16

Weeks	Actual	Predicted
2021-02-15 00:00:00	14	10
2021-02-22 00:00:00	13	13
2021-03-01 00:00:00	12	17
2021-03-08 00:00:00	13	17
2021-03-15 00:00:00	16	17
2021-03-22 00:00:00	18	11
2021-03-29 00:00:00	14	11
2021-04-05 00:00:00	12	16
2021-04-12 00:00:00	13	11
2021-04-19 00:00:00	18	18
2021-04-26 00:00:00	15	7
2021-05-03 00:00:00	20	15
2021-05-10 00:00:00	17	13
2021-05-17 00:00:00	16	17
2021-05-24 00:00:00	22	20
2021-05-31 00:00:00	12	15
2021-06-07 00:00:00	12	21
2021-06-14 00:00:00	15	16
2021-06-21 00:00:00	14	17
2021-06-28 00:00:00	16	19
2021-07-05 00:00:00	17	14
2021-07-12 00:00:00	15	22
2021-07-19 00:00:00	16	17
2021-07-26 00:00:00	18	13
2021-08-02 00:00:00	16	18
2021-08-09 00:00:00	17	18
2021-08-16 00:00:00	13	12
2021-08-23 00:00:00	14	13
2021-08-30 00:00:00	16	14
2021-09-06 00:00:00	19	21
2021-09-13 00:00:00	14	12

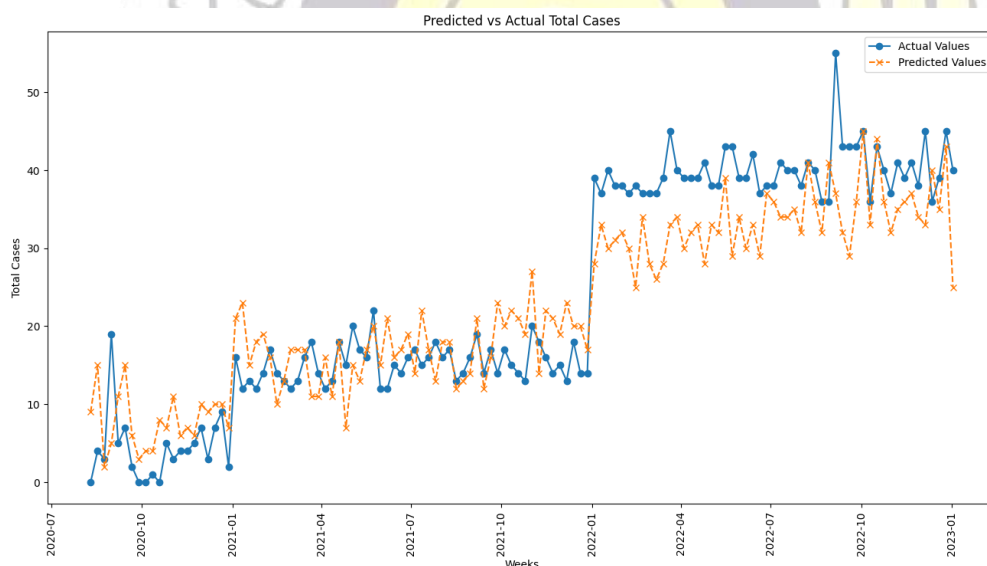
Weeks	Actual	Predicted
2021-09-20 00:00:00	17	16
2021-09-27 00:00:00	14	23
2021-10-04 00:00:00	17	20
2021-10-11 00:00:00	15	22
2021-10-18 00:00:00	14	21
2021-10-25 00:00:00	13	19
2021-11-01 00:00:00	20	27
2021-11-08 00:00:00	18	14
2021-11-15 00:00:00	16	22
2021-11-22 00:00:00	14	21
2021-11-29 00:00:00	15	19
2021-12-06 00:00:00	13	23
2021-12-13 00:00:00	18	20
2021-12-20 00:00:00	14	20
2021-12-27 00:00:00	14	17
2022-01-03 00:00:00	39	28
2022-01-10 00:00:00	37	33
2022-01-17 00:00:00	40	30
2022-01-24 00:00:00	38	31
2022-01-31 00:00:00	38	32
2022-02-07 00:00:00	37	30
2022-02-14 00:00:00	38	25
2022-02-21 00:00:00	37	34
2022-02-28 00:00:00	37	28
2022-03-07 00:00:00	37	26
2022-03-14 00:00:00	39	28
2022-03-21 00:00:00	45	33
2022-03-28 00:00:00	40	34
2022-04-04 00:00:00	39	30
2022-04-11 00:00:00	39	32
2022-04-18 00:00:00	39	33

