

**PREDIKSI JUMLAH KASUS MALARIA DI LABUHANBATU UTARA
MENGUNAKAN METODE GAUSSIAN PROCESS REGRESSION**

SKRIPSI

ADINDA KHAIRANI

201402019



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

MEDAN

2024

**PREDIKSI JUMLAH KASUS MALARIA DI LABUHANBATU UTARA
MENGUNAKAN METODE *GAUSSIAN PROCESS REGRESSION***

SKRIPSI

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

ADINDA KHAIRANI

201402019



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

MEDAN

2024

PERSETUJUAN

Judul : PREDIKSI JUMLAH KASUS MALARIA DI
LABUHANBATU UTARA MENGGUNAKAN
METODE *GAUSSIAN PROCESS REGRESSION*

Kategori : SKRIPSI

Nama : ADINDA KHAIRANI

Nomor Induk Mahasiswa : 201402019

Program Studi : TEKNOLOGI INFORMASI

Fakultas : ILMU KOMPUTER DAN TEKNOLOGI
INFORMASI

Medan, 16 Oktober 2024

Komisi Pembimbing

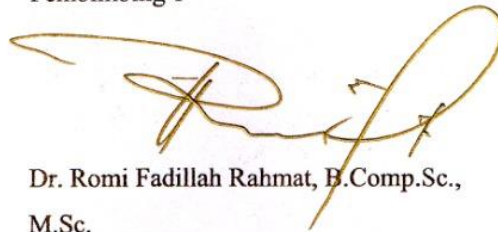
Pembimbing 2



Sarah Purnamawati, S.T.,
M.Sc.

NIP. 198302262010122003

Pembimbing 1



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

NIP.198603032010121004

Diketahui/disetujui oleh

Program Studi S1 Teknologi Informasi

Ketua Program Studi




Dedy Arisandi, S.T., M.Kom.
197908312009121002

PERNYATAAN**PREDIKSI JUMLAH KASUS MALARIA DI LABUHANBATU UTARA
MENGUNAKAN METODE *GAUSSIAN PROCESS REGRESSION*****SKRIPSI**

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

Medan, 16 Oktober 2024



Adinda Khairani

UCAPAN TERIMA KASIH

Puji dan syukur penulis sampaikan kepada Allah SWT, Tuhan satu-satunya yang telah memberikan nikmat, rahmat dan ridha-Nya sehingga penulis dapat menyelesaikan skripsi sebagai salah satu syarat lulus dan memperoleh gelar Sarjana Komputer pada program studi S1 Teknologi Informasi pada Fakultas Ilmu Komputer dan Teknologi Informasi di Universitas Sumatera Utara.

Penulisan skripsi ini terselesaikan karena dukungan moral dan materil yang diberikan oleh beberapa pihak kepada penulis. Penulis ingin menyampaikan ucapan terima kasih sebesar-besarnya sebagai bentuk rasa syukur atas doa, bantuan, motivasi dan kerjasama dari pihak yang memberikan dukungan dalam menyelesaikan skripsi ini, yakni kepada:

1. Allah SWT, atas segala rahmat dan hidayah-Nya yang selalu memberikan kekuatan dan kesabaran dalam menjalani setiap proses pengerjaan skripsi ini.
2. Keluarga penulis, Papa, Mama, Abang dan Kakak yang selalu mendoakan, mengingatkan dan memberikan dukungan kepada penulis untuk menyelesaikan skripsi ini.
3. Ibu Dr. Maya Silvi Lydia, B.Sc., M.Sc., selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Sumatera Utara.
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. Bapak Dr. Romi Fadillah Rahmat B.Comp.Sc., M.Sc., selaku Dosen Pembimbing pertama penulis dan Ibu Sarah Purnamawati, S.T., M.Sc., selaku Dosen Pembimbing kedua penulis yang telah meluangkan waktu untuk membimbing, mengarahkan, dan memberikan masukan yang sangat berharga selama proses penyusunan skripsi ini.
6. Bapak Ainul Hizriadi, S.Kom., M.Sc., selaku Dosen Penguji pertama penulis dan Ibu Fanindia Purnamasari, S.TI., M.IT., selaku Dosen Penguji kedua penulis yang telah memberikan kritik, saran, serta masukan yang sangat membangun dalam skripsi ini.

7. Staf dan pegawai Fakultas Ilmu Komputer dan Teknologi Informasi, yang telah memberikan bantuan administrasi selama perkuliahan berlangsung.
8. Teman-teman perkuliahan, Diah Paramitha, Putri Yanti Nahampun, Della Febriana, Monika Angelia Panjaitan, Irma Nathasya Putri dan Ruth Damayanthi Purba yang telah berjuang bersama dari awal sampai akhir perkuliahan ini.
9. Teman-teman seperbimbingan Muhammad Arief Fadhlhan, Herzinanda Putra, Levina Gunawan, Davita Chandra, Felix Kurnia Salim dan Cynthia Yapiter yang telah berjuang dan saling membantu dalam menyelesaikan skripsi ini.
10. Bang Rhama Permadi, dan Bang Iqbal Fakhriza selaku senior penulis yang telah memberikan bantuan dan masukan kepada penulis dalam mengerjakan skripsi.
11. Seluruh teman-teman Angkatan 2020 yang berjuang bersama dalam menyelesaikan skripsi.

Penulis menyadari, penelitian dan skripsi ini jauh dari kesempurnaan dan masih banyak ruang yang dapat diperbaiki di masa mendatang. Maka dari itu, penulis menerima kritik dan saran yang membangun dalam peningkatan penelitian dan skripsi ini.

Medan, 16 Oktober 2024

Penulis,



Adinda Khairani

201402019

ABSTRAK

Malaria adalah penyakit yang mengancam jiwa yang dapat ditemukan di negara-negara tropis. Malaria disebabkan oleh gigitan nyamuk *Anopheles* betina yang terinfeksi. Malaria banyak ditemukan di negara tropis, termasuk Indonesia. Di Sumatera Utara, beberapa daerah seperti Asahan, Batubara, dan Labuhanbatu Utara masih berstatus endemis sedang untuk malaria. Studi ini bertujuan untuk memprediksi jumlah kasus malaria di Labuhanbatu Utara menggunakan metode *Gaussian Process Regression* (GPR) sebagai tindakan penanggulangan perkembangan jumlah kasus malaria. Hasil penelitian ini diharapkan dapat memprediksi jumlah kasus malaria untuk menanggulangi perkembangan malaria, memperkirakan lebih awal periode yang berisiko mengalami lonjakan kasus, dan menjadi referensi bagi penelitian selanjutnya dalam memprediksi kejadian dengan menggunakan metode GPR. Data yang digunakan meliputi waktu mingguan, suhu maksimum, durasi sinar matahari, indeks jumlah hujan, indeks curah hujan, indeks tingkat banjir dan genangan air, indeks kepadatan dan pertambahan penduduk, serta total kasus malaria. Dalam mendapatkan jumlah data yang lebih banyak untuk meningkatkan kinerja model, penelitian ini menggunakan metode *Time-Series Generative Adversial Network* (TGAN). Penelitian ini menunjukkan bahwa metode GPR dengan *kernel* kombinasi *Matern* dan *Dot Product* mendapatkan hasil prediksi yang cukup akurat terhadap data kasus malaria. Didapatkan bahwa model yang dilatih dan diuji dengan menggunakan data sintesis mampu memberikan prediksi yang relatif akurat dan dekat dengan nilai aktual dalam skala yang lebih kecil. Ditunjukkan dengan nilai metrik evaluasi MSE sebesar 6.52, RMSE sebesar 2.55 dan MAE sebesar 1.71. Sedangkan data yang dilatih dan diuji dengan data asli memberikan nilai metrik evaluasi MSE sebesar 8.65, RMSE sebesar 2.94 dan MAE sebesar 2.28. Hal ini menunjukkan bahwa GPR lebih cocok digunakan dengan jumlah data yang lebih besar sehingga dapat meningkatkan akurasi prediksi.

Kata Kunci: Malaria, Prediksi, Jumlah Kasus, *Gaussian Process Regression* (GPR), *Kernel*, Evaluasi.

**PREDICTION OF THE NUMBER OF MALARIA CASES IN NORTH
LABUHANBATU USING THE GAUSSIAN PROCESS
REGRESSION METHOD**

ABSTRACT

Malaria is a life-threatening disease that can be found in tropical countries. Malaria is caused by the bite of an infected female Anopheles mosquito. Malaria is widely found in tropical countries, including Indonesia. In North Sumatra, some areas such as Asahan, Batubara, and Labuhanbatu Utara still have moderate endemic status for malaria. This study aims to predict the number of malaria cases in Labuhanbatu Utara using the Gaussian Process Regression (GPR) method as a countermeasure to the development of the number of malaria cases. The results of this study are expected to be able to predict the number of malaria cases to mitigate the development of malaria, estimate earlier the period at risk of experiencing a spike in cases, and become a reference for further research in predicting events using the GPR method. The data used includes weekly time, maximum temperature, sunshine duration, rainfall index, rainfall index, flood and inundation index, population density and growth index, and total malaria cases. In order to obtain a larger amount of data to improve model performance, this study used the Time-Series Generative Adversarial Network (TGAN) method. This study shows that the GPR method with a combination of Matern and Dot Product kernels gets fairly accurate prediction results on malaria case data. It was found that models trained and tested using synthesized data were able to provide relatively accurate predictions and close to actual values on a smaller scale. This is shown by the MSE evaluation metric value of 6.52, RMSE of 2.55 and MAE of 1.71. While the data trained and tested with the original data gives an evaluation metric value of MSE of 8.65, RMSE of 2.94 and MAE of 2.28. This shows that GPR is more suitable for use with a larger amount of data so that it can improve prediction accuracy.

Keywords: *Malaria, Prediction, Total Cases, Gaussian Process Regression (GPR), Kernel, Evaluation.*

DAFTAR ISI

PERSETUJUAN	iii
PERNYATAAN	iv
UCAPAN TERIMA KASIH	v
ABSTRAK	vii
<i>ABSTRACT</i>	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	3
1.5 Manfaat Penelitian	4
1.6 Metodologi Penelitian	4
1.7 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	7
2.1 Malaria	7
2.2 Jenis Malaria	7
2.2.1. <i>Plasmodium Falciparum</i>	7
2.2.2. <i>Plasmodium Vivax</i>	8
2.2.3. <i>Plasmodium Malariae</i>	8
2.2.4. <i>Plasmodium Ovale</i>	8
2.3 Iklim di Kabupaten Labuhanbatu Utara	9
2.4 Hubungan Faktor Iklim dengan Penyakit Malaria	10
2.5 <i>Time-Series Generative Adversial Network (TGAN)</i>	11
2.6 Regresi Nonparametrik	13
2.7 <i>Gaussian Process (GP)</i>	14
2.8 <i>Gaussian Process Regression (GPR)</i>	14
2.8.1. <i>Radial Basis Function (RBF)</i>	16
2.8.2. <i>Matern Kernel</i>	17

2.8.3.	<i>Rational Quadratic</i>	17
2.8.4.	<i>ExpSine Square</i>	18
2.8.5.	<i>White Kernel</i>	18
2.8.6.	<i>Dot Product Kernel</i>	19
2.9	<i>Grid Search</i>	19
2.10	<i>K-Fold Cross Validation (CV)</i>	19
2.11	Matriks Evaluasi	20
2.11.1	<i>Mean Square Error (MSE)</i>	21
2.11.2	<i>Root Mean Square Error (RMSE)</i>	21
2.11.3	<i>Mean Absolute Error (MAE)</i>	21
2.12	Penelitian Terdahulu	21
2.13	Perbedaan Penelitian	30
BAB 3 ANALISIS DAN PERANCANGAN SISTEM		7
3.1	Data	7
3.2	Analisis Sistem	7
3.2.1	<i>Data Collection</i>	34
3.2.2	<i>Data Pre-processing</i>	35
3.2.3	<i>Correlation Heatmap</i>	36
3.2.4	Implementasi <i>Time-Series Generative Adversial Network (TGAN)</i>	38
3.2.5	<i>Data Splitting</i>	39
3.2.6	Proses <i>Training</i>	39
3.2.7	Proses <i>Testing</i>	41
3.2.8	Proses Prediksi	41
3.3	Perancangan Antarmuka Sistem	41
3.3.1	Rancangan Tampilan Beranda	41
3.3.2	Rancangan Tampilan <i>Training</i>	42
3.3.3	Rancangan Tampilan <i>Testing</i>	43
3.3.4	Rancangan Tampilan Prediksi	44
BAB 4 IMPLEMENTASI DAN PENGUJIAN SISTEM		34
4.1.	Implementasi Sistem	34
4.1.1	Perangkat Keras dan Perangkat Lunak	34
4.1.2	Implementasi Data	34
4.1.3	Implementasi Tampilan Antarmuka	34

4.2. Pelatihan Model	52
4.3. Pengujian Model	59
4.4. Diskusi	67
BAB 5 KESIMPULAN DAN SARAN	47
5.1 Kesimpulan	47
5.2 Saran	47
DAFTAR PUSTAKA	71

DAFTAR TABEL

Tabel 2. 1 Tabel Penelitian Terdahulu	26
Tabel 3. 1 Dataset Penelitian Sebelumnya (Sumber: Fahmi <i>et al.</i> , 2022)	7
Tabel 3. 2 Kategori <i>Pearson Correlation Coefficient</i> (Sumber: Dahliani <i>et al.</i> , 2020)	37
Tabel 4. 1 Implementasi Model Training Pada <i>Kernel</i> Tunggal	52
Tabel 4. 2 Implementasi Model <i>Training</i> Pada <i>Kernel</i> Kombinasi RBF dengan <i>Kernel</i> Lainnya	53
Tabel 4. 3 Implementasi Model Training Pada <i>Kernel</i> Kombinasi <i>Matern</i> dengan <i>Kernel</i> Lainnya	55
Tabel 4. 4 Implementasi Model Training Pada <i>Kernel</i> Kombinasi <i>Rational Quadratic</i> dengan <i>Kernel</i> Lainnya	56
Tabel 4. 5 Implementasi Model Training Pada <i>Kernel</i> Kombinasi <i>ExpSine Square</i> dengan <i>Kernel</i> Lainnya	57
Tabel 4. 6 Hasil Terbaik dari Kombinasi <i>Kernel</i>	58
Tabel 4. 7 Hasil <i>K-Fold Cross Validation</i>	59
Tabel 4. 8 Hasil Pengujian Model dengan Data Sintesis	60
Tabel 4. 9 Hasil Perbedaan Nilai Prediksi dengan Nilai Aktual	62
Tabel 4. 10 Hasil Pengujian dengan Data Asli	65

DAFTAR GAMBAR

Gambar 2. 1 Peta Kabupaten Labuhanbatu Utara	10
Gambar 2. 2 Ilustrasi Cara Kerja TGAN (Sumber: Naveed <i>et al</i>)	12
Gambar 2. 3 Ilustrasi <i>Gaussian Process</i> (Sumber: Ding, 2018)	14
Gambar 2. 4 Arsitektur <i>Gaussian Process Regression</i> (GPR)	15
Gambar 2. 5 Ilustrasi Cara Kerja <i>K-Fold Cross Validation</i> (Sumber: Eric, 2023)	20
Gambar 3. 1 Arsitektur Umum Penelitian	34
Gambar 3. 2 Korelasi Antar Variabel Pada Dataset	37
Gambar 3. 3 Rancangan Tampilan Beranda	42
Gambar 3. 4 Rancangan Tampilan <i>Training</i>	42
Gambar 3. 5 Rancangan Tampilan Hasil <i>Training</i>	43
Gambar 3. 6 Rancangan Tampilan <i>Testing</i>	43
Gambar 3. 7 Rancangan Tampilan Hasil <i>Testing</i>	44
Gambar 3. 8 Rancangan Tampilan Prediksi	45
Gambar 3. 9 Rancangan Tampilan Hasil Prediksi	45
Gambar 4. 1 Tampilan Halaman Beranda	47
Gambar 4. 2 Tampilan Halaman <i>Training</i>	48
Gambar 4. 3 Tampilan Tabel Hasil <i>Training</i>	48
Gambar 4. 4 Tampilan Grafik Hasil Training	49
Gambar 4. 5 Tampilan Halaman <i>Testing</i>	49
Gambar 4. 6 Tampilan Tabel Hasil <i>Testing</i>	50
Gambar 4. 7 Tampilan Tabel Hasil <i>Testing</i>	50
Gambar 4. 8 Tampilan Grafik Hasil Testing	50
Gambar 4. 9 Tampilan Halaman Prediksi	51
Gambar 4. 10 Tampilan Tabel Hasil Prediksi	51
Gambar 4. 11 Tampilan Grafik Hasil Prediksi	52
Gambar 4. 12 Grafik Hasil Prediksi dengan Data Sintesis	62
Gambar 4. 13 Grafik Hasil Prediksi dengan Data Asli	66

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Malaria adalah penyakit yang mengancam jiwa yang dapat ditemukan di negara-negara tropis. Malaria disebabkan oleh gigitan nyamuk *Anopheles* betina yang terinfeksi. Nyamuk *Anopheles* betina yang terinfeksi menggigit dan memakan darah manusia. Selama proses makan, nyamuk *Anopheles* betina menularkan parasit *Plasmodium* (Health Organization, 2018). Tren peningkatan kasus malaria terus meningkat meskipun dengan laju yang lebih lambat (Roser, 2022). Pada tahun 2021, kasus malaria diperkirakan mencapai 247 juta kasus, dibandingkan dengan tahun 2020 dengan total 245 juta kasus dan 232 juta kasus pada tahun 2019, sedangkan jumlah kematian yang diakibatkan oleh malaria mengalami penurunan dari 625.000 pada tahun 2020 menjadi 619.000 pada tahun 2021 (Health Organization, 2022). Berdasarkan data tersebut, menunjukkan bahwa pencegahan malaria belum mengalami kemajuan yang signifikan untuk mencapai target *Sustainable Development Goals* (SDGs) 3.3, yakni penyakit menular. Pada poin 3.3 SDGs, target global pada kejadian malaria adalah mengurangi kasus dan jumlah kematian yang diakibatkan oleh malaria setidaknya 90% pada tahun 2030 (Health Organization, 2022).

Berdasarkan (Health Organization, 2022) Asia Tenggara masih menduduki peringkat kedua dalam penyebaran kasus malaria tertinggi dan Indonesia menjadi negara kedua setelah India yang menyumbangkan kasus malaria positif sebesar 811.636 kasus. Di Indonesia kasus malaria juga mengalami peningkatan. Berdasarkan data Kementerian Kesehatan (Kemenkes) pada tahun 2022 terdapat 415.140 kasus malaria. Angka tersebut meningkat 32,29% dibandingkan tahun 2021 sebanyak 304.607 kasus (Widi, 2022).

Dinas Kesehatan Sumatera Utara mencatat sepanjang tahun 2022, terdapat 5.133 kasus malaria di Sumatera Utara. Pada September 2022, tiga daerah di Sumatera Utara berstatus endemis sedang. Tiga daerah tersebut adalah Asahan, Batubara, dan Labuhanbatu Utara. Pada awal tahun 2023, Kementerian Kesehatan mengeliminasi lima provinsi yang dinyatakan bebas malaria. Lima provinsi tersebut adalah DKI Jakarta, Jawa Timur, Bali, Banten, dan Jawa Barat. Sumatera Utara masih belum dapat tereliminasi dalam penyebaran malaria dikarenakan masih terdapat 3 daerah dengan

kategori API (*Annual Parasite Incidence*) endemis sedang. API merupakan tren kasus positif malaria dengan jumlah penderita malaria perseribu penduduk dalam satu tahun. API dengan endemis rendah yakni jika API kurang dari 1 per 1000 penduduk. API dengan endemis sedang yaitu jika nilai API 1-5 per 1000 penduduk, sedangkan API dengan endemis tinggi yakni jika API lebih dari 5 per 1000 penduduk. Jika API kurang dari 1 per 1000 penduduk, maka suatu daerah dapat dieliminasi dari penyebaran penyakit malaria (Kementerian Kesehatan, 2022).

Pengaruh faktor iklim seperti suhu, curah hujan, dan kelembaban relatif mempengaruhi proses morfologi nyamuk. Suhu lingkungan dapat menyebabkan perkembangan parasit malaria yang lebih cepat sehingga menyebabkan insiden malaria yang lebih tinggi (Fischer et al., 2020). Namun, pengaruh faktor iklim terhadap kejadian malaria bervariasi dari satu negara ke negara lain dengan cara yang berbeda. Suhu memiliki hubungan linier yang tinggi dengan variabilitas malaria di keenam negara yang sudah diteliti. Tidak hanya suhu, curah hujan dan radiasi permukaan juga menunjukkan pengaruh terhadap variabilitas malaria (Nkiruka et al., 2021).