Weeks	Actual	Predicted
2022-04-25 00:00:00	41	28
2022-05-02 00:00:00	38	33
2022-05-09 00:00:00	38	32
2022-05-16 00:00:00	43	39
2022-05-23 00:00:00	43	29
2022-05-30 00:00:00	39	34
2022-06-06 00:00:00	39	30
2022-06-13 00:00:00	42	33
2022-06-20 00:00:00	37	29
2022-06-27 00:00:00	38	37
2022-07-04 00:00:00	38	36
2022-07-11 00:00:00	41	34
2022-07-18 00:00:00	40	34
2022-07-25 00:00:00	40	35
2022-08-01 00:00:00	38	32
2022-08-08 00:00:00	41	41
2022-08-15 00:00:00	40	36
2022-08-22 00:00:00	36	32
2022-08-29 00:00:00	36	41
2022-09-05 00:00:00	55	37
2022-09-12 00:00:00	43	32
2022-09-19 00:00:00	43	29
2022-09-26 00:00:00	43	36
2022-10-03 00:00:00	45	45
2022-10-10 00:00:00	36	33
2022-10-17 00:00:00	43	44
2022-10-24 00:00:00	40	36
2022-10-31 00:00:00	37	32
2022-11-07 00:00:00	41	35
2022-11-14 00:00:00	39	36
2022-11-21 00:00:00	41	37



Weeks	Actual	Predicted
2022-11-28 00:00:00	38	34
2022-12-05 00:00:00	45	33
2022-12-12 00:00:00	36	40
2022-12-19 00:00:00	39	35
2022-12-26 00:00:00	45	43
2023-01-02 00:00:00	40	25

Berdasarkan tabel 4.3.7, nilai aktual dengan jumlah total kasus di atas 35 menunjukkan performa yang meningkat secara signifikan dibandingkan dengan model *training* data tahun 2019-2022 yang ditunjukkan pada tabel 4.2.6. Hal ini ditunjukkan seperti pada 21 Februari 2022, total kasus malaria mingguan sebanyak 37 kasus diprediksi dengan jumlah 34 kasus. Perbandingan tersebut kemudian dibuat menjadi plot perbandingan dengan menggunakan *line plot*. Adapun plot perbandingan hasil prediksi dan nilai aktual terlihat pada gambar 4.3.5. Disimpulkan bahwa data sintesis meningkatkan akurasi model secara signifikan.



Gambar 4.3.5 Plot perbandingan nilai aktual dan prediksi *test data* 2015-2022

#### 4.3.4 Evaluasi

Setelah melakukan prediksi dengan menggunakan data *training* dan *testing* dengan *Mean Absolute Error* sebagai *cost function*, selanjutnya juga akan dilakukan evaluasi dengan menggunakan metrik evaluasi lainnya. Evaluasi hasil prediksi terlihat pada tabel 4.3.8. Hasil *test R<sup>2</sup>* pada model *training* ini adalah sebesar 0.808. Hal ini

menunjukkan bahwa model ini mampu menunjukkan 80.8% dari variabilitas data.

Tabel 4.3.8 Evaluasi hasil prediksi tahun 2015-2022

Metrics	Train Data	Test Data
MAE	2.686	5.320
MSE	14.040	40.910
RMSE	3.747	6.396
$R^2$	0.912	0.808

Penulis membandingkan nilai evaluasi hasil prediksi untuk data gabungan sintesis dan asli tahun 2015-2022 pada tabel 4.3.8 dengan evaluasi hasil prediksi untuk data asli tahun 2019-2022 seperti yang tercantum pada tabel 4.2.7, evaluasi hasil prediksi data tahun 2015-2022 memiliki nilai *test* MAE, MSE, dan RMSE yang jauh lebih rendah, serta nilai *test*  $R^2$  yang jauh lebih tinggi sehingga dapat disimpulkan bahwa data sintesis dapat meningkatkan akurasi model secara signifikan.

#### 4.4 Diskusi

Metode *Multilayer Perceptron-Whale Optimization Algorithm* telah diimplementasikan dalam penelitian ini untuk memprediksi angka penyebaran malaria di Kabupaten Asahan dengan memanfaatkan faktor-faktor meteorologi. Demi memperoleh model prediksi yang paling optimal, penulis telah melakukan beberapa eksperimen pada *hyperparameter* MLP-WOA, seperti *split data*, iterasi WOA, jumlah agen pencari (*whale population*), jumlah *hidden layer* dan *neuron*. Hasil prediksi yang paling optimal menunjukkan nilai *train* MAE, MSE, RMSE, dan  $R^2$  sebesar 2.686, 14.040, 3.747, dan 0.912 secara berturut-turut. Sementara itu, nilai *test* MAE, MSE, RMSE, dan  $R^2$  dari model prediksi ini adalah sebesar 5.320, 40.910, 6.396, dan 0.808 secara berturut-turut. Hasil prediksi tersebut menunjukkan bahwa model ini mampu menunjukkan variabilitas data, serta mempelajari pola kejadian malaria dengan cukup baik.

Penambahan data sintesis dengan menggunakan *Time-series Generative Adversarial Network* (TimeGAN) terbukti berhasil meningkatkan kemampuan model dalam mengenali pola data *time-series*. Dataset asli hanya memuat data kejadian malaria mingguan selama 4 tahun yang mana masih tergolong sedikit. Untuk

menghindari terjadinya *underfitting* atau *overfitting* pada saat *model training*, penulis menambahkan data sintesis 4 tahun sebelumnya dengan menggunakan TimeGAN.

Berdasarkan data yang digunakan, terdapat tren peningkatan kasus positif malaria pada periode waktu tertentu, seperti awal tahun 2016, November 2017, Oktober 2018, awal tahun 2019, awal tahun 2021, dan awal tahun 2022. Sementara itu, jumlah kasus positif pada periode waktu lainnya tergolong rendah. Peningkatan kasus positif berhubungan erat dengan faktor iklim. Kondisi lingkungan seperti curah hujan yang tinggi, banjir, genangan air, dan kepadatan penduduk mendukung perkembangbiakan nyamuk sehingga meningkatkan tren kasus positif malaria.

Untuk penelitian selanjutnya, penulis merekomendasikan untuk menggunakan lebih banyak data, terutama data historikal dengan tujuan dapat meningkatkan kinerja model. Selain itu, penulis juga menyarankan eksplorasi lebih jauh terkait parameter lain yang digunakan. Proses *hyperparameter tuning*, seperti eksperimen pada *learning rate*, *regularization*, *batch size*, dan *epoch*, juga dapat ditambahkan untuk meningkatkan performa model terutama dalam mengurangi *overfitting*. Secara keseluruhan, penelitian ini memberikan kontribusi penting dalam perkembangan prediksi angka kejadian malaria berdasarkan faktor iklim atau meteorologi. Walaupun demikian, penelitian ini berpotensi untuk perbaikan lebih lanjut terutama terkait jumlah data dan parameter model.