Pada tahun 2021, terdapat penelitian yang dilakukan untuk membandingkan metode pembelajaran mesin dalam memprediksi malaria dengan menggunakan dataset dari WHO, Kaggle, dan atmosfer curah hujan dan suhu dari portal iklim dunia. Parameter yang digunakan ialah persentase penduduk yang menggunakan sanitasi dasar, persentase penduduk yang menggunakan air minum standar, suhu rata-rata, curah hujan rata-rata, jumlah total kasus malaria yang dilaporkan setiap tahun dan jumlah insiden malaria (populasi berisiko per 1.000). Metode-metode yang dipakai pada penelitian ini adalah *Support Vector Machine* (SVM), *K-Nearest Neighbors* (KNN), *Random Forest*, *Decision Tree*, *Naive Bayes*, dan *Logistic Regression*. Dari keenam metode tersebut, *Random Forest* memiliki tingkat performa akurasi yang paling tinggi sebesar 97,72% dan nilai AUC sebesar 98% (Adamu & Singh, 2021).

Selain itu, terdapat penelitian yang dilakukan untuk memprediksi penularan malaria menggunakan metode pembelajaran mesin yaitu *Long Short Term Memory* (LSTM). Penelitian ini menggunakan dua kategori dataset, yaitu dataset klinis dan dataset lingkungan. Dataset klinis didapatkan dari Rural Health Centers (RHC) dan dataset lingkungan dikumpulkan dari satelit MODIS dan TRMM. Ketika suhu 28 dan kelembaban 60% kondisi ini menguntungkan untuk berkembangbiakan vektor nyamuk

Anopheles. Jika suhu di atas 30 dan di bawah 16, maka akan berdampak negatif terhadap perkembangbiakan nyamuk *Anopheles* (Santosh et al., 2020).

Adapun metode yang diusulkan dalam penelitian ini adalah *Gaussian Process Regression* (GPR). GPR merupakan model dengan pendekatan pembelajaran mesin klasik yang mengandalkan teori *Bayesian* dan teori pembelajaran statistik (Asante-Okyere et al., 2018). Dengan mengganti fungsi basis yang digunakan dalam identifikasi liner *Bayesian*, GPR dapat mengatasi masalah regresi yang rumit dengan ukuran sampel yang kecil dan dimensi yang tinggi (Taki et al., 2016). Dibandingkan dengan model peramalan tradisional, GPR memiliki keunggulan dalam penerapannya yang mudah, memperoleh *hyperparameter* secara adaptif, dan menghasilkan keluaran yang bersifat probabilitas (Sun et al., 2014). Oleh karena itu, penulis mengusulkan untuk melakukan penelitian dengan judul **Prediksi Jumlah Kasus Malaria di Labuhanbatu Utara Menggunakan Metode *Gaussian Process Regression***.

1.2 Rumusan Masalah

Berdasarkan data yang dikeluarkan Kementerian Kesehatan pada awal tahun 2023, Sumatera Utara masih belum dapat dieliminasi sebagai provinsi yang bebas malaria. Masih terdapat 11 daerah dengan kategori API endemis rendah, dan tiga daerah dengan endemis sedang. Labuhanbatu Utara adalah daerah yang berstatus endemis sedang dalam penyebaran malaria. Sebagai usaha untuk menurunkan jumlah kasus positif malaria di Labuhanbatu Utara, dibutuhkan suatu tindakan untuk menanggulangi perkembangan jumlah kasus positif malaria dengan segera. Tindakan tersebut adalah memprediksi jumlah kasus positif malaria untuk beberapa minggu mendatang guna mengidentifikasi terjadinya wabah.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah untuk memprediksi penyebaran jumlah kasus malaria di Labuhanbatu Utara dengan menggunakan metode *Gaussian Process Regression*.

1.4 Batasan Masalah

Agar penelitian terarah dan terfokus pada ruang lingkup permasalahan, maka ditentukan batasan masalah dalam penelitian ini. Adapun batasan masalah dalam penelitian ini sebagai berikut:

1. Wilayah penelitian ini adalah kabupaten Labuhanbatu Utara.
2. Data yang digunakan pada penelitian ini berformat .csv.
3. Parameter yang digunakan pada penelitian ini adalah suhu maksimum, durasi sinar matahari, *index* jumlah hujan, *index* curah hujan, *index* tingkat banjir dan genangan air, *index* kepadatan dan pertambahan penduduk dan total kasus malaria.

1.5 Manfaat Penelitian

Manfaat penelitian ini adalah sebagai berikut:

1. Membantu dalam menanggulangi perkembangan jumlah kasus positif malaria di Labuhanbatu Utara.
2. Memperkirakan lebih awal periode yang berisiko mengalami lonjakan kasus positif malaria.
3. Menjadi rujukan penelitian berikutnya dalam memprediksi suatu kejadian dengan menggunakan metode *Gaussian Proses Regression*.

1.6 Metodologi Penelitian

Adapun tahapan-tahapan yang dilakukan dalam penelitian ini adalah sebagai berikut:

1. Studi Literatur

Studi literatur dilakukan dengan mengumpulkan sumber referensi baik dari buku, jurnal, artikel, *report* publik, skripsi dan sumber-sumber lain terkait malaria dan metode *Gaussian Process Regression*.

2. Analisis Permasalahan

Tahap ini melakukan analisis yang didapatkan dari berbagai sumber referensi yang telah dikumpulkan sebelumnya untuk mendapatkan pemahaman lebih lanjut mengenai malaria dan metode *Gaussian Process Regression*.

3. Perancangan Sistem

Berdasarkan hasil dari studi literatur dan analisis permasalahan yang telah dilakukan, dibuatlah rancangan sistem berupa arsitektur umum untuk memprediksi jumlah kasus malaria.

4. Implementasi

Setelah sistem dibuat, dievaluasi dan diuji, sistem akan diimplementasikan untuk mendapatkan tujuan dan keluaran yang diharapkan. Penulis akan mengaplikasikan metode *Gaussian Process Regression* dengan menggunakan bahasa pemrograman *Python*.

5. Pengujian

Tahap ini melakukan pengujian sistem yang sudah dibuat terhadap hasil yang diperoleh dalam pengimplementasian GPR untuk menyelesaikan masalah yang diteliti.

6. Dokumentasi dan Penyusunan Laporan

Pada tahap ini, penulis akan melakukan dokumentasi dan penyusunan laporan yang berisikan proses penelitian dari awal hingga mendapatkan hasil akhir.

1.7 Sistematika Penulisan

Sistematika penulisan pada skripsi ini terdiri dari beberapa bagian, yaitu:

Bab 1: Pendahuluan

Bagian pendahuluan berisi uraian mengenai motivasi dan faktor yang menjadi penyebab dilakukannya penelitian ini. Setiap uraian dibagi menjadi beberapa bagian tambahan seperti latar belakang pelaksanaan penelitian, perumusan masalah sebagai landasan penelitian, tujuan penelitian, batasan masalah yang mengatur cakupan masalah yang diteliti, manfaat penelitian, metodologi dilakukannya penelitian dan sistematika penelitian sebagai tata cara penyusunan tulisan.

Bab 2: Landasan Teori

Bagian landasan teori memuat uraian terkait konsep atau teori yang digunakan pada analisis permasalahan dalam penelitian ini, yaitu malaria dan jenis-jenisnya, kondisi iklim di Sumatera Utara, hubungan antara faktor iklim dengan penyebaran malaria, kasus malaria impor, teknik-teknik yang akan digunakan dalam megolah data seperti *Time Series Generative Adversial Networks* (TGAN), konsep dasar metode GPR yaitu regresi nonparametrik, *Gaussian Process* (GP), *Gaussian Process Regression* (GPR), *Grid Search Cross Validation*, *K-Fold Cross Validation*, metode evaluasi untuk

mengukur kinerja model dan penelitian terdahulu yang relevan dengan masalah dan metode yang sama dengan penelitian ini serta perbedaan penelitian yang dilakukan dengan penelitian terdahulu.

Bab 3: Analisis dan Perancangan Sistem

Bagian analisis dan perancangan sistem memuat uraian terkait penggunaan data, arsitektur umum, proses *preprocessing*, pembangunan sistem dan sistem rancangan desain antarmuka.

Bab 4: Implementasi dan Pengujian Sistem

Bagian ini merupakan dokumentasi pengimplementasian dari tahapan analisis dan perancangan sistem. Selain itu, juga mencakup perangkat keras dan lunak yang diunakan untuk pengujian sistem serta evaluasi dari tahapan pengujian.

Bab 5: Kesimpulan dan Saran

Bagian kesimpulan dan saran memuat kesimpulan dari temuan penelitian yang telah dilakukan dan saran untuk penelitian berikutnya berdasarkan temuan kekurangan yang terdapat pada penelitian ini.

BAB 2

LANDASAN TEORI

2.1 Malaria

Penyakit malaria ditularkan ke manusia melalui gigitan nyamuk *Anopheles* betina yang terinfeksi *Plasmodium* yang ditularkan melalui parasit protozoa (Fischer et al., 2020). Nyamuk *Anopheles* betina memakan darah manusia untuk memproduksi telurnya. Agar nyamuk *Anopheles* terinfeksi *Plasmodium*, mereka harus menggigit atau menghisap darah dari penderita malaria, kemudian nyamuk yang sudah terinfeksi *Plasmodium* akan menyuntikan parasit melalui air liurnya dengan gigitan ke orang berikutnya dan siklus infeksi terus berlanjut (Direktorat Jenderal Pencegahan dan Pengendalian Penyakit, 2022). Setelah nyamuk *Anopheles* menggigit manusia, ada jangka waktu sebelum gejala-gejala pertama muncul. Masa antara gigitan nyamuk dengan gejala pertama muncul disebut dengan masa inkubasi. Masa inkubasi secara umum berkisar antara 7 hingga 30 hari. Berbagai *spesies* parasit penyebab malaria dapat menyebabkan masa inkubasi lebih pendek atau lebih panjang (Global Health, 2023).

Diketahui lebih dari 422 *spesies* nyamuk *Anopheles* hidup di dunia, tetapi yang dapat menjadi vektor penyakit malaria hanya 60 *spesies*. Di Indonesia, ditemukan sekitar 80 *spesies* nyamuk *Anopheles*, namun hanya 22 *spesies* yang berperan sebagai vektor malaria (Arsin, 2011). Terdapat empat *spesies* parasit di Indonesia yang secara rutin menginfeksi manusia yaitu, *Plasmodium falciparum*, *Plasmodium vivax*, *Plasmodium malariae*, dan *Plasmodium ovale* (Elyazar et al., 2011).

2.2 Jenis Malaria

Jenis-jenis malaria dipengaruhi oleh parasit yang ditularkan melalui gigitan nyamuk *Anopheles*. Berikut merupakan jenis-jenis malaria:

2.2.1. *Plasmodium Falciparum*

Plasmodium falciparum merupakan penyebab dari malaria tropika atau malaria *falciparum*. Malaria dengan jenis ini merupakan malaria berat dan berbahaya karena berkembang biak dengan cepat di dalam darah manusia (AK Githeko, EN ototo, 2012). Masa inkubasi *Plasmodium falciparum* sekitar 12 hari dengan gejala nyeri kepala, demam iregular, hingga dapat menimbulkan gagal ginjal (Apriliani, 2021).

2.2.2. *Plasmodium Vivax*

Plasmodium vivax menyebabkan malaria tertiana yang serangannya timbul berselang setiap tiga hari (Khayriah, 2020). Malaria dengan jenis ini telah menyebar hampir ke seluruh pulau di Indonesia dan merupakan malaria yang paling banyak di temukan di Indonesia dan secara klinis jauh lebih ringan dan jarang menimbulkan kematian dibandingkan dengan *Plasmodium falciparum* (Avichena & Anggriyani, 2023). Malaria tertiana memiliki distribusi geografis yang luas, mulai dari wilayah beriklim dingin, subtropik, hingga daerah tropik seperti Asia Tenggara. Masa inkubasi *Plasmodium vivax* adalah 12 – 17 hari dengan salah satu gejalanya adalah pembengkakan limpa dan splenomegali (Apriliani, 2021).

2.2.3. *Plasmodium Malariae*

Plasmodium malariae menyebabkan malaria quartana yang menyebabkan serangan demam setiap empat hari sekali. Malaria dengan jenis ini dapat terjadi di dataran rendah maupun dataran tinggi di daerah tropis (Avichena & Anggriyani, 2023). Biasanya malaria quartana berlangsung tanpa gejala dan ditemukan secara tidak sengaja, namun malaria jenis ini sering mengalami kekambuhan (Apriliani, 2021). Masa inkubasi *Plasmodium malariae* berkisar antara 18-40 hari, ini merupakan masa inkubasi paling lama diantara jenis malaria lainnya. Pada infeksi *Plasmodium malariae*, siklus reproduksi aseksual (pembelahan diri dalam tubuh manusia atau hewan) terjadi dalam waktu 72 jam yang mana setiap sel-sel membelah akan terjadi serangan demam (Putra, 2011).

2.2.4. *Plasmodium Ovale*

Plasmodium ovale menyebabkan malaria *ovale* dengan masa inkubasi 12-17 hari disertai gejala demam setiap 48 jam. Malaria *ovale* merupakan malaria yang relatif ringan dan dapat sembuh dengan sendirinya (Irwan, 2016). *Plasmodium ovale* sering dijumpai di Afrika yang beriklim tropik dan Pasifik Barat (Arsin, 2011). Sedangkan di Indonesia, *Plasmodium ovale* sejauh ini baru dilaporkan pernah terjadi di dua provinsi, yaitu Papua dan Nusa Tenggara Timur (Elyazar et al., 2011).

2.3 Iklim di Kabupaten Labuhanbatu Utara

Kabupaten Labuhanbatu Utara adalah salah satu kabupaten yang terletak di daerah Pantai Timur Provinsi Sumatera Utara dengan ibukota Aekkanopan. Labuhanbatu Utara secara geografis terletak pada 1058' 2050' 05'25' 100 Lintang Utara, 99 Bujur Timur yang memiliki ketinggian 0 – 700 meter di atas permukaan laut. Luas kabupaten Labuhanbatu Utara sebesar 354.580 Ha. Labuhanbatu Utara memiliki 8 kecamatan dengan 90 desa atau kelurahan.

Labuhanbatu Utara memiliki kondisi alam atau topografi dengan sebagian besar berupa perbukitan sempit dan terjal serta dataran rendah. Labuhanbatu Utara merupakan Daerah Aliran Sungai (DAS) yang terdiri dari Sungai Kualuh dan beberapa anak Sungai Bilah yang termasuk dalam Satuan Wilayah Sungai (SWS) Barumun atau Kualuh. Sungai ini mengalir dari Selatan ke Utara lalu bermuara ke Selat Malaka. Kondisi Sungai ini relatif baik dengan air yang cukup jernih. Musim kemarau dan musim hujan memengaruhi fluktuasi debit yang mana jika pada musim kemarau debit air mengecil sedangkan pada musim hujan debit air meningkat. Sungai ini berfungsi sebagai sumber air minum bagi penduduk setempat.

Labuhanbatu Utara memiliki rata-rata curah hujan bulanan sebesar 280.67 mm dengan 13.25 hari hujan. Bulan November merupakan rata-rata curah hujan tertinggi sebesar 428 mm dan bulan Februari curah hujan terendah sebesar 67 mm. Wilayah Labuhanbatu Utara termasuk ke dalam zona agroklimat D1 yang merupakan daerah dengan bulan basah (memiliki curah hujan bulanan > 200 mm) berurutan dengan 3 – 4 bulan dan bulan kering (memiliki curah hujan bulanan < 100 mm) berurutan selama < 2 bulan.

Kondisi iklim Labuhanbatu Utara secara umum beriklim tropis basah dengan ciri dua pertukaran angin yang disebabkan oleh angin Moonson Barat yang berhembus dari arah Utara (Asia Tenggara) dan melewati Selat Malaka yang membuatnya menjadi basah oleh kandungan air. Hal ini menyebabkan Labuhanbatu Utara bermusim hujan pada bulan April – September. Gambar 2.1 menunjukkan peta kabupaten Labuhanbatu Utara.



Gambar 2. 1 Peta Kabupaten Labuhanbatu Utara

2.4 Hubungan Faktor Iklim dengan Penyakit Malaria

Iklim merupakan faktor penting bagi berbagai jenis penyakit, salah satu pengaruh perubahan iklim terhadap potensi peningkatan kejadian timbulnya penyakit yang ditularkan oleh nyamuk seperti penyakit malaria dan demam berdarah. Perkembangbiakan nyamuk membutuhkan genangan air dan nyamuk dewasa membutuhkan tempat yang lembab agar dapat bertahan hidup. Penyakit yang disebabkan oleh nyamuk dan paling sensitif terhadap perubahan iklim jangka panjang adalah penyakit malaria (Achmadi, 2007).

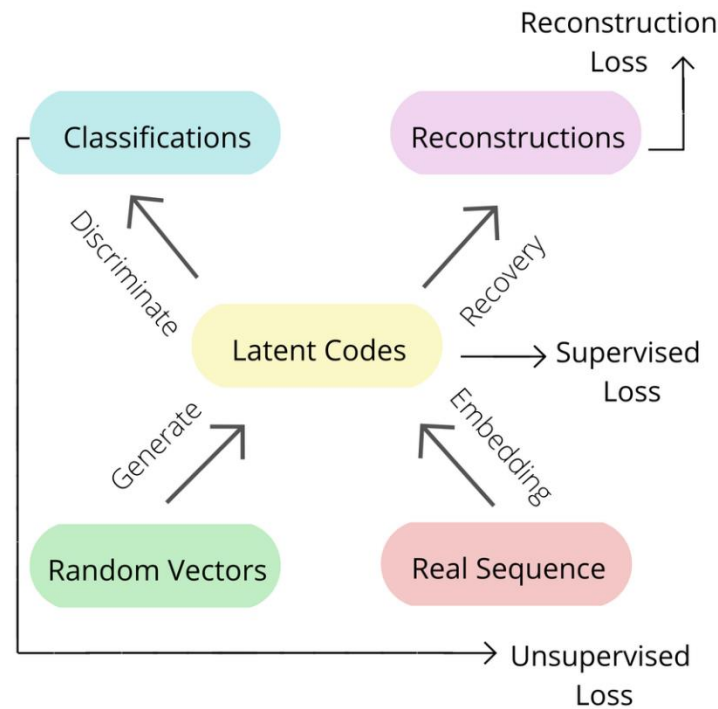
Peningkatan populasi dan aktivitas manusia akan menentukan laju perubahan iklim. Suhu merupakan faktor penentu yang penting dalam penularan penyakit malaria, meskipun curah hujan sangat penting dalam menyediakan habitat yang sesuai bagi nyamuk untuk berkembang biak, suhu menjadi pendorong utama dalam hidup nyamuk (Blanford et al., 2013). Suhu dapat menyebabkan perkembangan parasit malaria menjadi lebih cepat yang membuat penyebaran penyakit malaria lebih tinggi (Fischer et al., 2020). Selain suhu sebagai faktor penting dalam penyebaran penyakit malaria, faktor iklim seperti curah hujan dan kelembapan juga memberikan efek pada ekologi nyamuk dalam penyebaran malaria (Makinde et al., 2021).

Curah hujan mendorong pertumbuhan vegetasi yang menyediakan ruang berkembang biak bagi vektor (Sadoine et al., 2018). Curah hujan mempengaruhi keberadaan air di permukaan tanah, semakin tinggi intensitas hujan semakin besar potensi air sebagai genangan untuk habitat nyamuk. Hal ini menunjukkan, jika curah hujan tinggi maka populasi nyamuk penyebab malaria semakin meningkat (Sulasmi et al., 2019). Kelembapan udara juga berpengaruh terhadap penyebaran penyakit malaria. Jika kelembapan udara rendah, maka akan memperpendek umur nyamuk. Sebaliknya, pada kelembapan udara tinggi, nyamuk menjadi lebih aktif menggigit, sehingga akan meningkatkan penularan malaria (Arsin, 2011).

2.5 Time-Series Generative Adversial Network (TGAN)

Time-Series Generative Adversial Network (TGAN) merupakan model untuk menghasilkan data deret waktu sintesis dengan menggabungkan jaringan *adversial* generatif (GAN) dengan menggunakan *unsupervised* dan *supervised learning*. TGAN bertujuan untuk menangkap dinamika temporal dan hubungan dalam data secara efisien dengan melatih model dalam mode *loop* terbuka dan tertutup.

TGAN menghasilkan data sintesis yang mirip dengan data deret waktu asli dengan memperkenalkan konsep kerugian berjenjang (*stepwise supervised loss*) untuk memastikan model dapat menangkap distribusi data dalam setiap langkah waktu secara lebih akurat. TGAN menggunakan jaringan *embedding* yang dapat memetakan fitur asli ke representasi laten dan sebaliknya. Hal ini membantu mengurangi kompleksitas (dimensional tinggi) dari ruang pembelajaran model. Kerugian yang diawasi diminimalkan dengan melatih secara bersama-sama jaringan *embedding* dan jaringan generator, yang membantu generator untuk memahami dan mempelajari hubungan temporal dalam data. Pendekatan ini memungkinkan TGAN untuk menciptakan data deret waktu yang realistis dengan mempertahankan dinamika waktu (urutan waktu) dalam data. TGAN juga dapat menangani situasi di mana data yang dihasilkan mencakup data statis dan data deret waktu secara bersamaan (Yoon et al., n.d.). Gambar 2.2 mengilustrasikan cara kerja TGAN.



Gambar 2. 2 Ilustrasi Cara Kerja TGAN (Sumber: Naveed *et al*)

Pada Gambar 2.2, terdapat *random vectors* yang berfungsi sebagai *input* awal untuk memulai proses generasi data sintesis. *Latent codes* merupakan representasi tersembunyi yang dihasilkan dari *random vectors* atau dari data deret waktu asli, ini digunakan untuk menghasilkan dan merekonstruksi data. *Real sequence* merupakan data deret waktu asli yang digunakan sebagai referensi untuk melatih model. *Reconstructions* merupakan data yang dihasilkan kembali dari *latent codes* yang mendekati data asli dan *classifications* merupakan klasifikasi dari data yang dihasilkan untuk menentukan apakah data tersebut nyata atau sintesis.

Proses dimulai dari tahap *embedding*, data deret waktu asli (*real sequence*) dipetakan ke ruang laten (*latent codes*) melalui jaringan *embedding*. Ini adalah proses tanpa pengawasan (*unsupervised*) karena tidak memerlukan label atau supervisi tambahan. *Random vectors* akan digenerate untuk digunakan dalam menghasilkan *latent codes* yang kemudian akan digunakan untuk menghasilkan data sintesis. Selanjutnya tahap *recovery*, kode latent akan diubah kembali menjadi data deret waktu yang direkonstruksi (*reconstructions*). Proses ini diawasi (*supervised*) karena menggunakan data asli sebagai acuan untuk melatih model agar dapat merekonstruksi data yang mirip dengan data asli. Setelah itu, *supervised loss* dihitung untuk meminimalkan perbedaan antara data asli dan data yang direkonstruksi. Data yang

dihasilkan (baik dari vektor acak maupun dari rekonstruksi kode laten) diklasifikasikan untuk menentukan apakah data tersebut asli atau sintetis. Lalu, *reconstruction loss* dihitung untuk mengukur seberapa baik data yang dihasilkan mirip dengan data asli. Ini membantu dalam memperbaiki model untuk menghasilkan data yang lebih realistis.

2.6 Regresi Nonparametrik

Regresi digunakan untuk menganalisis pola hubungan antara satu atau lebih variabel, yaitu variabel respon dan variabel prediktor. Tujuan utama dalam melakukan regresi adalah mencari bentuk estimasi kurva regresi yang tepat. Regresi nonparametrik merupakan salah satu model regresi yang digunakan untuk mengetahui adanya pola hubungan antara variabel respon dengan variabel prediktor yang tidak diketahui bentuk kurva regresinya atau tidak terdapat informasi distribusi mengenai bentuk pola. Oleh karena itu, regresi nonparametrik memiliki fleksibilitas dalam mencari sendiri bentuk kurva regresinya. (Putri, 2018). Secara umum, model regresi nonparametrik dapat dituliskan pada persamaan berikut:

$$y_i = f(x_i) + \varepsilon_i, i = 1, 2, \dots, n \quad (1)$$

Dimana:

y_i : variabel respon

$f(x_i)$: nilai dari fungsi regresi yang tidak diketahui

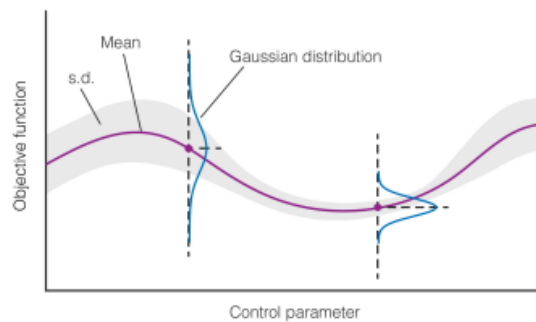
x_i : variabel prediktor

ε_i : error random yang diasumsikan identik, independen, dan berdistribusi normal dengan mean bernilai nol dan varian σ^2 .

Beberapa model regresi nonparametrik yang sering digunakan antara lain Histogram, *Kernel*, *Spline*, Polinomial Lokal, Deret Ortogonal, Deret *Fourier*, *K-Nearest Neighbor* (KNN), *Neural Network*, *Wavelets*, MARS, dan lainnya. Semua model memiliki kelebihan dan kekurangan serta motivasi tersendiri dalam memodelkan pola (Islamiyati, 2009).

2.7 Gaussian Process (GP)

Gaussian Process adalah model pembelajaran mesin probabilistik yang biasanya digunakan untuk tugas regresi atau klasifikasi yang diawasi. GP memiliki beberapa kelebihan seperti cocok digunakan untuk memprediksi model dengan *dataset* yang kecil dan berkorelasi, serta memberikan kuantifikasi ketidakpastian yang sangat penting dalam pengambilan keputusan (Tazi et al., 2023). GP menggunakan pendekatan probabilistik untuk mempelajari model nonparametrik yang dimana nonlinieritas diimplementasikan melalui *kernel*. GP menjadi metode yang populer untuk pemodelan non-parametrik, terutama untuk data *time series* (Cheng et al., 2019).



Gambar 2. 3 Ilustrasi *Gaussian Process* (Sumber: Ding, 2018)

GP merupakan model statistik yang menggeneralisasi Gaussian multivariat ke domain kontinu yang secara konseptual menentukan distribusi *over function*. Setiap titik data yang diamati dikaitkan dengan variabel acak yang terdistribusi secara normal. Garis ungu pada gambar di atas mewakili fungsi rata-rata posterior dan area yang diarsir mewakili standar deviasi. GP direpresentasikan dalam persamaan berikut:

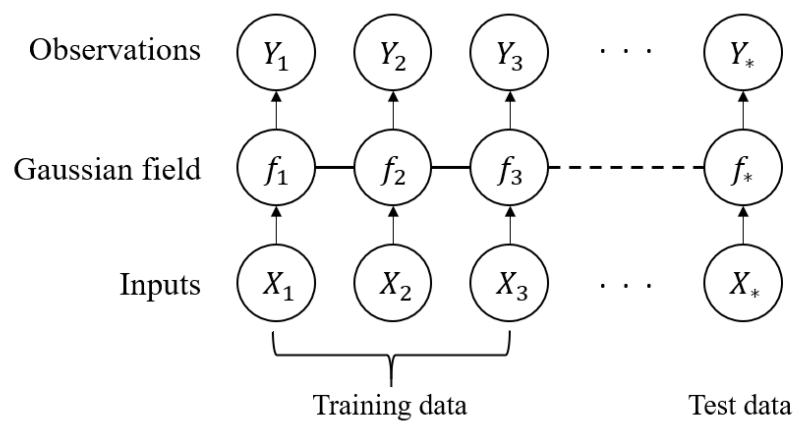
$$f(x) \sim GP(m(x), k(x, x')) \quad (2)$$

Dimana GP adalah *Gaussian Process*, $m(x)$ adalah fungsi rata-rata, dan $k(x, x')$ adalah fungsi kovarian.

2.8 Gaussian Process Regression (GPR)

Salah satu penerapan dari GP untuk melakukan regresi melalui *supervised learning* adalah GPR. GPR menggunakan konsep *Bayesian kernel*, dimana parameter *kernel*

ditentukan oleh pilihan fungsi kovarian atau *kernel*, serta data yang digunakan untuk membuat prediksi (Santosh, Ramesh, and Reddy 2020). GPR digunakan untuk menangani masalah regresi yang kompleks, seperti dimensi tinggi, ukuran sampel kecil, dan non-linieritas, serta memiliki kemampuan generalisasi yang kuat. Dibandingkan dengan *Neural Network* dan *Support Vector Machine*, GPR memiliki banyak keunggulan, yaitu realisasi yang mudah, akuisisi *hyperparameter* yang dapat beradaptasi dengan sendirinya, inferensi yang fleksibel terhadap non-parameter, dan signifikansi probabilitas dari *outputnya* (Zhao et al., 2019).



Gambar 2. 4 Arsitektur *Gaussian Process Regression* (GPR)
(Sumber: Oh & Lee, 2020)

Gambar 2.4 menggambarkan proses umum dari GPR. Variabel laten f_i diturunkan dari nilai *input* X_i dengan nilai observasi Y_i . Distribusi dari nilai observasi uji Y diestimasi menggunakan *Gaussian* f untuk nilai *input* X . Dataset untuk pelatihan digambarkan dengan dimana persamaan (2) dan (3) meringkas model GPR secara umum:

$$Y = f + \epsilon \quad \text{dimana } \epsilon \sim N(0, \sigma^2 I) \quad (3)$$

Dimana:

ϵ : parameter *noise* distribusi Gaussian

σ_y^2 : varian

I : matriks identitas yang dibangun berdasarkan dimensi data

f : nilai berbasis relasi transformasi untuk vektor *input* ekuivalen X
 y : *output* yang diamati vektor

Dalam GPR kita mengasumsikan GP sebelumnya yang ditentukan dengan menggunakan fungsi *mean* dan fungsi kovarian:

$$m(x) = E[f(x)] \quad (4)$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))] \quad (5)$$

$m(x)$ mewakili nilai yang diharapkan dari fungsi f pada titik matriks *input* x . $k(x, x')$ adalah ukuran tingkat kepercayaan untuk $m(x)$. Dengan fungsi-fungsi ini ditentukan, $f(x)$ dapat diambil sampelnya dari GP dengan persamaan 2 di atas.

$$f(x) \sim GP(m(x), k(x, x')) \quad (2)$$

Fungsi kovarian $k(x, x')$ diparameterkan oleh parameter *kernel* yang menggambarkan ketergantungan antara titik *input* yang diamati x dan titik *input* baru untuk memprediksi x . GPR memiliki beberapa fungsi kovarian seperti berikut:

2.8.1. Radial Basis Function (RBF)

RBF merupakan fungsi *kernel* yang sering digunakan. RBF juga sering dikenal dengan *quadratic exponential kernel*, *the squared exponential kernel*, dan *Gaussian kernel*. Fungsi *kernel* RBF dalam GPR adalah untuk menentukan kemiripan atau korelasi antara dua titik data, yang pada gilirannya membantu dalam memprediksi nilai di lokasi baru berdasarkan nilai-nilai yang sudah diketahui. Adapun persamaan matematis dari RBF adalah:

$$k(x, x') = \exp \left(- \frac{\|x - x'\|}{2l^2} \right) \quad (6)$$

Dimana l adalah panjang skala *kernel* yang mengontrol seberapa cepat korelasi antara titik-titik data menurun seiring bertambahnya jarak di antara mereka. *Kernel* RBF mampu menangkap pola non-linier dalam data karena fungsi eksponensialnya, yang membuatnya sangat fleksibel dalam menangkap berbagai bentuk dan variasi dalam data.

2.8.2. *Matern Kernel*

Matern kernel adalah jenis *kernel* yang digunakan dalam *Gaussian Process Regression* (GPR) yang merupakan keluarga *kernel* parametrik dengan fleksibilitas yang lebih besar dibandingkan dengan *kernel Radial Basis Function* (RBF). *Kernel* ini digunakan untuk mengukur kesamaan atau korelasi antara dua titik data dalam ruang *input*, dan berguna untuk berbagai aplikasi regresi karena dapat menangkap pola yang lebih kompleks dan memiliki kontrol yang lebih halus atas kelancaran fungsi yang diprediksi. Berikut fungsi *kernel Matern* secara matematis dapat dilihat dari persamaan 7.

$$k(x, x') = \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu} \|x - x'\|}{l} \right) K_\nu \left(\frac{\sqrt{2\nu} \|x - x'\|}{l} \right) \quad (7)$$

Dimana l merupakan panjang skala yang mengontrol seberapa cepat korelasi antara titik-titik data menurun seiring bertambahnya jarak di antara mereka, ν adalah parameter kehalusan yang mengontrol kelancaran dari fungsi yang diprediksi, $\Gamma(\nu)$ adalah fungsi Gamma dan K_ν adalah fungsi Bessel termodifikasi dari orde kedua jenis ketiga.

2.8.3. *Rational Quadratic*

Kernel Rational Quadratic dapat diartikan sebagai jumlah tak terbatas dari *kernel* RBF. *Kernel Rational Quadratic* akan menghasilkan prior yang cukup mulus pada fungsi sampel proses *Gaussian*. *Kernel* ini memberikan fleksibilitas yang lebih besar dalam menangkap variasi pada berbagai skala dalam data, membuatnya sangat berguna untuk menangani data dengan struktur yang kompleks. Persamaan 8 mendefinisikan *kernel Rational Quadratic* dalam bentuk matematis.

$$k(x, x') = \sigma_f^2 \left(1 + \frac{\|x - x'\|^2}{2\alpha l^2} \right)^{-\alpha} \quad (8)$$

Dimana l adalah skala panjang *kernel* yang mengontrol seberapa cepat korelasi antara titik-titik data menurun seiring bertambahnya jarak di antara mereka dan α yang merupakan parameter bentuk yang mengontrol relatif pentingnya variasi pada berbagai skala panjang.

2.8.4. *ExpSine Square*

ExpSine Square yang merupakan kependekan dari *Exponential Sine Squared Kernel* adalah jenis *kernel* yang digunakan dalam *Gaussian Process Regression* (GPR) untuk menangkap pola yang bersifat periodik dalam data. *Kernel* ini sangat berguna ketika data menunjukkan pola berulang dalam interval tertentu, seperti data musiman atau siklus. Berikut persamaan 9 mendefinisikan *ExpSine Square kernel* dalam bentuk matematis.

$$k(x, x') = \exp \left(-2 \sin^2 \left(\frac{\pi |x - x'|}{\rho} \right) / l^2 \right) \quad (9)$$

Dimana ρ adalah periode yang menunjukkan interval di mana pola berulang, l panjang skala yang mengontrol seberapa cepat korelasi menurun seiring bertambahnya jarak di antara titik-titik data dan $k(x, x')$ adalah jarak *Euclidean* antara dua titik x dan x' .

2.8.5. *White Kernel*

White kernel juga dikenal sebagai *White Noise Kernel* adalah jenis *kernel* yang digunakan dalam *Gaussian Process Regression* (GPR) untuk menangani komponen kebisingan dalam data. *Kernel* ini membantu memodelkan variabilitas atau ketidakpastian yang ada dalam pengukuran data yang mungkin disebabkan oleh faktor acak atau kebisingan instrumen. Persamaan 10 mendefinisikan *white kernel* dalam bentuk matematis.

$$k(x, x') = \sigma^2 I_n \quad (10)$$

Dimana σ^2 adalah varian dari jenis kebisingan dan I_n adalah matriks identitas.

2.8.6. Dot Product Kernel

Dot Product Kernel juga dikenal sebagai *Linier Kernel*, adalah salah satu *kernel* paling sederhana dan paling dasar yang digunakan dalam metode berbasis *kernel* seperti *Gaussian Process Regression* (GPR). *Kernel* ini memodelkan hubungan linier antara data dan cocok untuk dataset di mana hubungan antar fitur adalah linier. Persamaan 11 mendefinisikan *Dot Product kernel* dalam bentuk matematis.

$$k(x, x') = x \cdot x' = \sum_{i=1} x_i x'_i \quad (11)$$

Dimana x dan x' adalah vektor fitur dari dua titik data, \cdot adalah operasi *Dot Product* dan x_i dan x'_i adalah elemen ke- i dari vektor fitur x dan x' .

2.9 Grid Search

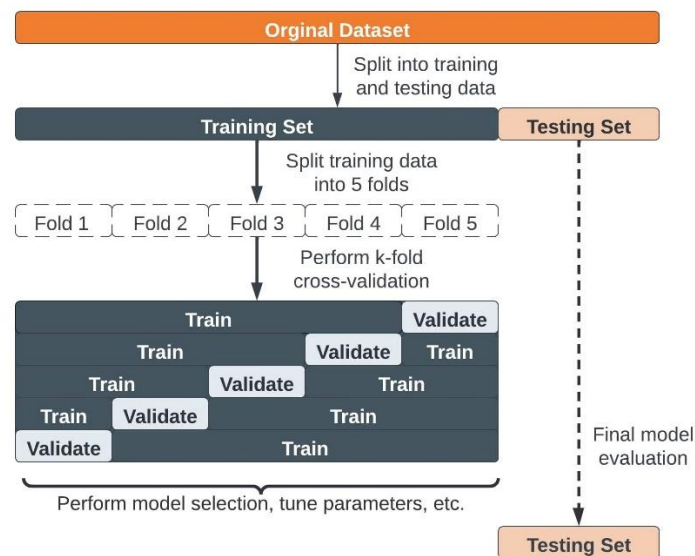
Grid search merupakan metode yang dilakukan untuk mencari kombinasi optimal dari *hyperparameter* dengan cara melakukan pencarian secara sistematis pada *grid* atau kisi parameter yang telah ditentukan sebelumnya (Nugraha & Sasongko, 2022). Dengan menggunakan *grid search*, model dapat mengidentifikasi kombinasi *hyperparameter* yang paling sesuai dengan data, sehingga meningkatkan kemampuan model untuk menggeneralisasi dan memberikan prediksi yang akurat. Setelah melakukan kombinasi *hyperparameter* dalam *grid* dapat dievaluasi dengan menggunakan *K-Fold Cross Validation*.

2.10 K-Fold Cross Validation (CV)

K-Fold Cross Validation (CV) adalah salah satu teknik validasi silang yang umum digunakan dalam pembelajaran mesin untuk mengevaluasi kinerja model secara objektif. Teknik ini membagi data menjadi K *subset* atau *fold* yang mana K adalah bilangan bulat yang ditentukan sebelumnya. *K-Fold Cross Validation* dapat secara

efektif memperkirakan kesalahan prediksi dengan menghubungkan kesalahan pelatihan dan dapat mengurangi bias ke bawah dalam kesalahan pelatihan.

K-Fold Cross Validation dapat menggunakan data *training* dan data validasi. Data *training* dibagi menjadi *K fold*, misalnya dibagi menjadi 5 *fold*, namun data validasi tidak terlibat dalam proses *K-Fold Cross Validation*. Karena tujuan utamanya adalah mengevaluasi model pada data yang belum pernah dilihat sebelumnya. Oleh karena itu, data validasi biasanya dipisahkan dari awal dan tidak terpengaruh oleh *K-Fold Cross Validation*. Pada iterasi pertama, *fold 1* menggunakan data validasi sementara *fold 2 -5* menggunakan data *training*. Iterasi kedua, *fold 2* menggunakan data validasi sementara *fold 1, 3, 4, 5* menggunakan data *training* dan seterusnya. Setelah iterasi selesai, hasil validasi dari setiap *fold* akan dihitung menggunakan metrik evaluasi untuk memberikan gambaran kinerja model secara keseluruhan.



Gambar 2. 5 Ilustrasi Cara Kerja *K-Fold Cross Validation* (Sumber: Eric, 2023)

2.11 Matriks Evaluasi

Model yang sudah dilatih menggunakan GPR akan dievaluasi untuk memastikan model yang dibuat dapat memberikan prediksi yang akurat. Evaluasi model menggunakan tiga matriks evaluasi, yaitu *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE), dan *Mean Absolute Error* (MAE).

2.11.1 Mean Square Error (MSE)

MSE memberikan gambaran tentang seberapa baik model dalam menggambarkan variabilitas data aktual dengan menghitung rata-rata dari kuadrat selisih antara nilai prediksi (\hat{y}_i) dengan nilai aktual (y_i). Semakin rendah nilai MSE, semakin baik model dapat memprediksi data aktual. Persamaan 12 mendefinisikan MSE dalam bentuk matematis.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (12)$$

2.11.2 Root Mean Square Error (RMSE)

RMSE mengukur seberapa baik model regresi yang dibuat dalam memprediksi nilai sebenarnya dari suatu variabel berkelanjutan. Semakin kecil nilai RMSE, maka hasil prediksi akan semakin akurat. RMSE menghitung rata-rata dari selisih kuadrat antara nilai prediksi (\hat{y}_i) dan nilai aktual (y_i) kemudian diambil akar kuadratnya. Secara sistematis, RMSE didefinisikan dengan persamaan 13.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (13)$$

2.11.3 Mean Absolute Error (MAE)

MAE mengukur rata-rata dari selisih absolut antara nilai prediksi (\hat{y}_i) dan nilai observasi sebenarnya (y_i). Semakin kecil nilai MAE, semakin baik kualitas model yang dibuat. MAE menggambarkan rata-rata dari kesalahan absolut tanpa memperhitungkan arah kesalahan. MAE dapat dihitung dengan menggunakan persamaan 14.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (14)$$

2.12 Penelitian Terdahulu

Prediksi kejadian malaria dengan menggunakan data kejadian malaria dan data meteorologi sudah banyak dilakukan sebelumnya. Santosh et al., 2020 melakukan prediksi malaria dengan menggunakan *Long Short Term Memory* (LSTM). Penelitian

ini menggunakan data klinis dan data lingkungan. Data klinis dikumpulkan dari keempat lokasi geografis (Awaraopeta, Khammam, Aswapuram, dan Venkatapuram), dari pusat-pusat kesehatan pedesaan (RHC) yang didirikan oleh pemerintah Telangana dengan variabel jumlah rata-rata bulanan kasus positif, kasus tanpa gejala, usia rata-rata, laki-laki, dan perempuan. Data dikumpulkan dari bulan Menunjukkan 1995 – Desember 2018. Data lingkungan dikumpulkan dari satelit MODIS dan TRMM dengan variabel curah hujan, suhu, kelembapan relatif, dan indeks vegetasi. Hasil dari penelitian ini adalah Venkatapuram dan Awaraopeta mencatat jumlah kasus tertinggi. Di lokasi tersebut curah hujan rata-rata adalah 1,6 mm dan suhu rata-rata di siang dan malam hari sedikit berbeda 3°C dibandingkan dengan lokasi lainnya.