## BAB 5

### KESIMPULAN DAN SARAN

#### 5.1 Kesimpulan

Kesimpulan terkait prediksi angka penyebaran malaria menggunakan metode *Multilayer Perceptron-Whale Optimization Algorithm* adalah sebagai berikut:

1. Metode MLP-WOA dapat melakukan prediksi angka penyebaran malaria dengan memanfaatkan faktor meteorologi dan memperoleh nilai *test* MAE, MSE, RMSE, dan  $R^2$  sebesar 5.320, 40.910, 6.396, dan 0.808 secara berturut-turut.
2. Penggunaan *Time-series Generative Adversarial Networks* (TimeGAN) menggunakan *library python* yaitu *ydata-synthetic* membantu menambahkan data historikal dan meningkatkan performa model dengan cukup signifikan.
3. Berdasarkan data kejadian malaria dalam penelitian ini, curah hujan, banjir dan genangan air, serta kepadatan penduduk menunjukkan korelasi positif terhadap tren kejadian malaria dan memiliki pengaruh signifikan terhadap kasus positif malaria di Kabupaten Asahan.

#### 5.2 Saran

Saran yang dapat dipertimbangkan untuk penelitian selanjutnya meliputi:

1. Penggunaan lebih banyak data historikal sehingga model dapat mempelajari tren data dengan lebih baik dan mengurangi tingkat *error*.
2. Mempertimbangkan faktor pendukung lainnya, seperti kasus impor, tingkat pendidikan, indeks vegetasi, ataupun faktor sosial-ekonomi lain yang berkaitan dengan penyebaran malaria.
3. Menggunakan data dari lokasi yang berbeda agar model dapat digeneralisasikan dan diimplementasikan pada wilayah lain.
4. Mengeksplor lebih jauh terkait penggunaan hyperparameter lain, seperti *learning rate*, *regularization*, *batch size*, dan *epoch* untuk mengurangi *overfitting*.



## DAFTAR PUSTAKA

- Adamu, Y. A., & Singh, J. (2021). Malaria prediction model using advanced ensemble machine learning techniques. *Journal of Medical Pharmaceutical and Allied Sciences*, 10(6), 3794–3801. <https://doi.org/10.22270/jmpas.V10I6.1701>
- Adeola, A. M., Botai, J. O., Olwoch, J. M., De W Rautenbach, H. C. J., Adisa, O. M., De Jager, C., Botai, C. M., & Aaron, M. T. (2019). Predicting malaria cases using remotely sensed environmental variables in Nkomazi, South Africa. *Geospatial Health*, 14(1). <https://doi.org/10.4081/gh.2019.676>
- Aheto, J. M. K., Duah, H. O., Agbadi, P., & Nakua, E. K. (2021). A predictive model, and predictors of under-five child malaria prevalence in Ghana: How do LASSO, Ridge and Elastic net regression approaches compare? *Preventive Medicine Reports*, 23. <https://doi.org/10.1016/j.pmedr.2021.101475>
- Alameer, Z., Elaziz, M. A., Ewees, A. A., Ye, H., & Jianhua, Z. (2019). Forecasting gold price fluctuations using improved multilayer perceptron neural network and whale optimization algorithm. *Resources Policy*, 61, 250–260. <https://doi.org/10.1016/j.resourpol.2019.02.014>
- Badan Pusat Statistik. (2024). Kejadian Malaria Per 1000 Orang 2019-2021. Badan Pusat Statistik. Diakses pada 31 Mei 2024. <https://www.bps.go.id/id/statistics-table/2/MTM5MyMy/kejadian-malaria-per-1000-orang.html>
- Badan Pusat Statistik. (2024). Jumlah Kabupaten/Kota yang mencapai eliminasi Malaria. Diakses pada 31 Mei 2024. <https://www.bps.go.id/id/statistics-table/2/MTc2NCMy/jumlah-kabupaten-kota-yang-mencapai-eliminasi-malaria.html>
- Badan Pusat Statistik. (2024). Jumlah Penduduk menurut Jenis Kelamin dan Kabupaten/Kota (Jiwa), 2021-2023. Badan Pusat Statistik. Diakses pada 31 Mei 2024. <https://sumut.bps.go.id/indicator/12/65/1/jumlah-penduduk-menurut-jenis-kelamin-dan-kabupaten-kota.html>
- Basheer, I. A., & Hajmeer, M. N. (2000). Artificial neural networks: fundamentals, computing, design, and application. *Journal of Microbiological Methods*, 43(1), 3–31. [https://doi.org/10.1016/s0167-7012\(00\)00201-3](https://doi.org/10.1016/s0167-7012(00)00201-3)