Prediksi malaria dengan membandingkan metode *machine learning* dilakukan oleh (Yadav et al. 2021). Metode yang digunakan adalah *Naïve Bayes*, *Linier Regression*, *Decission Tree*, *Random Forest*, *Support Vector Machine* (SVM), dan *Artificial Neural Network* (ANN). Dataset pertama yang digunakan berisi catatan medis pasien yang tinggal di tempat yang berbeda di Senegal. Dataset kedua berisi catatan medis pasien yang tinggal di area yang sama di Diourbel, Thies, dan Fatick. Dalam mengevaluasi model, penelitian ini menggunakan lima matriks yaitu *precision*, *recall*, *F-measure*, *Specificity*, dan *Receiver Operating Characteristic* (ROC). Dari perbandingan keenam metode tersebut, metode RF, SVM dengan *gaussian kernel*, dan ANN memberikan akurasi tertinggi untuk memprediksi malaria dengan *precision*, *re-call*, dan *F-measure* setidaknya sama dengan 92%, 85%, dan 89%.

Penelitian dengan membandingkan metode prediksi juga dilakukan di Burundi dengan menggunakan *General Linier Model* (GLM) dan *Artificial Neural Network* (ANN) (Mfisimana et al. 2022). Penelitian ini menggunakan data malaria yang dikumpulkan dari berbagai provinsi di Burundi dari 2010-2017. Data tersebut terdiri dari beberapa variabel yaitu total kasus malaria, kasus anak balita dan ibu hamil. Data iklim bulanan dengan variabel ketinggian, curah hujan, suhu, dan kelembapan juga digunakan untuk memprediksi malaria. Selain itu, penelitian ini juga menggunakan data demografi dengan variabel kepadatan penduduk, tingkat pendidikan, jumlah ibu hamil, jumlah anak balita, dan jumlah jumlah kelambu berinsektisida (ITN) yang diberikan. Hasil dari penelitian ini menunjukkan bahwa metode ANN memiliki kinerja yang lebih baik dalam memprediksi total kasus malaria dibandingkan dengan GLM. Dibuktikan dengan data yang diberikan berjumlah 7,82 juta, kemudian dengan menggunakan

metode ANN total prediksi malaria bulanan berjumlah 7,81 juta sedangkan metode GLM berjumlah 7,48 juta (Mfisimana *et al.*, 2022).

Prediksi malaria dengan menggunakan metode time series dilakukan oleh (Wang *et al.* 2019). Penelitian ini menggunakan metode time series tradisional ARIMA dan STL+ARIMA. Selain itu, penelitian ini juga menggunakan *deep learning* dengan metode *Back-Propagation Neural Network* (BP-NN) dan LSTM. Penggunaan *ensemble learning* dengan metode *Stacking* juga digunakan di penelitian ini. Dataset yang digunakan merupakan laporan bulanan insiden malaria di China dari tahun 2007-2017 yang diperoleh dari *National Scientific Data Sharing Platform for Population and Health*. Penelitian ini juga menggunakan dataset meteorologi yang diperoleh dari *National Meteorological Information Center* yang terdiri dari variabel suhu, kelembaban relatif, tekanan udara, tekanan uap air, tingkat kelembaban, kecepatan angin, curah hujan, durasi penyinaran matahari, dan hari dengan curah hujan harian $\geq 0,1$ mm. Matriks yang digunakan untuk mengevaluasi prediksi dari model yang digunakan adalah *Root Mean Square Error* (RMSE), *Mean Absolute Scaled Error* (MASE), dan *Mean Absolute Deviation* (MAD). Metode *Stacking* menunjukkan kinerja yang baik daripada model deret waktu tradisional ARIMA dan STL+ARIMA, serta metode *deep learning* BP-ANN dan LSTM.

(Nkiruka, Prasad, and Clement 2021) melakukan penelitian dimulai dengan melakukan rekayasa fitur untuk mengidentifikasi faktor iklim yang mempengaruhi kejadian malaria. Lalu diikuti dengan proses pengelompokan *K-Means Clustering* dalam membagi dataset ke dalam beberapa cluster yang berhubungan serta membersihkan data dan mendeteksi *outlier*. Metode klasifikasi yang digunakan dalam penelitian ini adalah XGBoost. Data klinis yang digunakan merupakan data dari enam negara di wilayah Afrika Sub-Sahara (Burkina Faso, Mali, Republik Niger, Nigeria, Kamerun, dan Republik Demokratik Dongo) didapatkan dari WHO data repository. Data klinis berisi nilai normalisasi dari insiden malaria yang dikonfirmasi tahunan per 1000 penduduk. Data iklim didapatkan dari *National Centre for Atmospheric Research* (NCAR) berdasarkan pengamatan selama 28 tahun (1990-2017) dengan beberapa variabel yaitu tekanan atmosfer, suhu permukaan, curah hujan, radiasi permukaan dan kelembapan relatif. Hasil penelitian menunjukkan bahwa hubungan antara kejadian malaria dan variabilitas iklim bervariasi dari satu wilayah dengan wilayah lainnya, variabel curah hujan, suhu, dan radiasi permukaan secara signifikan berkontribusi

terhadap wabah malaria. Penelitian ini juga memvalidasi kinerja model yang dilakukan dengan membandingkan dengan model klasifikasi ML lainnya seperti *Naïve Bayes*, *Support Vector Machine* (SVM), dan *Logistic Regression* (LR). XGBoost masih menghasilkan akurasi tertinggi dengan rata-rata nilai 95,8% di enam negara yang dipilih jika dibandingkan dengan model lainnya.

Penelitian prediksi malaria pada Kabupaten Batubara yang dilakukan oleh (Vicalina, 2023) menggunakan metode *Gated Recurrent Unit* (GRU) dengan memanfaatkan data meteorologi. Data meteorologi dikumpulkan dari *website* resmi Badan Meteorologi, Klimatologi dan Geofisika (BMKG). Penelitian ini juga mensintesis data dengan menggunakan metode *Conditional Tabular Generative Adversarial Network* (CTGAN). CTGAN membantu dalam menghasilkan kualitas data yang cukup baik dalam proses pembentukan model untuk memprediksi kejadian malaria. Hasil dari penelitian ini menunjukkan metode GRU dapat melakukan prediksi kejadian malaria dengan nilai MAE sebesar 6,03 dan RMSE sebesar 9,99.

(Vidya and Hari 2023) melakukan penelitian untuk memprediksi lalu lintas penumpang bus agar penjadwalan bus dan manajemen lalu lintas menjadi tepat yang berdampak pada peningkatan efisiensi dan keamanan penumpang. Penelitian ini menggunakan satu set rute bus Pune Mahanagar Parivahan Mahamandal Limited (PMPL). Fungsi *kernel* yang digunakan pada penelitian ini adalah fungsi *kernel* kuadrat eksponensial yang dapat menghasilkan prior yang mulus pada fungsi yang diambil sampelnya dari proses *Gaussian*. Validasi kinerja model menggunakan RMSE, estimasi kepadatan *kernel*, dan distribusi tingkat energi. Selain itu, penelitian ini juga membandingkan kinerja GPR dengan model *Student-t* dan KRR. RMSE dari ketiga model yang dicoba menunjukkan bahwa GPR memiliki nilai paling rendah yaitu 5,33, sedangkan *Student-t* sebesar 13,36 dan KRR sebesar 24,61. Hal ini menunjukkan bahwa GPR memiliki kinerja yang lebih baik karena kesalahannya berkurang sebesar 8 kali lipat dibandingkan dua metode lainnya. GPR hanya membutuhkan waktu 170 detik untuk menyelesaikan proses pelatihan dan prediksi, menunjukkan bahwa GPR bekerja dengan baik dalam akurasi prediksi bahkan dengan data yang terbatas.

Penelitian dengan menggunakan GPR juga dilakukan oleh (Zhao et al., 2019) dalam memprediksi kandungan sulfat pada danau-danau di Tiongkok. Data yang digunakan pada penelitian ini adalah data pemantauan kualitas air di bagian Tengah dan hilir Sungai Yangtze dari tahun 2007-2009. Sebanyak 515 sampel pemantauan kualitas air

dipilih sebagai sampel percobaan. Suhu air, transparansi, Ph, oksigen terlarut, konduktivitas, klorofil, fosfor total, nitrogen total, dan nitrogen amonia digunakan sebagai *input* model serta kandungan sulfat pada air sebagai *output* model. Penilaian kinerja model menggunakan MAPE, MAE, dan MSE. Sedangkan untuk fungsi *kernel* menggunakan eksponensial, *mattern 5/2*, *square* eksponensial, dan *Rational Quadratic*. Adapun hasil dari pengujian terhadap empat fungsi *kernel* tersebut adalah metode GPR dengan fungsi *kernel* eksponensial adalah yang paling cocok digunakan untuk memprediksi kandungan sulfat di danau karena fungsi *kernel* eksponensial mendapatkan nilai MAE 5,0464 dan nilai RMSE 7,269 yang lebih rendah dibandingkan ketiga fungsi *kernel* lainnya.

Penerapan GPR dalam mitigasi gas hidrat diteliti oleh (Suresh et al., 2021) untuk meminimalkan pertumbuhan gas hidrat. Parameter *input* yang digunakan adalah konsentrasi dan tekanan gas hidrat karbon dioksida (CO₂) dan Metana (CH₄), sedangkan parameter *output*nya adalah rata-rata *depression temperature* (ADP). Pengujian kinerja model menggunakan *Coefficient of Determination*, R^2 , dan MSE. Selain itu, dalam membandingkan GPR untuk memprediksi *output* ADT penelitian ini juga menggunakan LSSVM dan ANN. Fungsi *kernel* yang digunakan adalah *Radial Basis Function* (RBF). Hasil dari penelitian ini adalah metode GPR memberikan hasil terbaik dalam hal R^2 yaitu 97,25% dan 96,71% dibandingkan dengan model ANN dan LSSVM. Nilai MSE sebesar 0.019 untuk data *training* dan 0.023 untuk data *testing* yang diperoleh merupakan yang terendah untuk model GPR yang mengindikasikan bahwa model ini memberikan kesalahan prediksi ADT yang paling rendah dibandingkan dengan dua model ML lainnya.

Penelitian lainnya menggunakan GPR untuk mengestimasi kondisi Kesehatan baterai Li-ion dilakukan oleh (Dong et al., 2022). Dataset yang digunakan pada penelitian ini adalah data baterai publik yang disediakan oleh NASA dan Oxford yang terdiri dari baterai silinder LG Chem 18650 dan baterai lithium. Penelitian ini menggabungkan fungsi *kernel* polinomial dan fungsi *kernel* SE. Pengujian kinerja model menggunakan MAE, MAPE, dan RMSE. Selain itu, metode SVM, ANN, LSTM, dan ELM dipilih untuk perbandingan dengan GPR. Hasilnya adalah metode GPR memiliki indeks error paling rendah dengan rata-rata nilai MAE, MAPE, dan RMSE masing-masing 0,61%, 0,62%, dan 0,73%.

Tabel 2. 1 Tabel Penelitian Terdahulu

No	Peneliti	Judul	Metode	Keterangan
1	Thakur Santosh, Dharavath Ramesh, Damodar Reddy	<i>LSTM Based Prediction of Malaria Abundances Using Big Data</i>	Long Short Term Memory (LSTM)	Penelitian ini menggunakan LSTM dalam memprediksi malaria dengan hasil lokasi Venkatapuram dan Awaraopeta mencatat jumlah kasus malaria tertinggi dibandingkan dua lokasi geografis lainnya.
2	Samir S. Yadav, Vinod J Kadam, Shivajirao M. Jadhav, Sagar Jagtap, Prasad R. Pathak	<i>Machine Learning Based Malaria Prediction using Clinical Findings</i>	Naïve Bayes, Linier Regression, Decission Tree, Random Forest, Support Vector Machine dan Artificial Neural Network	Penelitian membandingkan 6 metode pembelajaran mesin dalam memprediksi malaria dengan data klinis. Metode RF, SVM, dan ANN memberikan akurasi tertinggi yaitu 92%, 85%, dan 89%.
3	Lionel Divin Mfisimana, Emile Nibayisabe, Kingsley Badu, David Niyukuri	<i>Exploring Predictive Frameworks for Malaria in Burundi</i>	General Linier Model (GLM) dan Artificial Neural Network (ANN)	Penelitian membandingkan metode prediksi GLM dan ANN. Data yang diberikan pada penelitian ini berjumlah 7,81 juta dan metode ANN menunjukkan hasil yang baik dengan menunjukkan total prediksi malaria bulanan berjumlah 7,81 juta dibandingkan dengan GLM yang berjumlah 7,48 juta.

No	Peneliti	Judul	Metode	Keterangan
4	Mengyang Wang, Hui Wang, Jiao Wang, Hongwei Liu, Rui Lu, Tongqing Duan, Xiaowen Gong, Siyuan Feng, Yuanyuan Liu, Zhuang Cui, Changping Li, Jun Ma	<i>A Novel Model for Malaria Prediction Based on Ensemble Algorithms</i>	ARIMA, STL+ARIM, Back-Propagation Neural Network (BP-NN), Long Short Term Memory (LSTM), Ensamble Learning	Penelitian membandingkan metode time series dengan ensemble learning dalam memprediksi malaria. Hasil dari penelitian adalah metode Stacking menunjukkan hasil yang baik daripada metode lainnya dengan nilai RMSE 6,8, MASE dengan nilai 0,24, dan MAD dengan nilai 4,62.
5	Odu Nkiruka, Rajesh Prasad, Onime Clement	<i>Prediction of malaria incidence using climate variability and machine learning</i>	XGBoost, Naïve Bayes, Support Vector Machine (SVM), dan Logistic Regression (LR)	Penelitian bertujuan membandingkan metode XGBoost dengan metode pembelajaran mesin lainnya dalam memprediksi malaria dengan menggunakan data variabilitas iklim. XGBoost memperoleh akurasi 82% yang merupakan akurasi tertinggi dibandingkan dengan tiga metode lainnya.
6	Vicalina A	Prediksi Kejadian Malaria dengan Pemanfaatan Data Meteorologi	Gated Recurrent Unit (GRU)	Penelitian ini menggunakan data meteorologi yang disintesis dengan CTGAN. Hasil dari penelitian ini menunjukkan bahwa GRU dapat melakukan prediksi kejadian malaria dengan

No	Peneliti	Judul	Metode	Keterangan
		Menggunakan <i>Gated Recurrent Unit</i>		memanfaatkan data meteorologi dengan nilai MAE sebesar 6,03 dan RMSE sebesar 9,99.
7	Vidya, Hari	<i>Prediction of Bus Passenger Traffic using Gaussian Process Regression</i>	<i>Gaussian Process Regression (GPR), Student-t, dan Kernel Ridge Regression (KRR)</i>	Penelitian bertujuan untuk memprediksi lalu lintas penumpang bus agar penjadwalan bus dan manajemen lalu lintas menjadi tepat dan meningkatkan efisiensi serta keamanan penumpang. Dari ketiga metode yang digunakan, GPR memiliki nilai RMSE yang paling rendah yaitu 5,33, menunjukkan bahwa GPR memiliki kinerja yang lebih baik karena kesalahannya berkurang sebesar 8 kali lipat dibandingkan dua metode lainnya.
8	Jingying Zhao, Hai Guo, Min Han, Haoran Tang, Xiaoni Li	<i>Gaussian Process Regression for Prediction of Sulfate Content in Lakes Of China</i>	<i>Gaussian Process Regression (GPR)</i>	Penelitian dilakukan untuk memprediksi kandungan sulfat pada danau di Tiongkok. Penilaian kinerja model menggunakan MAPE, MAE, dan MSE dengan membandingkan fungsi <i>kernel</i> eksponensial, <i>mattern 5/2</i> , <i>square</i>

No	Peneliti	Judul	Metode	Keterangan
				<i>eksponensial</i> , dan <i>Rational Quadratic</i> . Hasilnya adalah metode GPR dengan fungsi <i>kernel</i> eksponensial memiliki nilai MAPE, MAE, dan MSE paling rendah dibandingkan dengan fungsi <i>kernel</i> lainnya.
9	Sachin Dev Duresh, Ali Qasim, Bhajan Lal, Syed Muhammad Imran, Khor Siak Foo	<i>Application of Gaussian Process Regression (GPR) in Gas Hydrate Mitigation</i>	<i>Gaussian Process Regression (GPR), Artificial Neural Network (ANN), Least Square Version of Support Vector Machine (LSSVM)</i>	Penelitian bertujuan untuk meminimalkan pertumbuhan gas hidrat. Pengujian kinerja model menggunakan <i>Coefficient of Determination</i> , R^2 , dan MSE. Penelitian juga membandingkan GPR dengan ANN dan LSSVM. Fungsi <i>kernel</i> yang digunakan adalah <i>Radial Basis Function</i> (RBF). Hasil dari penelitian ini adalah metode GPR memberikan hasil terbaik dalam hal R^2 dan <i>Coefficient of Determination</i> yaitu 97,25% dan 96,71% dibandingkan dengan model ANN dan LSSVM.

No	Peneliti	Judul	Metode	Keterangan
10	Hao Dong, Ling Mao1, Keqing Qu, Jinbin Zhao, Fen Li, Lei Jiang	<i>State of Health Estimation for Li-ion Batteries using Improved Gaussian Process Regression and Multiple Health Indicators</i>	<i>Gaussian Process Regression (GPR), Support Vector Machine (SVM), Artificial Neural Network (ANN), Long Short Term Memory (LSTM), Extreme Learning Machine (ELM).</i>	Penelitian bertujuan untuk mengestimasi kondisi kesehatan baterai Li-ion dengan menggabungkan fungsi <i>kernel</i> polinomial dan fungsi <i>kernel square eksopensial</i> . Pengujian kinerja model menggunakan MAE, MAPE, dan RMSE. Selain itu, metode GPR dibandingkan dengan metode pembelajaran mesin lainnya dengan hasil bahwa metode GPR memiliki nilai MAE, MAPE, dan RMSE dengan masing-masing nilai 0,61%, 0,62%, dan 0,73%. Hal ini menunjukkan bahwa metode GPR memiliki indeks eror paling rendah.

2.13 Perbedaan Penelitian

Prediksi kejadian malaria telah banyak dilakukan penelitian dengan menggunakan berbagai model dan data. Penelitian yang dilakukan oleh (Yadav et al., 2021) memanfaatkan beberapa metode *machine learning* dalam memprediksi apakah seseorang menderita malaria atau tidak dengan menggunakan dataset rekam medis pasien di Senegal dan dataset rekam medis pasien yang tinggal di wilayah yang sama. Penelitian ini tidak menggunakan data iklim atau lingkungan sebagai parameter lain dalam memprediksi kejadian malaria.

Perbedaan penelitian yang dilakukan oleh (Mfisimana et al., 2022) terletak pada metode yang digunakan. Penelitian yang dilakukan menggunakan metode *General Linier Model* (GLM) dan *Artificial Neural Network* (ANN) dengan objek penelitian provinsi Burundi. Dataset yang digunakan memiliki rentang waktu dari tahun 2010 – 2017 dengan variabel jumlah kasus malaria antara anak-anak berumur 5 tahun dan ibu hamil, serta *Insecticide Treated Nets* (ITN). Selain itu, penelitian ini juga menggunakan data iklim bulanan dan data demografi yang termasuk didalamnya perkiraan populasi tahunan.

Penelitian yang dilakukan oleh (Wang et al., 2019) menggunakan metodel *time series* tradisional dan *deep learning* dalam memprediksi kejadian malaria dengan data laporan kejadian malaria bulanan dari tahun 2007 – 2017 dan data meteorologi dengan variabel suhu, kelembapan relatif, tekanan udara, kecepatan angin, curah hujan dan durasi sinar matahari. Penelitian ini menggunakan lokasi Yunnan, China dalam memprediksi malaria.

Perbedaan penelitian yang dilakukan oleh (Nkiruka et al., 2021) adalah penelitian ini menggunakan variabel independen yang diwakili oleh variabel iklim dengan satu variabel dependen yang diwakili oleh peningkatan atau penurunan kasus kejadian malaria. Atribut data tersebut adalah curah hujan, radiasi permukaan, suhu, tekanan atmosfer dan kelembapan relatif. Hasil dari penelitian ini mengelompokkan kejadian malaria ke dalam kelompok kejadian tinggi dan rendah berdasarkan faktor iklim.