- Bates, D. M., Mächler, M., Bolker, B. M., & Walker, S. (2014). Package LME4: Linear Mixed-Effects models using EIGEN and S4. ResearchGate. [https://www.researchgate.net/publication/279236477\\_Package\\_Lme4\\_Linear\\_Mixed-Effects\\_Models\\_Using\\_Eigen\\_and\\_S4](https://www.researchgate.net/publication/279236477_Package_Lme4_Linear_Mixed-Effects_Models_Using_Eigen_and_S4)
- Biologijk. (2017). 4 Jenis Plasmodium Penyebab Penyakit Malaria, Gejala, Cara Penularan, Pengobatan dan Pencegahannya. Diakses pada 26 Januari 2024. <https://www.biologijk.com/2017/10/jenis-plasmodium-penyebab-penyakit-malaria.html>
- BPS Kabupaten Asahan. (2018). Kabupaten Asahan dalam Angka. <https://northsumatrainvest.id/data/pdf/publication/2018-BPS-Kabupaten%20Asahan%20Dalam%20Angka-compressed.pdf>
- Brophy, E., Wang, Z., She, Q., & Ward, T. (2021). Generative adversarial networks in time series: A survey and taxonomy. <http://arxiv.org/abs/2107.11098>
- Chen, S., Whiteman, A., Li, A., Rapp, T., Delmelle, E., Chen, G., Brown, C. L., Robinson, P., Coffman, M. J., Janies, D., & Dulin, M. (2019). An operational machine learning approach to predict mosquito abundance based on socioeconomic and landscape patterns. *Landscape Ecology*, 34(6), 1295–1311. <https://doi.org/10.1007/s10980-019-00839-2>
- De Souza, V. L. T., Marques, B. a. D. Batagelo, H. C., & Góis, J. P. (2023). A review on Generative Adversarial Networks for image generation. *Computers & Graphics*, 114, 13–25. <https://doi.org/10.1016/j.cag.2023.05.010>
- Diop, L., Samadianfard, S., Bodian, A., Yaseen, Z. M., Ghorbani, M. A., & Salimi, H. (2020). Annual Rainfall Forecasting Using Hybrid Artificial Intelligence Model: Integration of Multilayer Perceptron with Whale Optimization Algorithm. *Water Resources Management*, 34(2), 733–746. <https://doi.org/10.1007/s11269-019-02473-8>
- Erlita. (2017). 4 Jenis-jenis Malaria – Penyebab, Gejala dan Pengobatan. HaloSehat. Diakses pada 23 Maret 2024. HaloSehat. <https://halosehat.com/penyakit/malaria/jenis-jenis-malaria>
- Ewees, A. A., Elaziz, M. A., Alameer, Z., Ye, H., & Jianhua, Z. (2020). Improving multilayer perceptron neural network using chaotic grasshopper optimization algorithm to forecast iron ore price volatility. *Resources Policy*, 65. <https://doi.org/10.1016/j.resourpol.2019.101555>