Penelitian yang dilakukan oleh (Vicalina, 2023) memiliki perbedaan penelitian. Penelitian ini dilakukan dengan memanfaatkan data meteorologi yang memiliki parameter rata-rata suhu udara, kelembapan rata-rata, curah hujan, lamanya penyinaran matahari dan kecepatan angin maksimum. Prediksi malaria pada penelitian ini dilakukan dengan menggunakan metode *Gated Recurrent Unit* (GRU). Hasil dari penelitian ini menunjukkan bahwa GRU dapat melakukan prediksi kejadian malaria dengan memanfaatkan data meteorologi dengan nilai MAE sebesar 6,03 dan RMSE sebesar 9,99.

Berdasarkan perbandingan yang telah diuraikan di atas, penelitian ini memiliki perbedaan dengan penelitian yang sudah dilakukan sebelumnya. Perbedaan penelitian tersebut adalah objek penelitian, batasan penelitian, data penelitian, dan metode penelitian. Penelitian ini menggunakan wilayah kejadian malaria di kabupaten Labuhanbatu Utara. Data yang digunakan akan disintesis menggunakan TGAN agar

model dapat bekerja dengan lebih baik. Data yang digunakan pada penelitian ini memiliki parameter suhu maksimum, durasi sinar matahari, *index* jumlah hujan, *index* curah hujan, *index* tingkat banjir dan genangan air, *index* kepadatan dan penambahan penduduk dan total kasus malaria.

BAB 3

ANALISIS DAN PERANCANGAN SISTEM

3.1 Data

Penelitian ini menggunakan sumber data dari penelitian sebelumnya (Fahmi et al., 2022) dengan judul “*Spatial analysis to evaluate risk of malaria in Northern Sumatera, Indonesia*”. Dataset ini berformat *Comma Separetd Value* (CSV) dengan jumlah data sebesar 210 data. Setiap data memiliki parameter sebagai berikut *weeks*, suhu maksimum (TR_max), durasi sinar matahari (DR_M), *index* jumlah hujan (CR), *index* curah hujan (CH), *index* tingkat banjir dan genangan air (F_I), *index* kepadatan dan pertambahan penduduk (PI_Den) dan total kasus positif malaria (*total_case*). Berikut Tabel 3.1 yang merupakan contoh dataset yang digunakan pada penelitian ini.

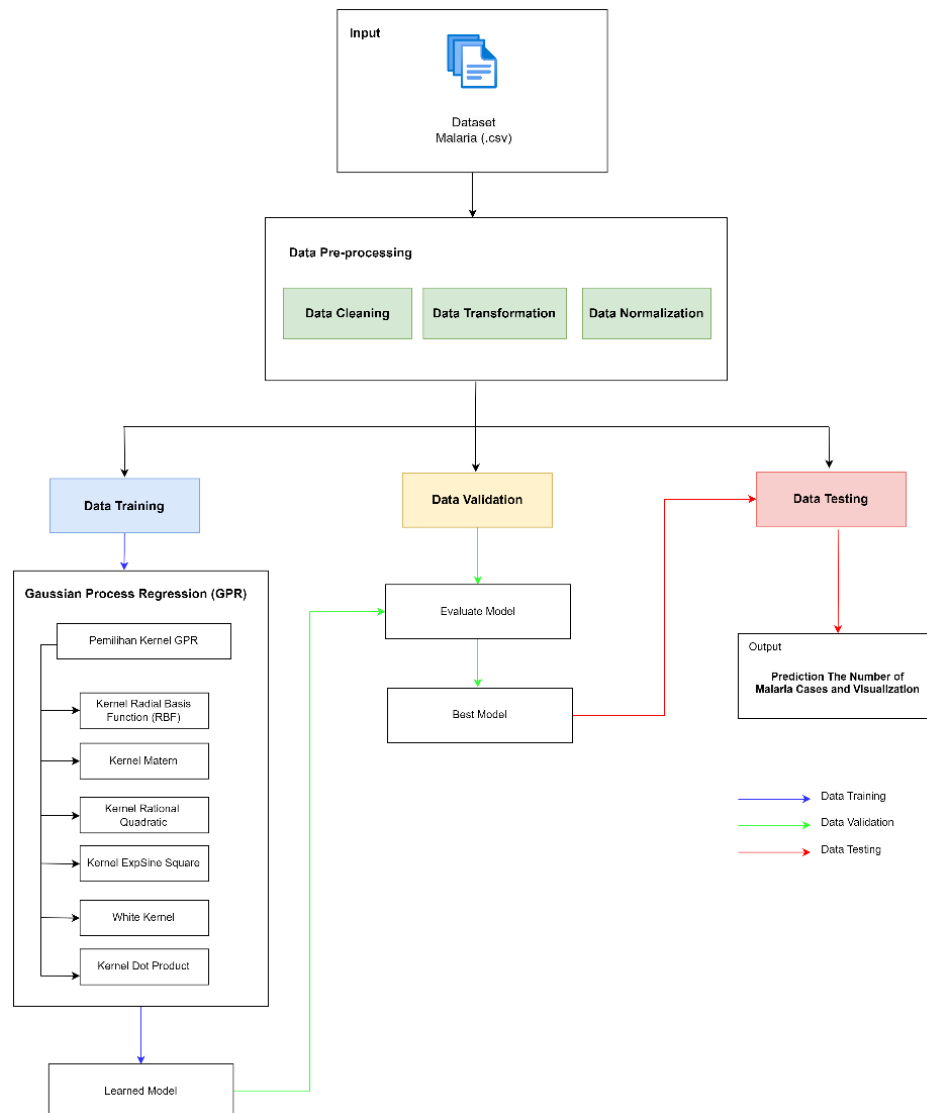
Tabel 3. 1 Dataset Penelitian Sebelumnya (Sumber: Fahmi *et al.*, 2022)

Weeks	TR_max	DR_M	CR	CH	F_I	PI_Den	Total case
31-12-2018	3.16	4.83	1.01	0.22	1.77	6.05	1
07-01-2019	3.31	6.53	1.51	0.25	1.66	6.05	2
14-01-2019	3.21	6.23	2.01	0.45	1.28	6.05	2
21-01-2019	3.32	4.25	2.01	0.34	1.09	6.05	1
28-01-2019	3.21	4.67	1.01	0.19	1.97	6.06	1
...
2-01-2023	3.02	3.86	1.51	3.68	4.40	9.24	16

3.2 Analisis Sistem

Tahapan untuk mengembangkan sistem prediksi dengan menggunakan metode *Gaussian Process Regression* (GPR) pada penelitian ini terdiri dari beberapa tahapan yang harus dilakukan. Tahap pertama dilakukan dengan mengumpulkan sumber data. Data yang dikumpulkan dari penelitian sebelumnya akan disintesis untuk mendapatkan jumlah data yang lebih banyak. Sintesis data dilakukan dengan menggunakan metode *Time-Series Generative Adversial Network* (TGAN) dari *python package y-data synthetic*. Kemudian data akan di *pre-processing* untuk pengoptimalan dataset saat digunakan. Setelah dataset selesai dilakukan *preprocessing*, data akan dibagi menjadi

tiga bagian yaitu data *training*, data *validation* dan data *testing*. Data *training* akan dimanfaatkan untuk melakukan proses pembelajaran dan pembentukan model menggunakan *Gaussian Process Regression* (GPR). Selanjutnya, model yang sudah dibuat akan dilakukan validasi dengan menggunakan data *validation* dan akan diuji dengan menggunakan data *testing*. Gambar 3.1 merupakan proses dari setiap tahapan.



Gambar 3. 1 Arsitektur Umum Penelitian

3.2.1 Data Collection

Tahap pertama dalam melakukan proses prediksi jumlah kasus malaria dilakukan dengan pengumpulan data. Data yang dikumpulkan berasal dari penelitian sebelumnya (Fahmi et al., 2022) terkait analisis risiko penyebaran penyakit malaria Dataset dari

penelitian sebelumnya berjumlah 210 data dengan rentang waktu dari 31-12-2018 sampai dengan 2-1-2023 yang selanjutnya akan digunakan untuk sintesis data. Sintesis data dilakukan dengan menggunakan metode *Time-Series Generative Adversial Network* (TGAN).

3.2.2 Data *Pre-processing*

Pada tahap ini dilakukan pra-pemrosesan data untuk memastikan bahwa data yang dipakai dalam model prediksi memiliki kualitas yang baik dan tidak mengandung kesalahan. Terdapat beberapa tahapan yang dilakukan pada *pre-processing*, yaitu:

3.2.2.1 Data *Cleaning*

Dataset yang dikumpulkan masih memiliki *missing value* pada beberapa parameter. Tahap ini akan melakukan penanganan *missing value* dengan menghitung nilai rata-rata dari nilai yang tersedia dalam suatu parameter dan kemudian akan menggunakan nilai rata-rata tersebut untuk mengisi nilai kosong dalam suatu parameter. Berikut *pseudocode* dalam menangani *missing value*.

```

FUNCTION fill_missing_with_mean(data):
  For each column in data:
    if column has missing values:
      calculate mean of non-missing values in the column
      replace missing values in the column with the calculated
      mean
  return data

```

3.2.2.2 Data *Transformation*

Pada dataset, parameter waktu berisi waktu dalam bentuk harian. Oleh karena itu, pada tahap pembersihan data, nilai pada parameter waktu akan diubah menjadi periode mingguan. Setelah itu, dibuat kolom baru untuk menghitung total kasus positif malaria dalam periode mingguan. Hal ini dilakukan karena penyebaran malaria dipengaruhi oleh faktor-faktor iklim yang menunjukkan pola mingguan yang lebih konsisten daripada harian. Selain itu, nyamuk *Anopheles* betina memiliki siklus hidup yang biasanya lebih dari seminggu. Data mingguan memungkinkan untuk lebih baik menyesuaikan analisis dengan siklus hidup nyamuk, sehingga hubungan antara kondisi iklim dan peningkatan kasus malaria bisa lebih jelas

terlihat. Sehingga dengan mengubah data menjadi mingguan, dapat menangkan perubahan iklim dan siklus hidup nyamuk *Anopheles* dengan lebih baik. Berikut *pseudocode* untuk mengubah waktu harian menjadi periode mingguan.

```

DEFINE datetime to weeks:
  FUNCTION datetime_to_weeks(datetime_column):
    Initialize an empty list weeks_column
    For each datetime in datetime_column:
      Calculate the week number from datetime
      Append week number to weeks_column
    Return weeks_column

```

3.2.2.3 Data Normalization

Setelah data dibersihkan, dilakukan normalisasi data. Normalisasi data dilakukan dengan menyesuaikan nilai-nilai pada data ke dalam skala yang sama untuk mencegah variabel dengan skala besar mendominasi model. Proses normalisasi membantu meningkatkan performa model dalam mengidentifikasi pola. Normalisasi yang dilakukan menggunakan *MinMaxScaler* yang akan mengubah atau "menskalakan" data sehingga nilai-nilai dalam data tersebut berada dalam rentang tertentu, yaitu antara 0 dan 1.

```

DEFINE MinMaxScaler:
  FUNCTION min_max_scaler(data_column):
    Initialize min_value as the minimum value of data_column
    Initialize max_value as the maximum value of data_column
    Initialize an empty list scaled_column
    For each value in data_column:
      Calculate scaled_value using the formula:
        
$$\text{scaled\_value} = (\text{value} - \text{min\_value}) / (\text{max\_value} - \text{min\_value})$$

      Append scaled_value to scaled_column
    Return scaled_column

```

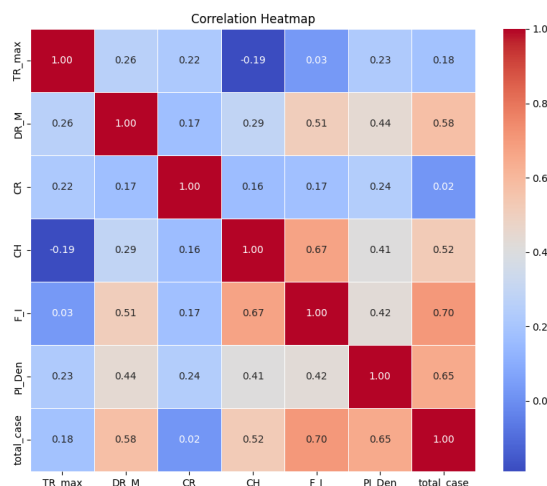
3.2.3 Correlation Heatmap

Dalam melihat hubungan linier antara variabel-variabel pada dataset, dilakukan *correlation heatmap*. Korelasi adalah ukuran statistik yang menunjukkan sejauh mana dua variabel bergerak bersama. *Heatmap* memberikan cara yang mudah untuk melihat

dan memahami matriks korelasi, dengan menggunakan warna untuk mewakili kekuatan dan arah hubungan antara variabel. Korelasi positif ($0 - 1$) menunjukkan bahwa ketika suatu variabel meningkat, variabel lainnya juga cenderung meningkat. Korelasi negatif ($0 - (-1)$) menunjukkan bahwa suatu variabel meningkat, variabel lainnya cenderung menurun. Korelasi nol (0) menunjukkan tidak ada hubungan yang linier antara variabel-variabel tersebut. Dalam mengkategorikan hubungan antara variabel digunakan *Pearson Correlation Coefficient*. Tabel 3.2 menunjukkan kategori *Pearson Correlation Coefficient* pada dataset.

Tabel 3. 2 Kategori *Pearson Correlation Coefficient* (Sumber: Dahliani *et al.*, 2020)

Interval Koefisien (r)	Kategori
0.00 – 0.20	Sangat Rendah
0.21 – 0.40	Rendah
0.41 – 0.60	Cukup Signifikan
0.61 – 0.80	Cukup Tinggi
0.81 – 0.99	Tinggi
1	Korelasi Sempurna



Gambar 3. 2 Korelasi Antar Variabel Pada Dataset

Pada Gambar 3.2, menunjukkan hasil bahwa jumlah kasus malaria (*total_case*) memiliki korelasi dengan kategori sangat rendah pada variabel suhu maksimum (*TR_max*) dengan nilai 0.18 dan variabel *index* jumlah hujan (*CR*) dengan nilai 0.02. Sedangkan korelasi dengan kategori cukup signifikan didapatkan pada variabel durasi

sinar matahari (DR_M) dengan nilai 0.58 dan variabel *index* curah hujan (CH) dengan nilai 0.52. Korelasi dengan kategori cukup tinggi didapatkan pada variabel *index* tingkat banjir dan genangan air (F_I) dengan nilai 0.70 dan variabel *index* kepadatan dan pertumbuhan penduduk (PI_Den) dengan nilai 0.65.

Dari hasil *correlation heatmap*, menunjukkan bahwa beberapa faktor lingkungan memiliki hubungan yang lebih kuat dengan jumlah kasus malaria dibandingkan dengan faktor lainnya. Faktor seperti durasi sinar matahari, curah hujan, tingkat banjir/genangan air, dan kepadatan penduduk menunjukkan korelasi yang lebih signifikan terhadap jumlah kasus malaria. Sementara itu, suhu maksimum dan jumlah hujan menunjukkan korelasi yang sangat rendah, menunjukkan bahwa faktor-faktor ini kurang berpengaruh terhadap jumlah kasus malaria.

3.2.4 Implementasi *Time-Series Generative Adversial Network* (TGAN)

Setelah dilakukan tahap *pre-processing*, data akan disintesis dengan menggunakan metode TGAN. Tujuan utama dari proses ini adalah untuk menghasilkan data sintesis yang memiliki karakteristik dan pola yang mirip dengan data asli. Data akan diatur ke dalam bentuk *sequence* (urutan) dengan panjang tertentu yang memungkinkan model untuk mempelajari pola dari urutan data dalam interval waktu tertentu. Setelah itu, model generator dan diskriminator dibangun. Generator adalah model yang bertugas menghasilkan data sintesis yang meniru data asli, sementara diskriminator adalah model yang bertugas membedakan antara data asli dan data sintesis yang dihasilkan oleh generator.

Selama proses pelatihan (*training*), generator dan diskriminator dilatih secara bersamaan. Generator berusaha untuk menipu diskriminator dengan menghasilkan data sintesis yang semakin menyerupai data asli, sementara diskriminator belajar untuk semakin baik dalam membedakan data asli dari data sintesis. Proses ini diulang selama sejumlah iterasi di mana pada setiap iterasi, generator dan diskriminator diperbarui untuk meningkatkan kemampuannya masing-masing. Setelah model selesai dilatih, generator digunakan untuk menghasilkan data sintesis baru. Proses ini dimulai dengan memberikan *noise* acak sebagai *input* ke generator, yang kemudian menghasilkan data sintesis berdasarkan pola yang telah dipelajari selama pelatihan. Data sintesis yang dihasilkan kemudian di-denormalisasi kembali ke skala aslinya, sehingga dapat lebih mudah dibandingkan dengan data asli.

Sintesis data yang dilakukan dimulai dari periode waktu 5-1-2015 sampai dengan 24-12-2018 dan akan menghasilkan 418 data. Hal ini dilakukan untuk dapat menciptakan basis data yang lebih luas dan lengkap, yang membantu model untuk belajar dari tren yang lebih panjang. Ini penting untuk meningkatkan akurasi prediksi, karena model dapat menangkap pola yang lebih beragam. Selain itu, dilakukan untuk mengisi kekosongan data dan memastikan bahwa model memiliki gambaran lengkap dari awal periode hingga data asli yang ada.

3.2.5 Data *Splitting*

Data yang sudah melewati tahap *pre-processing* akan dibagi untuk memisahkan dataset menjadi beberapa bagian yang berbeda. Pada penelitian ini, data akan dibagi menjadi tiga bagian yaitu data *training*, data *validation* dan data *testing*. Data *training* digunakan untuk melatih model agar dapat mengenali pola dalam data dan mempelajari parameter-parameter yang optimal. Data *validation* digunakan untuk mengevaluasi performa model selama proses *training* sehingga dapat mengoptimalkan performa model dan mencegah *overfitting*. Data *testing* digunakan untuk menguji performa model dalam melakukan prediksi dan melakukan perbandingan prediksi dengan nilai sebenarnya.

Penelitian ini menggunakan dua dataset. Pada dataset sintesis, jumlah data *training* yang digunakan sebanyak 292 data, data *validation* sebanyak 84 data dan data *testing* sebanyak 42 data. Pada dataset awal, jumlah data *training* yang digunakan sebanyak 147 data, data *validation* sebanyak 42 data dan data *testing* sebanyak 21 data.

3.2.6 Proses *Training*

Pada proses *training*, metode GPR menggunakan data *training* untuk memodelkan dan memahami pola serta hubungan antara variabel fitur dan variabel target. Selanjutnya, penulis melakukan eksplorasi untuk menentukan parameter-parameter yang akan digunakan untuk menggunakan *kernel*. Setiap *kernel* pada metode GPR memiliki beberapa parameter seperti pada *kernel Radial Basis Function* (RBF) yang memiliki parameter *length scale* atau panjang skala yang menentukan seberapa jauh titik data harus berjarak satu sama lain sebelum korelasi antara data menjadi kecil sehingga dapat diabaikan. Pada *kernel Matern* terdapat parameter *length scale* dan nu *values* yang berfungsi untuk mengontrol kehalusan (*smoothness*) dari fungsi korelasi. *Kernel Rational Quadratic* memiliki parameter *length scale* yang mengatur seberapa banyak

panjang skala yang dapat bervariasi. *Kernel ExpSine Square* memiliki *parameter length scale* dan *periodicity* yang menentukan jarak antara puncak-puncak data berulang. *White kernel* memiliki *parameter length scale* dan *noise level* yang menunjukkan seberapa banyak kebisingan acak pada data dan *kernel Dot Product* yang memiliki *parameter length scale* dan σ_0 yang mengatur tingkat pengaruh bias dalam hubungan linier antara fitur-fitur *input*. Kemudian model akan dibuat berdasarkan parameter-parameter yang sudah ditentukan sebelumnya. Dalam menemukan kombinasi *hyperparameter* terbaik di setiap *kernel*, penulis menerapkan metode *Grid Search*. *Grid Search* mengatur *hyperparameter* dengan mencoba semua kombinasi yang mungkin dari *hyperparameter* yang ditentukan. Berikut *pseudocode* dalam mencari parameter terbaik pada *kernel*.

```

DEFINE RANGE OF PARAMETERS
matern_nu_values = [0.5, 1.5, 2.5]
matern_length_scale_values = [1.0, 10.0, 100.0]
dot_product_sigma_0_values = [0.1, 1.0, 10.0]
for nu in matern_nu_values:
    for length_scale in matern_length_scale_values:
        for sigma_0 in dot_product_sigma_0_values:
            kernel = Matern(nu=nu, length_scale=length_scale) +
                DotProduct(sigma_0=sigma_0)
            gpr_model = GaussianProcessRegressor(kernel=kernel)
            gpr_model.fit(train_data, train_labels)

```

Model akan dibuat berdasarkan eksplorasi penggunaan *kernel* tunggal dan *kernel* kombinasi dengan *hyperparameter* yang sudah ditentukan dengan *grid search*. Setelah menentukan *kernel* dan parameter yang akan digunakan, model akan divalidasi untuk mengukur performa model GPR dengan menggunakan *K-Fold Cross Validation*. *K-Fold Cross Validation* merupakan metode validasi silang yang digunakan untuk mengevaluasi kinerja model untuk mengukur seberapa baik model dalam memprediksi data. *K-Fold Cross Validation* bekerja dengan cara membagi data sebanyak K *subset* atau yang dikenal dengan *fold*. Setelah itu, proses validasi dijalankan sebanyak lima kali yang mana setiap iterasi salah satu *fold* dipilih untuk validasi dan sisanya digunakan untuk melatih model. Setelah dilatih, model akan dievaluasi dan akan menghasilkan nilai MSE, RMSE dan MAE.