- Fahmi, F., Pasaribu, A. P., Theodora, M., & Wangdi, K. (2022). Spatial analysis to evaluate risk of malaria in Northern Sumatera, Indonesia. *Malaria Journal*, 21(1). <https://doi.org/10.1186/s12936-022-04262-y>
- Fisher, T. M., Rojas–Galeano, S., & Fernandez-Reyes, D. (2023). Specialist hybrid models with asymmetric training for malaria prevalence prediction. *Frontiers in Public Health*, 11. <https://doi.org/10.3389/fpubh.2023.1207624>
- Garrido-Cardenas, J. A., Cebrián-Carmona, J., González-Cerón, L., Manzano-Agugliaro, F., & Mesa-Valle, C. (2019). Analysis of global research on malaria and *Plasmodium vivax*. *International Journal of Environmental Research and Public Health*, 16(11). <https://doi.org/10.3390/ijerph16111928>
- Gurney, K. (1997). An introduction to neural networks.
- Haddawy, P., Hasan, A. H., Kasantikul, R., Lawpoolsri, S., Sa-Angchai, P., Kaewkungwal, J., & Singhasivanon, P. (2018). Spatiotemporal Bayesian networks for malaria prediction. *Artificial Intelligence in Medicine*, 84, 127–138. <https://doi.org/10.1016/j.artmed.2017.12.002>
- Hecht-Nielsen, R. (1992). Theory of the Backpropagation Neural Network (pp. 65–93). Elsevier. <https://doi.org/10.1016/b978-0-12-741252-8.50010-8>
- Holland, J. H. (1992). Genetic algorithms. *Scientific American*, 267(1), 66–72. <https://doi.org/10.1038/scientificamerican0792-66>
- Hussien, H. H. (2019). Malaria's association with climatic variables and an epidemic early warning system using historical data from Gezira State, Sudan. *Heliyon*, 5(3). <https://doi.org/10.1016/j.heliyon.2019.e01375>
- Kalipe, G., Gautham, V., & Behera, R. K. (2018). Predicting Malarial Outbreak using Machine Learning and Deep Learning Approach: A Review and Analysis. In *Proceedings - 2018 International Conference on Information Technology, ICIT 2018* (pp. 33–38). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICIT.2018.00019>
- Kennedy, J., & Eberhart, R. C. (1995). Particle Swarm Optimization. <https://doi.org/10.1109/icnn.1995.488968>
- Kim, Y., Ratnam, J. V., Doi, T., Morioka, Y., Behera, S., Tsuzuki, A., Minakawa, N., Sweijd, N., Kruger, P., Maharaj, R., Imai, C. C., Ng, C. F. S., Chung, Y., & Hashizume, M. (2019). Malaria predictions based on seasonal climate forecasts