3.2.7 Proses *Testing*

Proses *testing* dilakukan untuk menguji model yang sudah dilatih dan divalidasi dengan data *testing*. Proses ini melakukan prediksi dan membandingkan nilai prediksi dengan nilai aktual berdasarkan data *testing*. Selanjutnya, akan diukur hasil *testing* tersebut dengan metrik evaluasi RMSE, MSE, dan MAE.

3.2.8 Proses Prediksi

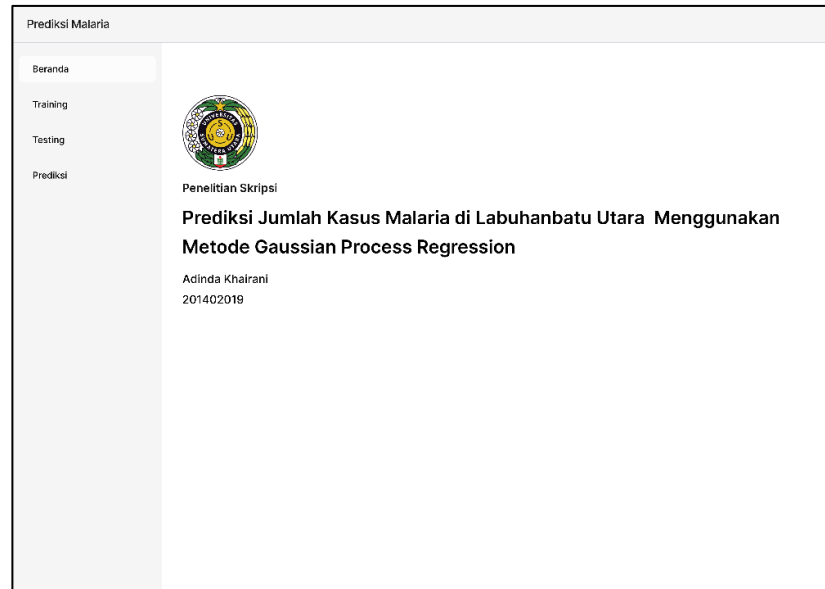
Proses prediksi melakukan prediksi jumlah kasus malaria diluar dataset dengan menggunakan model yang telah dibuat dengan GPR. Model tersebut telah melewati proses pengujian data sebelum digunakan untuk memprediksi. Pada proses ini akan mengeluarkan output prediksi malaria selama 12 minggu ke depan dimulai dari 9-1-2023, 16-1-2023, 23-1-2023, 30-1-2023, 6-2-2023, 13-2-2023, 20-2-2023, 27-2-2023-, 6-3-2023, 13-3-2023, 20-3-2023 dan 27-3-2023.

3.3 Perancangan Antarmuka Sistem

Perancangan antarmuka sistem dilakukan untuk menampilkan hasil proses *training*, *testing* dan prediksi ke dalam tampilan sistem berbasis website yang lebih mudah digunakan oleh pengguna. Tahap ini bertujuan untuk merancang sistem sebelum diimpelentasikan agar mendapatkan gambaran bagaimana tampilan sistem dibuat.

3.3.1 Rancangan Tampilan Beranda

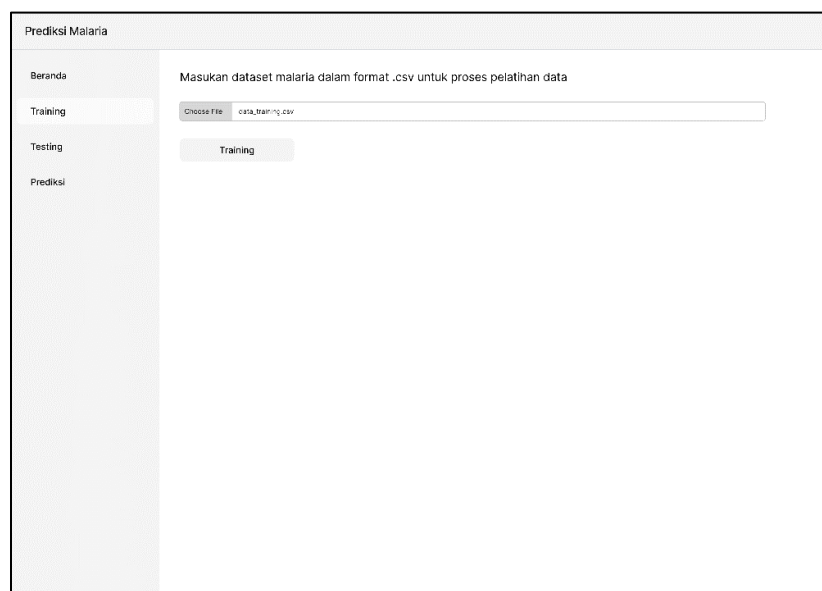
Rancangan tampilan beranda adalah tampilan awal ketika pengguna menjalankan sistem ini. Pada *layout* kiri terdapat *sidebar* yang menampilkan menu *training*, *testing*, dan prediksi. Ketika pengguna mengakses salah satu menu pada *sidebar*, akan diarahkan ke halaman menu yang pengguna akses. Kemudian, pada *layout* kanan menampilkan informasi berupa logo universitas, judul penelitian, nama dan NIM penulis. Perancangan tampilan beranda dapat dilihat pada Gambar 3.3.



Gambar 3. 3 Rancangan Tampilan Beranda

3.3.2 Rancangan Tampilan *Training*

Pada rancangan tampilan *training*, akan menampilkan *form input* data untuk pengguna mengunggah dataset berformat .csv yang akan dilatih. Kemudian pengguna menekan tombol *training*, lalu proses pelatihan akan berjalan. Berikut perancangan tampilan *training* pada Gambar 3.4.



Gambar 3. 4 Rancangan Tampilan *Training*

Setelah proses pelatihan selesai, sistem akan menampilkan tabel hasil *training* dan grafik metrik evaluasi berdasarkan dari setiap hasil *training*. Perancangan tampilan hasil *training* dapat dilihat pada Gambar 3.5.



Gambar 3. 5 Rancangan Tampilan Hasil *Training*

3.3.3 Rancangan Tampilan *Testing*

Tampilan *testing* akan menampilkan *form input* data untuk pengguna mengunggah dataset berformat .csv yang akan dilakukan pengujian oleh sistem. Setelah pengguna menekan tombol *testing*, sistem akan melakukan proses pelatihan data. Perancangan tampilan halaman *testing* dapat dilihat pada Gambar 3.6.

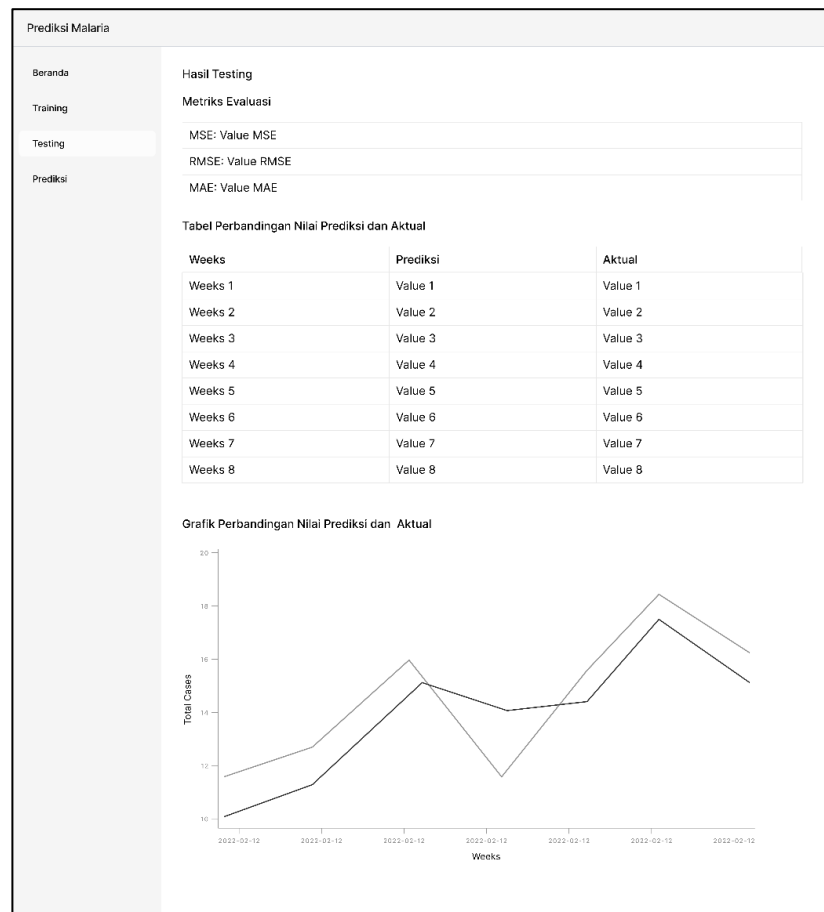
The screenshot shows the 'Prediksi Malaria' interface. On the left is a sidebar with 'Beranda', 'Training', 'Testing', and 'Prediksi'. The 'Testing' section is active, displaying a form for uploading a dataset.

The form contains the following elements:

- A header text: "Masukan dataset malaria dalam format .csv untuk proses pengujian data"
- A file input field with a "Choose File" button and the text "cara_testing.csv".
- A "Testing" button.

Gambar 3. 6 Rancangan Tampilan *Testing*

Setelah proses pelatihan selesai, sistem akan menampilkan hasil pengujian berupa tabel metrik evaluasi hasil pengujian, tabel perbandingan yang mempunyai kolom *weeks*, nilai *total case* prediksi dan aktual. Kemudian di bawahnya terdapat grafik hasil perbandingan nilai *total case* prediksi dan aktual. Perancangan tampilan *testing* dapat dilihat pada Gambar 3.7.



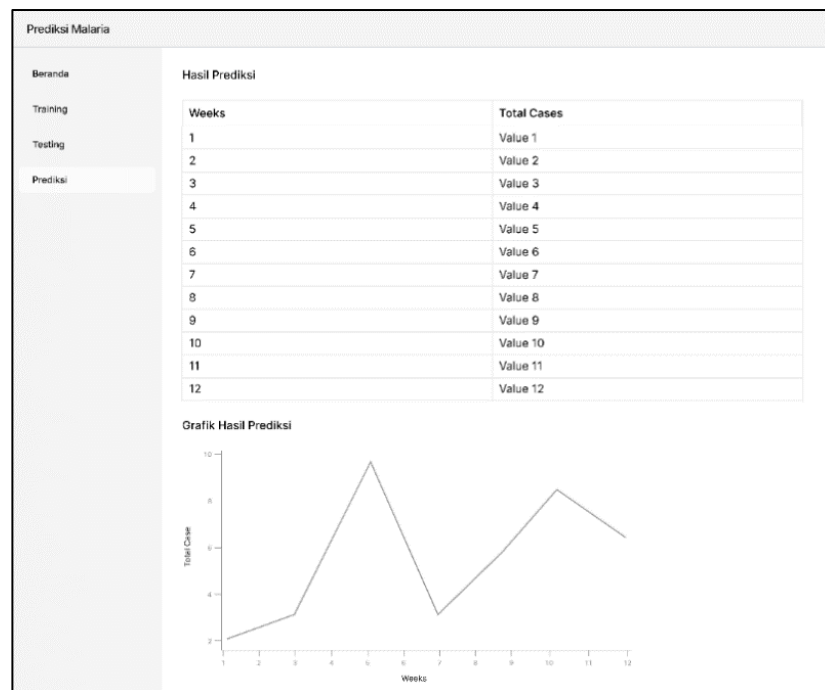
Gambar 3. 7 Rancangan Tampilan Hasil *Testing*

3.3.4 Rancangan Tampilan Prediksi

Tampilan prediksi akan memprediksi jumlah kasus malaria di luar dataset selama 12 minggu ke depan. Tampilan ini menampilkan *form input* data yang mengharuskan pengguna mengunggah data berformat .csv. Proses prediksi akan berjalan Ketika pengguna mengklik tombol prediksi. Perancangan tampilan prediksi dapat dilihat pada Gambar 3.8.

Gambar 3. 8 Rancangan Tampilan Prediksi

Setelah proses prediksi selesai, tampilan akan menyajikan tabel hasil prediksi yang memiliki kolom *weeks* dan nilai *total case*. Prediksi dimulai dari minggu pertama sampai minggu ke-12. Selanjutnya, sistem akan menampilkan grafik hasil prediksi. Tampilan prediksi dapat dilihat pada Gambar 3.9.



Gambar 3. 9 Rancangan Tampilan Hasil Prediksi

BAB 4

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Implementasi Sistem

Pada tahap implementasi sistem, GPR akan diterapkan untuk memprediksi jumlah kasus malaria di Labuhanbatu Utara. Bagian ini akan menjelaskan penerapan sistem yang sudah direncanakan pada bagian sebelumnya termasuk di dalamnya penjelasan terkait perangkat keras dan perangkat lunak yang dibutuhkan dalam mengimplementasikan sistem.

4.1.1 Perangkat Keras dan Perangkat Lunak

Pengimplementasian sistem prediksi jumlah kasus malaria membutuhkan perangkat keras dan perangkat lunak dengan spesifikasi sebagai berikut:

1. Processor Intel(R) Core (TM) i3-1005G1 CPU @ 1.20GHz 1.19 GHz
2. RAM dengan kapasitas sebesar 4 GB
3. SSD dengan kapasitas sebesar 250 GB
4. Sistem Operasi Windows 11 Home-64 bit
5. Google Colab
6. Visual Studio Code
7. Bahasa pemrograman Python versi 3.10.5 dengan *library* pendukung yang digunakan seperti *scikit-learn* versi 1.5.1, *numpy* versi 2.0.0, *pandas* versi 2.2.2, *scipy* versi 1.14.0, *matplotlib* versi 3.9.1 dan *seaborn* versi 0.13.2.

4.1.2 Implementasi Data

Data yang digunakan pada penelitian ini berformat .csv dengan jumlah data sebanyak 418 data. Data yang digunakan pada sistem sudah melewati proses *pre-processing* agar data yang digunakan memiliki kualitas yang lebih baik sehingga performa dan akurasi model yang digunakan pada sistem menjadi lebih optimal.

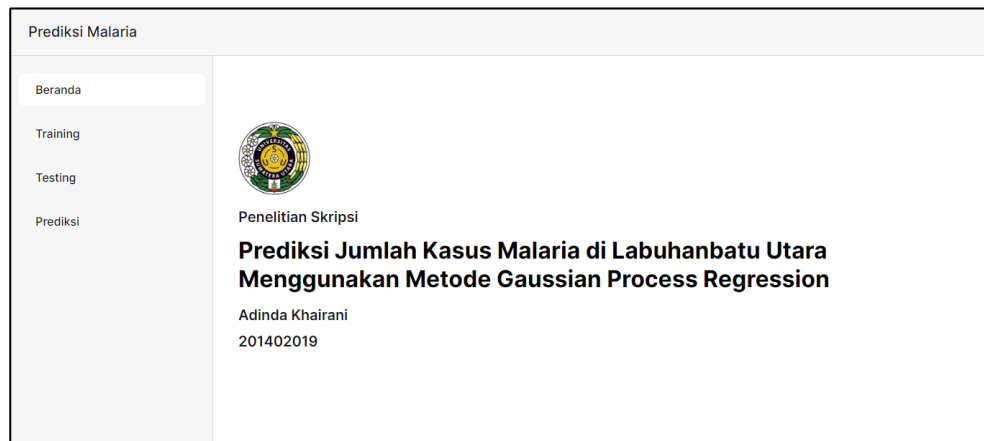
4.1.3 Implementasi Tampilan Antarmuka

Setelah dilakukan perancangan antarmuka sistem pada bab sebelumnya, rancangan tersebut akan diimplementasikan ke dalam sistem untuk memastikan bahwa fungsionalitas dan pengalaman pengguna sesuai dengan rancangan yang telah

direncanakan. Sistem ini dibuat dengan menggunakan *framework* Flask untuk bagian *backend* dan HTML, CSS, serta JavaScript untuk bagian antarmuka atau *frontend*. Tampilan antarmuka memiliki tampilan beranda, tampilan *training*, tampilan *testing* dan tampilan prediksi.

4.1.3.1 Tampilan Halaman Beranda

Tampilan halaman beranda merupakan halaman awal ketika pengguna mengakses sistem ini. Pada tampilan beranda di *layout* kiri menampilkan *sidebar* menu-menu yang dapat diakses pengguna. *Layout* kanan menampilkan logo universitas, judul penelitian, nama dan NIM penulis. Gambar 4.1 menunjukkan tampilan halaman beranda sistem.



Gambar 4. 1 Tampilan Halaman Beranda

4.1.3.2 Tampilan Halaman Training

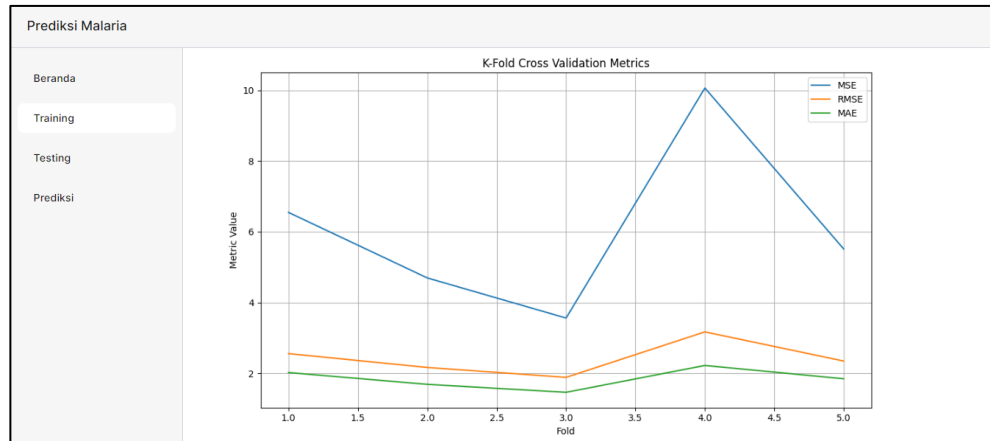
Halaman *training* berfungsi untuk melatih data yang akan pengguna unggah pada sistem ini. Halaman *training*, menampilkan *form input* untuk pengguna mengunggah dataset berformat .csv. Gambar 4.2 menunjukkan tampilan halaman *training*.

Gambar 4. 2 Tampilan Halaman *Training*

Setelah pengguna mengunggah dataset berformat .csv dan menekan tombol *training*, proses pelatihan data akan berlangsung. Setelah selesai, akan menampilkan halaman hasil *training* yang memuat tabel hasil *training* yang memiliki kolom *fold*, MSE, RMSE, dan MAE. Di bawahnya terdapat grafik hasil pelatihan setiap *fold*. Gambar 4.3 menunjukkan tampilan halaman hasil *training* dalam bentuk tabel dan Gambar 4.4 menunjukan tampilan halaman hasil *training* dalam bentuk grafik.

Fold	MSE	RMSE	MAE
1	6.54	2.55	2.02
2	3.69	2.16	1.69
3	3.56	1.88	1.46
4	10.06	3.17	2.22
5	5.51	2.34	1.84

Gambar 4. 3 Tampilan Tabel Hasil *Training*



Gambar 4. 4 Tampilan Grafik Hasil Training

4.1.3.3 Tampilan Halaman Testing

Halaman *testing* berfungsi untuk menguji data yang sudah dilatih pada halaman sebelumnya. Pengguna diharuskan mengunggah data berformat .csv pada *form input* yang ada pada halaman *testing*. Gambar 4.5 menunjukkan tampilan halaman *testing*.

The screenshot shows the "Prediksi Malaria" application interface. On the left is a sidebar with navigation links: Beranda, Training, Testing (highlighted), and Prediksi. The main content area has a heading "Masukan dataset malaria dalam format .csv untuk proses pengujian data". Below this heading is a file upload section with a "Choose File" button and a text input field containing "data_testing.csv". At the bottom of this section is a purple "Testing" button.