- in South Africa: A time series distributed lag nonlinear model. *Scientific Reports*, 9(1), 1–10. <https://doi.org/10.1038/s41598-019-53838-3>
- Kurnianingsih, K., Wirasatriya, A., Lazuardi, L., Kubota, N., & Ng, N. (2020). IOD and ENSO-Related Time Series Variability and Forecasting of Dengue and Malaria Incidence in Indonesia. IOD And ENSO-Related Time Series Variability and Forecasting of Dengue and Malaria Incidence in Indonesia. <https://doi.org/10.1109/ccs49175.2020.9231358>
- Mbunge, E., Millham, R. C., Sibiya, M. N., & Takavarasha, S. (2022). Application of machine learning models to predict malaria using malaria cases and environmental risk factors. In 2022 Conference on Information Communications Technology and Society, ICTAS 2022 - Proceedings. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICTAS53252.2022.9744657>
- Mfisimana, L. D., Nibayisabe, E., Badu, K., & Niyukuri, D. (2022). Exploring predictive frameworks for malaria in Burundi. *Infectious Disease Modelling*, 7(2), 33–44. <https://doi.org/10.1016/j.idm.2022.03.003>
- Mirjalili, S. & Lewis, A. (2014). Grey Wolf Optimizer. *Advances in Engineering Software*, 69, 46–61. <https://doi.org/10.1016/j.advengsoft.2013.12.007>
- Mirjalili, S., & Lewis, A. (2016). The Whale Optimization Algorithm. *Advances in Engineering Software*, 95, 51–67. <https://doi.org/10.1016/j.advengsoft.2016.01.008>
- Mopuri, R., Kakarla, S. G., Mutheneni, S. R., Kadiri, M. R., & Kumaraswamy, S. (2020). Climate based malaria forecasting system for Andhra Pradesh, India. *Journal of Parasitic Diseases*, 44(3), 497–510. <https://doi.org/10.1007/s12639-020-01216-6>
- Munif, A. (2009). Nyamuk Vektor Malaria dan Hubungannya Dengan Aktivitas Kehidupan Manusia Di Indonesia. *ASPIRATOR - Journal of Vector-borne Disease Studies*, 1(2), 53351. <https://doi.org/10.22435/aspirator.v1i2.2936>
- Nkiruka, O., Prasad, R., & Clement, O. (2021). Prediction of malaria incidence using climate variability and machine learning. *Informatics in Medicine Unlocked*, 22. <https://doi.org/10.1016/j.imu.2020.100508>
- Rokom. (2022). Kejar Target Bebas Malaria 2030, Kemenkes Tetapkan 5 Regional Target Eliminasi. Sehat Negeriku.



- <https://sehatnegeriku.kemkes.go.id/baca/rilis-media/20220422/1439692/kejar-target-bebas-malaria-2030-kemenkes-tetapkan-5-regional-target-eliminasi/>
- Sadoine, M. L., Smargiassi, A., Ridde, V., Tusting, L. S., & Zinszer, K. (2018). The associations between malaria, interventions, and the environment: a systematic review and meta-analysis. *Malaria Journal*, 17(1). <https://doi.org/10.1186/s12936-018-2220-x>
- Samadianfard, S., Hashemi, S., Kargar, K., Izadyar, M., Mostafaeipour, A., Mosavi, A., Nabipour, N., & Shamshirband, S. (2020). Wind speed prediction using a hybrid model of the multi-layer perceptron and whale optimization algorithm. *Energy Reports*, 6, 1147–1159. <https://doi.org/10.1016/j.egyr.2020.05.001>
- Santosh, T., Ramesh, D., & Reddy, D. (2020). LSTM based prediction of malaria abundances using big data. *Computers in Biology and Medicine*, 124. <https://doi.org/10.1016/j.combiomed.2020.103859>
- Sianturi, A. A. (2023). Letak Geografis, Batas Wilayah, serta Iklim di Sumatera Utara. *detikSumut*. Diakses pada 30 Januari 2024. <https://www.detik.com/sumut/berita/d-7076810/letak-geografis-batas-wilayah-septa-iklim-di-sumatera-utara>
- Sinaga, N. (2022). Tiga Kabupaten di Sumut Endemis Sedang Malaria, Pasokan Obat Terbatas. *kompas.id*. <https://www.kompas.id/baca/nusantara/2022/09/27/tiga-kabupaten-sumut-endemis-sedang-malaria-pasokan-obat-terbatas>
- Singh, A., Mehra, M., Kumar, A., Niranjannaik, M., Priya, D. B., & Gaurav, K. (2023). Leveraging hybrid machine learning and data fusion for accurate mapping of malaria cases using meteorological variables in western India. *Intelligent Systems With Applications*, 17, 200164. <https://doi.org/10.1016/j.iswa.2022.200164>
- Ssempiira, J., Kissa, J., Nambuusi, B., Mukooyo, E., Opigo, J., Makumbi, F., Kasasa, S., & Vounatsou, P. (2018). Interactions between climatic changes and intervention effects on malaria spatio-temporal dynamics in Uganda. *Parasite Epidemiology and Control*, 3(3), e00070. <https://doi.org/10.1016/j.parepi.2018.e00070>
- Svozil, D., Kvasnička, V., & Pospíchal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems*, 39(1), 43–62. [https://doi.org/10.1016/s0169-7439\(97\)00061-0](https://doi.org/10.1016/s0169-7439(97)00061-0)

- Thakur, S., & Dharavath, R. (2019). Artificial neural network based prediction of malaria abundances using big data: A knowledge capturing approach. *Clinical Epidemiology and Global Health*, 7(1), 121–126. <https://doi.org/10.1016/j.cegh.2018.03.001>
- Thirumagal, E., & Saruladha, K. (2021). GAN models in natural language processing and image translation. In *Elsevier eBooks* (pp. 17–57). <https://doi.org/10.1016/b978-0-12-823519-5.00001-4>
- Wang, M., Wang, H., Wang, J., Liu, H., Lu, R., Duan, T., et al. (2019). A novel model for malaria prediction based on ensemble algorithms. *PLoS ONE*, 14(12). <https://doi.org/10.1371/journal.pone.0226910>
- World Health Organization. (2023). Malaria. World Health Organization. Diakses pada 26 Januari 2024. <https://www.who.int/news-room/fact-sheets/detail/malaria>
- World Health Organization. (2021). Global technical strategy for malaria 2016-2030, 2021. World Health Organization. Diakses pada 26 Januari 2024. <https://www.who.int/publications/i/item/9789240031357>
- World Health Organization. (2023). World malaria report 2023, 1–356. <https://www.who.int/teams/global-malaria-programme/reports/world-malaria-report-2023>
- Wiese, M., Knobloch, R., Korn, R., & Kretschmer, P. (2020). Quant GANs: deep generation of financial time series. *Quantitative Finance*, 20(9), 1419–1440. <https://doi.org/10.1080/14697688.2020.1730426>
- Xu, L., Skoularidou, M., Cuesta-Infante, A., & Veeramachaneni, K. (2019). Modeling Tabular data using Conditional GAN. *arXiv (Cornell University)*, 32, 7333–7343. <https://arxiv.org/pdf/1907.00503.pdf>
- Yu, L., Zhang, W., Wang, J., & Yu, Y. (2017). SeqGAN: Sequence Generative Adversarial Nets with Policy Gradient. *AAAI Conference on Artificial Intelligence*, 31(1). <https://doi.org/10.1609/aaai.v31i1.10804>
- Yoon, J., Jarrett, D., & van der Schaar, M. (2019). Time-series generative adversarial networks. *Advances in Neural Information Processing Systems*, 32(NeurIPS), 1–11.
- Zhang, M., Wu, Q., Chen, H., Heidari, A. A., Cai, Z., Li, J., Md. Abdelrahim, E., & Mansour, R. F. (2023). Whale optimization with random contraction and Rosenbrock method for COVID-19 disease prediction. *Biomedical Signal*

*Processing and Control*, 83(August 2022).  
<https://doi.org/10.1016/j.bspc.2023.104638>

- Zhang, Y., Bi, P., & Hiller, J. E. (2010). Meteorological variables and malaria in a Chinese temperate city: A twenty-year time-series data analysis. *Environment International*, 36(5), 439–445. <https://doi.org/10.1016/j.envint.2010.03.005>
- Zhou, T., Jiang, Z., Liu, X., & Tan, K. (2020). Research on the long-term and short-term forecasts of navigable river's water-level fluctuation based on the adaptive multilayer perceptron. *Journal of Hydrology*. Elsevier B.V. <https://doi.org/10.1016/j.jhydrol.2020.125285>
- Zinszer, K., Kigozi, R., Charland, K., Dorsey, G., Brewer, T. F., Brownstein, J. S., Kamya, M. R., & Buckeridge, D. L. (2015). Forecasting malaria in a highly endemic country using environmental and clinical predictors. *Malaria Journal*, 14(1). <https://doi.org/10.1186/s12936-015-0758-4>