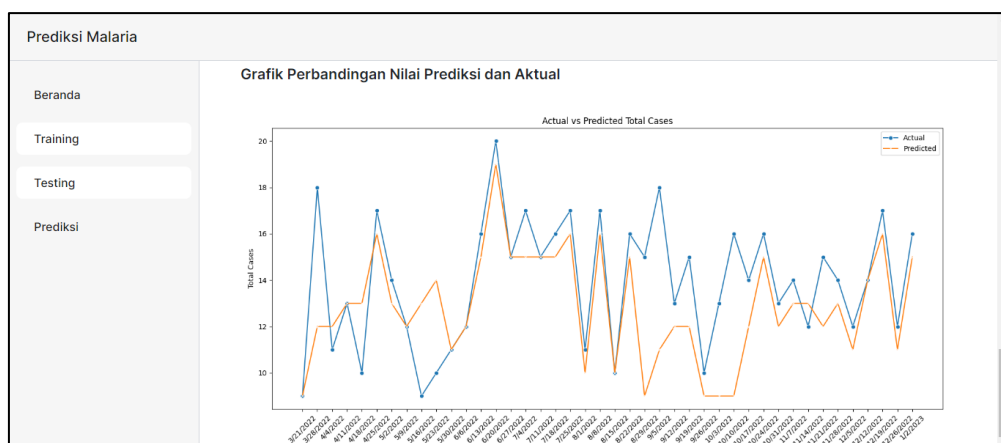
Gambar 4. 5 Tampilan Halaman Testing

Setelah pengguna menekan tombol *testing*, proses pengujian data akan berlangsung. Setelah itu, halaman akan dialihkan ke halaman hasil *testing*. Pada halaman hasil *testing* akan menunjukkan tabel metrik evaluasi yang memiliki kolom MSE, RMSE dan MAE, tabel perbandingan nilai prediksi dan aktual *total case* dan grafik dari hasil perbandingan. Gambar 4.6 dan Gambar 4.7 menunjukkan tampilan hasil *testing* serta Gambar 4.8 menunjukkan tampilan grafik hasil *testing*.

Prediksi Malaria														
Beranda	Hasil Testing													
Training	Metrik Evaluasi													
Testing	<div>MSE: 5.484821369146707</div> <div>RMSE: 2.341969549150182</div> <div>MAE: 1.2789513364499194</div>													
Prediksi	Tabel Perbandingan Nilai Prediksi dan Aktual <table> <tr> <th>Weeks</th><th>Prediksi</th><th>Aktual</th></tr> <tr> <td>3/21/2022</td><td>9</td><td>9</td></tr> <tr> <td>3/28/2022</td><td>12</td><td>18</td></tr> <tr> <td>4/4/2022</td><td>12</td><td>11</td></tr> </table>		Weeks	Prediksi	Aktual	3/21/2022	9	9	3/28/2022	12	18	4/4/2022	12	11
Weeks	Prediksi	Aktual												
3/21/2022	9	9												
3/28/2022	12	18												
4/4/2022	12	11												

Gambar 4. 6 Tampilan Tabel Hasil *Testing*

Prediksi Malaria				
Beranda		4/4/2022	12	11
		4/11/2022	13	13
Training		4/18/2022	13	10
Testing		4/25/2022	16	17
		5/2/2022	13	14
Prediksi		5/9/2022	12	12
		5/16/2022	13	9
		5/23/2022	14	10
		5/30/2022	11	11
		6/6/2022	12	12
		6/13/2022	15	16
		6/20/2022	19	20
		6/27/2022	15	15

Gambar 4. 7 Tampilan Tabel Hasil *Testing*

Gambar 4. 8 Tampilan Grafik Hasil Testing

4.1.3.4 Tampilan Halaman Prediksi

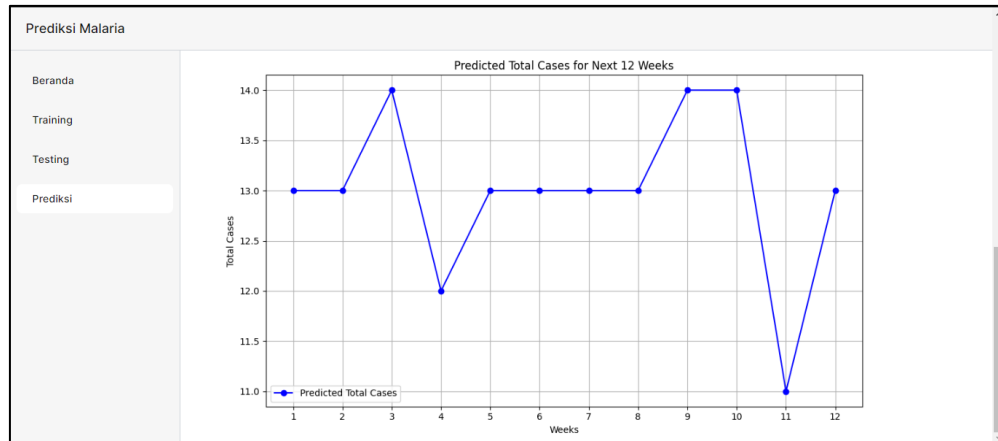
Halaman prediksi berfungsi untuk memprediksi jumlah kasus malaria. Halaman ini akan menampilkan *form input* untuk pengguna mengunggah dataset berformat .csv. Gambar 4.9 menunjukkan tampilan halaman prediksi.

Gambar 4. 9 Tampilan Halaman Prediksi

Setelah pengguna mengunggah dataset untuk diprediksi, pengguna akan menekan tombol prediksi. Lalu proses prediksi akan berjalan. Kemudian akan menampilkan tabel hasil prediksi yang memiliki kolom *weeks* dan *total case*. Di bawahnya terdapat grafik berdasarkan hasil prediksi. Gambar 4.10 menunjukan tampilan tabel hasil prediksi dan Gambar 4.11 menunjukkan grafik hasil prediksi.

Weeks	Prediksi
1	13
2	13
3	14
4	12
5	13
6	13
7	13
8	13

Gambar 4. 10 Tampilan Tabel Hasil Prediksi



Gambar 4. 11 Tampilan Grafik Hasil Prediksi

4.2. Pelatihan Model

Penulis melakukan eksplorasi dalam pemilihan parameter yang optimal di setiap *kernel* pada model *training* GPR. *Kernel* yang digunakan pada penelitian ini adalah *kernel Radial Basis Function* (RBF), *kernel Matern*, *kernel Rational Quadratic*, *kernel ExpSine Square*, *White kernel* dan *Dot Product kernel*. Setiap *kernel* memiliki parameter yang berbeda-beda. Inisialisasi parameter setiap *kernel* menggunakan metode *Grid Search Cross Validation* (CV). Setelah *kernel* digunakan pada model *training*, akan dihitung kinerjanya menggunakan MSE, RMSE dan MAE. Tabel 4.1 menunjukkan implementasi model *training* pada setiap parameter di *kernel* tunggal

Tabel 4. 1 Implementasi Model Training Pada *Kernel* Tunggal

<i>Kernel</i>	Parameter		MSE	RMSE	MAE
RBF	length scale	0.5	12.87	3.53	2.59
<i>Matern</i>	nu values	0.5	7.61	2.71	0.65
	length scale	1.0			
<i>Matern</i>	nu values	1.5	8.98	2.94	2.15
	length scale	1.0			
<i>Matern</i>	nu values	2.5	33.12	6.06	4.03
	length scale	10			
<i>Rational Quadratic</i>	alpha	0.1	8.6	2.88	0.60
	length scale	1.0			

<i>Kernel</i>	Parameter		MSE	RMSE	MAE
<i>ExpSine Square</i>	periodicity	0.1	12.87	3.53	2.59
	length scale	1.0			
<i>White Kernel</i>	noise level	0.01	64.15	7.99	6.56
<i>Dot Product</i>	sigma 0	0.1	7.82	2.76	2.11
	sigma 0	1.0	7.88	2.77	2.10
	sigma 0	10	7.80	2.76	2.10

Tabel 4.1 menunjukkan hasil yang cukup bagus pada *kernel Matern* dengan *nu values* 0.5 dan *length scale* 1.0. *Kernel Matern* dengan parameter tersebut menghasilkan nilai MSE sebesar 7.61, RMSE sebesar 2.71 dan MAE sebesar 0.65. Hasil ini merupakan hasil yang paling bagus di antara *kernel* tunggal lainnya.

Setelah melakukan pengimplementasian model pada *kernel* tunggal, penulis melakukan implementasi model pada *kernel* kombinasi. *Kernel* kombinasi melakukan penggabungan antara satu *kernel kernel* dengan *kernel* lainnya. Tabel 4.2 menunjukkan hasil pengimplementasian model *training* pada *kernel* kombinasi RBF dan *kernel* lainnya.

Tabel 4. 2 Implementasi Model *Training* Pada *Kernel* Kombinasi RBF dengan *Kernel* Lainnya

<i>Kernel</i>	Parameter		MSE	RMSE	MAE
RBF <i>Matern</i>	length scale	0.1	8.38	2.85	2.14
	nu values	0.5			
	length scale	0.1			
RBF <i>Matern</i>	length scale	0.1	7.68	2.72	2.01
	nu values	1.5			
	length scale	0.1			
RBF <i>Matern</i>	length scale	0.1	7.56	2.85	2.14
	nu values	2.5			
	length scale	0.1			

<i>Kernel</i>	<i>Parameter</i>	<i>MSE</i>	<i>RMSE</i>	<i>MAE</i>	<i>Kernel</i>
RBF <i>Rational</i> <i>Quadratic</i>	length scale	0.1	8.10	2.80	2.09
	length scale	10			
	alpha	0.1			
RBF <i>Rational</i> <i>Quadratic</i>	length scale	1.0	7.66	2.72	2.04
	length scale	0.1			
	alpha	0.1			
RBF <i>Rational</i> <i>Quadratic</i>	length scale	10	8.06	2.79	2.09
	length scale	1.0			
	alpha	1.0			
RBF <i>ExpSine Square</i>	length scale	0.1	7.48	2.68	1.98
	length scale	0.1			
	alpha	0.5			
RBF	length scale	0.1	20.07	4.45	3.57
White Kernel	noise level	0.1			
RBF	length scale	0.1	7.65	2.73	2.06
Dot Product	sigma 0	0.1			

Dari hasil *training* dengan menggabungkan *kernel* RBF dengan *kernel* lainnya, didapatkan hasil terbaik pada penggabungan *kernel* RBF dengan *kernel* *ExpSine Square* pada parameter *length scale* RBF 0.1 dan *length scale* *ExpSine Square* 0.1 dan *alpha* 0.5. Hasil evaluasi dari penggabungan *kernel* tersebut adalah MSE sebesar 7.48, RMSE sebesar 2.68 dan MAE sebesar 1.98. Hal ini menunjukkan bahwa pengaruh nilai *length scale* pada *kernel* yang sama menghasilkan variasi kinerja yang cukup signifikan.

Selanjutnya, penulis melakukan pengimplementasian gabungan *kernel* *Matern* dengan *kernel* lainnya. Tabel 4.3 menunjukkan hasil implementasi *kernel* *Matern* dengan *kernel* lainnya.

Tabel 4. 3 Implementasi Model Training Pada *Kernel* Kombinasi *Matern* dengan *Kernel* Lainnya

<i>Kernel</i>	Parameter		MSE	RMSE	MAE
<i>Matern</i> <i>Rational</i> <i>Quadratic</i>	nu values	0.5	8.15	2.81	2.11
	length scale	0.1			
	alpha	0.1			
	length scale	0.1			
<i>Matern</i> <i>Rational</i> <i>Quadratic</i>	nu values	1.5	7.98	2.78	2.09
	length scale	1.0			
	alpha	0.1			
	length scale	0.1			
<i>Matern</i> <i>Rational</i> <i>Quadratic</i>	nu values	2.5	7.81	2.75	2.06
	length scale	1.0			
	alpha	0.1			
	length scale	0.1			
<i>Matern</i> <i>ExpSine Square</i>	nu values	0.5	8.21	2.81	2.12
	length scale	0.1			
	periodicity	0.1			
	length scale	0.1			
<i>Matern</i> <i>ExpSine Square</i>	nu values	0.5	7.66	2.72	2.01
	length scale	1.0			
	periodicity	0.1			
	length scale	0.1			
<i>Matern</i> <i>White Kernel</i>	nu values	0.5	20.06	4.45	3.56
	length scale	0.1			
	noise level	0.01			
<i>Matern</i> <i>White Kernel</i>	nu values	1.5	19.95	4.44	3.56
	length scale	0.1			
	noise level	0.1			

Kernel	Parameter		MSE	RMSE	MAE
<i>Matern</i> <i>Dot Product</i>	nu values	0.5	6.99	2.61	1.96
	length scale	0.1			
	sigma 0	0.1			
<i>Matern</i> <i>Dot Product</i>	nu values	1.5	7.29	2.66	2.01
	length scale	0.1			
	sigma 0	10			

Dari hasil *training* dengan menggabungkan *kernel Matern* dengan *kernel* lainnya, didapatkan hasil terbaik pada penggabungan *kernel Matern* dengan *kernel Dot Product* pada parameter *nu values* 0.5, *length scale Matern* 0.1 dan *sigma 0 Dot Product* 0.1. Hasil evaluasi dari penggabungan *kernel* tersebut adalah MSE sebesar 6.99, RMSE sebesar 2.61 dan MAE sebesar 1.96. Kombinasi *kernel Matern* dan *Dot Product* memberikan hasil terbaik karena kemampuan mereka untuk menangani berbagai jenis pola dalam data dan fleksibilitas dalam menangani variasi.

Tabel 4. 4 Implementasi Model Training Pada *Kernel* Kombinasi *Rational Quadratic* dengan *Kernel* Lainnya

Kernel	Parameter		MSE	RMSE	MAE
<i>Rational</i> <i>Quadratic</i> <i>ExpSine Square</i>	alpha	0.1	8.07	2.79	2.09
	length scale	0.1			
	periodicity	0.1			
	length scale	0.1			
<i>Rational</i> <i>Quadratic</i> <i>White Kernel</i>	alpha	0.1	19.91	4.44	3.55
	length scale	0.1			
	noise level	0.1			
<i>Rational</i> <i>Quadratic</i> <i>Dot Product</i>	alpha	0.1	7.64	2.73	2.08
	length scale	0.1			
	sigma 0	0.1			

Kernel	Parameter		MSE	RMSE	MAE
<i>Rational Quadratic Dot Product</i>	alpha	0.1	7.03	2.64	1.99
	length scale	10			
	sigma 0	1.0			
<i>Rational Quadratic Dot Product</i>	alpha	1.0	7.01	2.61	1.97
	length scale	10			
	sigma 0	0.1			

Dari hasil *training* dengan menggabungkan *kernel Rational Quadratic* dengan *kernel* lainnya, didapatkan hasil terbaik pada penggabungan *kernel Rational Quadratic* dengan *kernel Dot Product* pada parameter *alpha* 1.0, *length scale Rational Quadratic* 10 dan *sigma 0 Dot Product* 0.1. Hasil evaluasi dari penggabungan *kernel* tersebut adalah MSE sebesar 7.01, RMSE sebesar 2.61 dan MAE sebesar 1.97. *Length scale Rational Quadratic* yang lebih besar memberikan kemampuan model untuk menangkap tren umum dalam data yang dapat meningkatkan akurasi prediksi.

Tabel 4. 5 Implementasi Model Training Pada *Kernel Kombinasi ExpSine Square* dengan *Kernel Lainnya*

Kernel	Parameter		MSE	RMSE	MAE
<i>ExpSine Square White Kernel</i>	periodicity	0.1	20.07	4.45	3.57
	length scale	0.1			
	noise level	0.1			
<i>ExpSine Square Dot Product</i>	periodicity	0.1	7.64	2.73	2.08
	length scale	0.1			
	sigma 0	0.1			

Dari hasil *training* dengan menggabungkan *kernel ExpSine Square* dengan *kernel* lainnya, didapatkan hasil terbaik pada penggabungan *kernel ExpSine Square* dengan *kernel Dot Product* pada parameter *periodicity* 0.1 *length scale ExpSine Square* 0.1 dan *sigma 0 Dot Product* 0.1. Hasil evaluasi dari penggabungan *kernel* tersebut adalah MSE sebesar 7.64, RMSE sebesar 2.73 dan MAE sebesar 2.08. Hal ini menunjukkan bahwa

kombinasi *kernel* ini cukup efektif dalam menangkap pola periodik dalam data tetapi tidak pada kombinasi *White Kernel*.

Setelah melakukan kombinasi antara satu *kernel* dengan *kernel* lainnya, penulis merangkum hasil terbaik dari setiap *kernel*nya. Tabel 4.6 menampilkan hasil terbaik dari setiap *kernel*nya.

Tabel 4. 6 Hasil Terbaik dari Kombinasi *Kernel*

<i>Kernel</i>	Parameter		MSE	RMSE	MAE
RBF <i>ExpSine Square</i>	length scale	0.1	7.48	2.68	1.98
	length scale	0.1			
	alpha	0.5			
<i>Matern</i> <i>Dot Product</i>	nu values	0.5	6.99	2.61	1.96
	length scale	0.1			
	sigma 0	0.1			
<i>Rational</i> <i>Quadratic</i> <i>Dot Product</i>	alpha	1.0	7.01	2.61	1.97
	length scale	10			
	sigma 0	0.1			
<i>ExpSine Square</i> <i>Dot Product</i>	periodicity	0.1	7.64	2.73	2.08
	length scale	0.1			
	sigma 0	0.1			

Hasil dari kombinasi terbaik antara satu *kernel* dengan *kernel* lainnya didapatkan bahwa kombinasi *Kernel Matern* dengan *kernel Dot Product* pada parameter *nu values* 0.5, *length scale Matern* 0.1 dan *sigma 0* 0.1 dengan hasil MSE sebesar 6.99, RMSE sebesar 2.61 dan MAE sebesar 1.96. Hal ini menunjukkan bahwa kombinasi antara dua *kernel* ini sangat efektif dalam menangkap pola dalam data dengan baik. Kombinasi ini menggabungkan kemampuan *kernel Matern* untuk menangkap non-linieritas dengan kemampuan *kernel Dot Product* untuk menangkap hubungan linier, sehingga cocok untuk data yang heterogen. Selanjutnya kombinasi *kernel Matern* dengan *kernel Dot Product* pada parameter tersebut akan digunakan untuk melatih model pada penelitian ini.

Setelah mendapatkan kombinasi *kernel* dan parameternya, selanjutnya dilakukan *K-Fold Cross Validation* untuk melatih model dengan membagi data menjadi lima *fold* yang mana setiap iterasinya empat *fold* akan digunakan untuk melatih model dengan data *training* dan satu *fold* akan digunakan untuk memvalidasi model dengan data *validation*. Setiap iterasi akan mengacak pembagian *fold* untuk melatih dan memvalidasi model. Tabel 4.7 menunjukkan hasil *K-Fold Cross Validation* yang telah dilakukan.

Tabel 4. 7 Hasil *K-Fold Cross Validation*

Fold	MSE	RMSE	MAE
1	6.54	2.55	2.02
2	3.69	2.16	1.69
3	3.56	1.88	1.46
4	10.06	3.17	2.22
5	5.51	2.34	1.84

Dari Tabel 4.7, menunjukkan bahwa Fold 3 memiliki kinerja yang sangat baik dengan nilai MSE sebesar 3.56, RMSE sebesar 1.88 dan MAE sebesar 1.46. Selanjutnya dilakukan perhitungan rata-rata pada kelima *fold* tersebut untuk digunakan sebagai data uji.

4.3. Pengujian Model

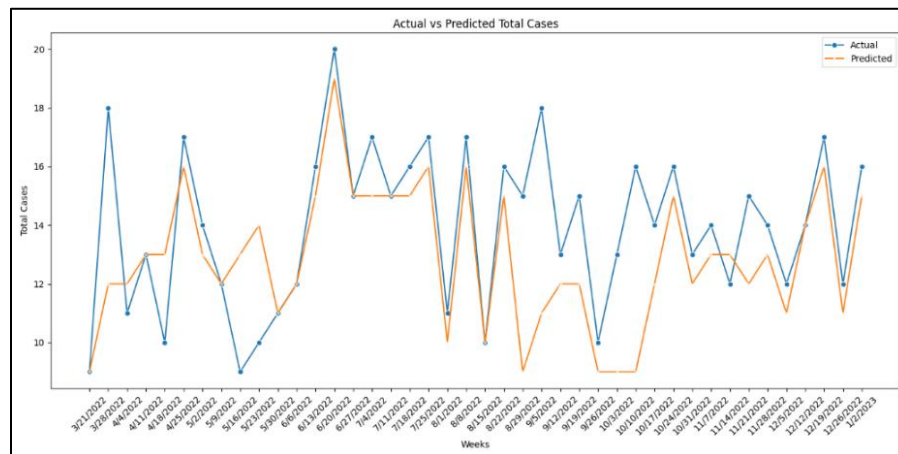
Setelah melatih model dengan menggunakan kombinasi *kernel Matern* dan *kernel Dot Product*, dilakukan pengujian untuk mengetahui kinerja model sebelum dilakukannya prediksi dengan membandingkan nilai prediksi dengan nilai aktual. Pengujian model dilakukan dengan menggunakan data yang sudah disintesis dan data asli dari penelitian sebelumnya. Selanjutnya, pengujian model akan dinilai dengan menggunakan tiga metrik evaluasi yaitu *Mean Square Error* (MSE), *Root Mean Square Error* (RMSE) dan *Mean Absolute Error* (MAE). Model terbaik akan menghasilkan nilai yang rendah untuk ketiga metrik evaluasi tersebut. Tabel 4.8 merupakan hasil pengujian model.

Tabel 4. 8 Hasil Pengujian Model dengan Data Sintesis

Weeks	Prediksi	Aktual
21/03/2022	9	9
28/03/2022	12	18
04/04/2022	12	11
11/04/2022	13	13
18/04/2022	13	10
25/04/2022	16	17
02/05/2022	13	14
09/05/2022	12	12
16/05/2022	13	9
23/05/2022	14	10
30/05/2022	11	11
06/06/2022	12	12
13/06/2022	15	16
20/06/2022	19	20
27/06/2022	15	15
04/07/2022	15	17
11/07/2022	15	15
18/07/2022	15	16
25/07/2022	16	17
01/08/2022	10	11
08/08/2022	16	17
15/08/2022	10	10
22/08/2022	15	16
29/08/2022	9	15
05/09/2022	11	18
12/09/2022	12	13
19/09/2022	12	15
26/09/2022	9	10
03/10/2022	9	13
10/10/2022	9	16

Weeks	Prediksi	Aktual
21/03/2022	9	9
28/03/2022	12	18
04/04/2022	12	11
11/04/2022	13	13
18/04/2022	13	10
25/04/2022	16	17
02/05/2022	13	14
09/05/2022	12	12
16/05/2022	13	9
23/05/2022	14	10
30/05/2022	11	11
06/06/2022	12	12
13/06/2022	15	16
20/06/2022	19	20
27/06/2022	15	15
04/07/2022	15	17
11/07/2022	15	15
18/07/2022	15	16
25/07/2022	16	17
17/10/2022	12	14
24/10/2022	15	16
31/10/2022	12	13
07/11/2022	13	14
14/11/2022	13	12
21/11/2022	12	15
28/11/2022	13	14
05/12/2022	11	12
12/12/2022	14	14
19/12/2022	16	17
26/12/2022	11	12
02/01/2023	15	16

Hasil pengujian dengan menggunakan data sintesis juga dapat dilihat melalui grafik pada Gambar 4.12.



Gambar 4. 12 Grafik Hasil Prediksi dengan Data Sintesis

Hasil perbandingan nilai prediksi dan nilai aktual *total case* pada hasil pengujian menunjukkan bahwa beberapa minggu nilai prediksi cukup dekat dengan nilai aktual seperti pada minggu 21/03/2022, 27/06/2023 dan 11/07/2022. Terdapat juga minggu-minggu yang memiliki perbedaan yang cukup signifikan seperti pada minggu 28/03/2022, 29/08/2022, 05/09/2022 dan 10/10/2022. Berdasarkan hasil ini, menunjukkan bahwa model yang digunakan cenderung memperkirakan dengan baik, meskipun ada variasi dalam akurasi prediksi dari waktu ke waktu.

Dari hasil perbandingan tersebut, digunakan untuk menghitung metrik evaluasi MSE, RMSE dan MAE dengan menghitung terlebih dahulu perbedaan nilai prediksi dan nilai aktual. Tabel 4.9 menunjukkan hasil perbedaan nilai prediksi dan nilai aktual setiap minggunya.

Tabel 4. 9 Hasil Perbedaan Nilai Prediksi dengan Nilai Aktual

Weeks	Prediksi	Aktual	$y_i - \hat{y}_i$
21/03/2022	9	9	0
28/03/2022	12	18	6
04/04/2022	12	11	1
11/04/2022	13	13	0

Weeks	Prediksi	Aktual	$y_i - \hat{y}_i$
18/04/2022	13	10	3
25/04/2022	16	17	1
02/05/2022	13	14	1
09/05/2022	12	12	0
16/05/2022	13	9	4
23/05/2022	14	10	4
30/05/2022	11	11	0
06/06/2022	12	12	0
13/06/2022	15	16	1
20/06/2022	19	20	1
27/06/2022	15	15	0
04/07/2022	15	17	2
11/07/2022	15	15	0
18/07/2022	15	16	1
25/07/2022	16	17	1
01/08/2022	10	11	1
08/08/2022	16	17	1
15/08/2022	10	10	0
22/08/2022	15	16	1
29/08/2022	9	15	6
05/09/2022	11	18	7
12/09/2022	12	13	1
19/09/2022	12	15	3
26/09/2022	9	10	1
03/10/2022	9	13	4
10/10/2022	9	16	7
17/10/2022	12	14	2
24/10/2022	15	16	1
31/10/2022	12	13	1
07/11/2022	13	14	1
14/11/2022	13	12	1

Weeks	Prediksi	Aktual	$y_i - \hat{y}_i$
21/11/2022	12	15	3
28/11/2022	13	14	1
05/12/2022	11	12	1
12/12/2022	14	14	0
19/12/2022	16	17	1
26/12/2022	11	12	1
02/01/2023	15	16	1

Dari hasil selisih nilai prediksi dan nilai aktual pada tabel di atas, dilakukan perhitungan untuk mencari nilai MSE, RMSE dan MAE. Berikut perhitungan untuk mencari nilai MSE.

$$\begin{aligned}
 \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\
 &= \frac{1}{42} |0^2 + 6^2 + 1^2 + 0^2 + 3^2 + 1^2 + 1^2 + 0^2 + 4^2 + 4^2 + 0^2 + 0^2 + \\
 &\quad 1^2 + 1^2 + 0^2 + 2^2 + 0^2 + 1^2 + 1^2 + 1^2 + 1^2 + 0^2 + 1^2 + 6^2 + 7^2 + 1^2 + \\
 &\quad 3^2 + 1^2 + 4^2 + 7^2 + 2^2 + 1^2 + 1^2 + 1^2 + 1^2 + 3^2 + 1^2 + 1^2 + 0^2 + 1^2 + 1^2 + \\
 &\quad 1^2| \\
 &= \frac{1}{42} |0 + 36 + 1 + 0 + 9 + 1 + 1 + 0 + 16 + 16 + 0 + 0 + 1 + 1 + 0 + \\
 &\quad 4 + 0 + 1 + 1 + 1 + 1 + 0 + 1 + 36 + 49 + 1 + 9 + 1 + 16 + 49 + 4 + 1 + \\
 &\quad 1 + 1 + 1 + 9 + 1 + 1 + 0 + 1 + 1 + 1| \\
 &= \frac{1}{42} \times 274 \\
 &= 6.52
 \end{aligned}$$

Berikut perhitungan untuk mencari RMSE dan MAE pada hasil *testing*:

$$\begin{aligned}
 \text{RMSE} &= \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \\
 &= \sqrt{\frac{1}{42} \times 274} \\
 &= 2.55
 \end{aligned}$$

$$\begin{aligned}
 \text{MAE} &= \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \\
 &= \frac{1}{42} \times 72 \\
 &= 1.71
 \end{aligned}$$

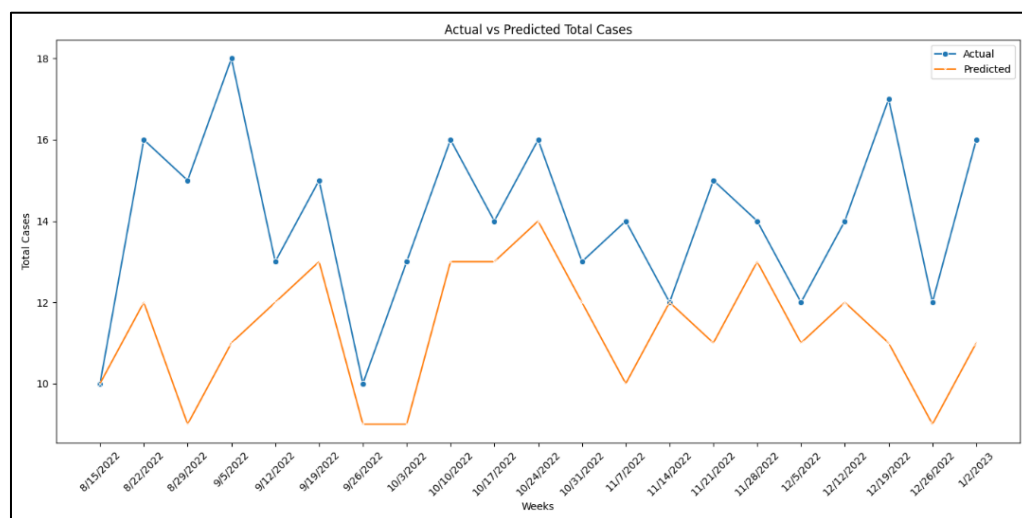
Dari hasil perhitungan di atas, didapatkan hasil pengujian model dengan metrik evaluasi MSE sebesar 6.52, RMSE sebesar 2.55 dan MAE sebesar 1.71. Hal ini menunjukkan bahwa meskipun ada variasi besar antara prediksi dan nilai aktual yang ditunjukkan pada hasil MAE, model mampu memberikan prediksi yang relatif akurat dan dekat dengan nilai aktual dalam skala yang lebih kecil. Selanjutnya, dilakukan pengujian model dengan menggunakan data asli dari penelitian sebelumnya. Tabel 4.10 merupakan hasil pengujian model dengan data asli.

Tabel 4. 10 Hasil Pengujian dengan Data Asli

Weeks	Prediksi	Aktual
15/08/2022	10	10
22/08/2022	12	16
29/08/2022	9	15
05/09/2022	11	18
12/09/2022	12	13
19/09/2022	13	15
26/09/2022	9	10
03/10/2022	9	13
10/10/2022	13	16
17/10/2022	13	14
24/10/2022	14	16
31/10/2022	12	13
07/11/2022	10	14
14/11/2022	12	12
21/11/2022	11	15
28/11/2022	13	14
05/12/2022	11	12

Weeks	Prediksi	Aktual
12/12/2022	12	14
19/12/2022	11	17
26/12/2022	9	12
02/01/2023	11	16

Hasil pengujian dengan menggunakan data asli juga dapat dilihat melalui grafik pada gambar 4. 13.



Gambar 4. 13 Grafik Hasil Prediksi dengan Data Asli

Hasil perbandingan nilai prediksi dan nilai aktual *total case* pada hasil pengujian menggunakan data asli menunjukkan bahwa terdapat beberapa minggu nilai prediksi cukup dekat dengan nilai aktual seperti pada minggu 15/08/2022 dan 11/04/2022. Namun, terdapat juga minggu-minggu yang memiliki perbedaan yang cukup signifikan seperti pada minggu 29/08/2022 dan 19/12/2022 dengan selisih sebesar 4 dan 05/09/2022 dengan selisih sebesar 7. Berdasarkan hasil ini, model pengujian dengan data asli cenderung merespon dengan baik dalam beberapa periode berturut-turut tetapi juga menunjukkan kelemahan dalam beberapa periode lainnya, terutama saat terjadi fluktuasi besar pada data aktual. Dari hasil pengujian dengan data asli, didapatkan nilai metrik evaluasi MSE sebesar 8.65, RMSE sebesar 2.94 dan MAE sebesar 2.28.

4.4. Diskusi

Penelitian ini menggunakan *Gaussian Process Regression* (GPR) untuk memprediksi jumlah kasus malaria di Kabupaten Labuhanbatu Utara. Data yang digunakan pada penelitian ini berasal dari penelitian sebelumnya oleh (Fahmi et al., 2022). Namun, karena keterbatasan data yang didapatkan dari data asli sebesar 210 data, maka dilakukanlah sintesis data dengan menggunakan *Time-Series Generative Adversial Netwrok* (TGAN). Sintesis data dilakukan untuk menambah jumlah data dan memperkaya dataset dengan variabilitas yang lebih tinggi, yang akan membantu model memahami lebih banyak pola. Sehingga didapatkan data sebanyak 418 data yang dimulai dari periode waktu tahun 2015 – 2018. Dalam melihat hubungan linier antara variabel-variabel pada dataset, *Correlation heatmap* digunakan pada penelitian ini dengan hasil bahwa durasi sinar matahari, curah hujan, tingkat banjir dan genangan air, serta kepadatan dan pertumbuhan penduduk menunjukkan korelasi yang lebih signifikan terhadap jumlah kasus malaria. Sementara itu, suhu maksimum dan jumlah hujan menunjukkan korelasi yang sangat rendah, menunjukkan bahwa faktor-faktor ini kurang berpengaruh terhadap jumlah kasus malaria. Hal ini dapat disebabkan adanya faktor-faktor lain yang lebih dominan dalam memengaruhi penyebaran malaria di Labuhanbatu Utara seperti durasi sinar matahari dan lainnya serta kondisi lingkungan dan pola hidup masyarakat Labuhanbatu Utara yang berbeda. Studi yang dilakukan oleh (Okiring et al., 2021) menyatakan bahwa suhu dan jumlah hujan dapat memiliki dampak yang bervariasi terhadap penyebaran malaria. Faktor-faktor ini mungkin tidak selalu menunjukkan korelasi yang kuat dengan kasus malaria, terutama ketika kondisi lingkungan lainnya seperti kelembapan atau vegetasi lebih dominan.

Berbagai *kernel* digunakan dalam model GPR ini, termasuk *Radial Basis Function* (RBF), *Matern*, *Rational Quadratic*, *ExpSine Square*, *White Kernel*, dan *Dot Product*. Setiap *kernel* dieksplorasi untuk menemukan parameter optimal yang memberikan performa terbaik dalam hal akurasi prediksi. Didapatkan bahwa kombinasi *kernel Matern* dengan *kernel Dot Product* pada parameter *nu values* 0.5, *length scale Matern* 0.1 dan *sigma* 0 0.1 menunjukkan bahwa kombinasi *kernel* ini efektif dalam menangkap pola dalam data dengan baik. *Kernel Matern* cocok untuk menangkap pola yang lebih halus dan fleksibel dalam data yang tidak selalu linier. Sedangkan *kernel Dot Product* membantu menangkap hubungan linier yang mungkin ada di beberapa variabel. Sedangkan *kernel-kernel* lain tidak seimbang dalam menangani variasi pola data yang

linier dan non-linier secara bersamaan. Misalnya, *kernel* RBF cenderung lebih efektif untuk menangani pola non-linier, tetapi kurang optimal ketika ada hubungan linier yang signifikan di data. Sebaliknya, *kernel* seperti *White Kernel* terlalu sederhana dan hanya menangani komponen kebisingan, bukan pola utama dalam data. Sehingga kombinasi *kernel Matern* dengan *kernel Dot Product* membuat model menjadi lebih fleksibel dan mampu memprediksi dengan lebih akurat, baik untuk data yang memiliki hubungan linier maupun non-linier. Hal ini sesuai dengan variabel-variabel iklim yang digunakan dalam prediksi kasus malaria menunjukkan pola yang kompleks dan bervariasi.

Evaluasi performa model menggunakan metrik *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE) dan *Mean Absolute Error* (MAE). Pada pengujian model menggunakan kombinasi *kernel* tersebut, didapatkan bahwa model yang dilatih dan diuji dengan menggunakan data sintesis mampu memberikan prediksi yang relatif akurat dan dekat dengan nilai aktual dalam skala yang lebih kecil. Ditunjukkan dengan nilai metrik evaluasi MSE sebesar 6.52, RMSE sebesar 2.55 dan MAE sebesar 1.71. Sedangkan data yang dilatih dan diuji dengan data asli memberikan nilai metrik evaluasi MSE sebesar 8.65, RMSE sebesar 2.94 dan MAE sebesar 2.28. Dengan hasil metrik evaluasi yang lebih baik pada data sintesis, menunjukkan bahwa GPR lebih cocok digunakan pada data yang berjumlah banyak. GPR mengandalkan *kernel* untuk mengukur korelasi antara titik data. Dengan lebih banyak data, *kernel* dapat mempelajari hubungan yang lebih kaya dan memberikan prediksi yang lebih akurat. Hal ini sesuai dengan penjelasan (Rasmussen & Williams, n.d.) yang menjelaskan bagaimana GPR dapat memanfaatkan jumlah data yang lebih besar untuk meningkatkan akurasi prediksi, asalkan *kernel* yang dipilih mampu menangkap struktur data dengan baik.

Penggunaan metode GPR pada penelitian ini memberikan hasil prediksi yang cukup akurat meskipun terdapat variabilitas dalam data. GPR mampu menangani variabilitas ini dengan baik karena model ini dapat memahami perubahan dalam pola data. Model bisa menangkap hubungan antara fluktuasi iklim dengan jumlah kasus malaria, sehingga prediksinya tetap akurat meskipun data iklimnya bervariasi. Namun, penggunaan *kernel* pada metode GPR perlu dilakukan *trial* dan *error* beberapa kali untuk menemukan *kernel* yang tepat dan sesuai dengan dataset. Jika *kernel* yang digunakan tidak cocok dengan data, maka hasil prediksinya bisa tidak akurat. Pemilihan

kernel yang optimal bisa menjadi tantangan tersendiri. Selain itu, tidak ada satu *kernel* yang secara universal cocok untuk semua jenis data. Sehingga perlu mencoba berbagai jenis *kernel* agar bisa menemukan kombinasi yang memberikan performa terbaik.

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Setelah melakukan proses penelitian untuk memprediksi jumlah kasus malaria di Labuhanbatu Utara dengan menggunakan *Gaussian Process Regression* (GPR) dapat diambil kesimpulan sebagai berikut.

1. Model *Gaussian Process Regression* (GPR) pada kombinasi *kernel Matern* dan *Dot Product* dengan parameter *nu values* 0.5, *length scale Matern* 0.1 dan *sigma* 0 0.1 yang digunakan dalam penelitian ini menunjukkan hasil prediksi yang cukup akurat terhadap data kasus malaria. Didapatkan bahwa model yang dilatih dan diuji dengan menggunakan data sintesis mampu memberikan prediksi yang relatif akurat dan dekat dengan nilai aktual dalam skala yang lebih kecil. Ditunjukkan dengan nilai metrik evaluasi MSE sebesar 6.52, RMSE sebesar 2.55 dan MAE sebesar 1.71. Sedangkan data yang dilatih dan diuji dengan data asli memberikan nilai metrik evaluasi MSE sebesar 8.65, RMSE sebesar 2.94 dan MAE sebesar 2.28. Hal ini menunjukkan bahwa GPR lebih cocok digunakan dengan jumlah data yang lebih besar sehingga dapat meningkatkan akurasi prediksi.
2. Penggunaan *Time-Series Generative Adversial Network* (TGAN) dapat membantu menyintesis data malaria berdasarkan data asli. TGAN mampu menghasilkan data sintesis yang cukup akurat dengan mempertimbangkan ketergantungan temporal yang ada dalam dataset asli.
3. Jumlah kasus malaria menunjukkan korelasi yang signifikan pada faktor lingkungan seperti durasi sinar matahari, curah hujan, tingkat banjir dan genangan air, serta kepadatan dan pertumbuhan penduduk. Sementara itu, suhu maksimum dan jumlah hujan menunjukkan korelasi yang sangat rendah, yang berarti faktor-faktor ini kurang berpengaruh terhadap jumlah kasus malaria.

5.2 Saran

Dalam meningkatkan kualitas penelitian, dapat dituliskan saran untuk pengembangan penelitian selanjutnya. Adapun saran yang dapat dilakukan adalah sebagai berikut:

1. Penelitian berikutnya dapat menggunakan penggabungan *kernel* yang lebih kompleks pada metode *Gaussian Process Regression* (GPR) untuk menghasilkan nilai prediksi yang lebih baik dan akurat.
2. Data yang digunakan dapat ditingkatkan lagi dalam segi kuantitas data yang digunakan dalam pelatihan model. Sehingga pengumpulan data yang lebih lengkap dan representatif dapat meningkatkan akurasi dan generalisasi model prediksi.
3. Penelitian lebih lanjut dapat dilakukan dengan mengintegrasikan faktor-faktor lain yang mempengaruhi penyebaran malaria seperti mobilitas penduduk dan intervensi kesehatan masyarakat untuk meningkatkan akurasi prediksi.
4. Penelitian lebih lanjut dapat dilakukan dengan mengeksplorasi *kernel* lain atau kombinasi beberapa *kernel* yang mungkin memberikan hasil yang lebih baik. *Kernel-kernel* yang belum dieksplorasi dalam penelitian ini mungkin memiliki potensi untuk meningkatkan performa model.

DAFTAR PUSTAKA

- Achmadi, U. F. (2007). Dampak Perubahan Iklim Dalam Perspektif Kesehatan Lingkungan. *Seminar Kipnas IX Jakarta*.
- 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>
- AK Githeko, EN ototo, Y. G. (2012). Progress towards understanding the ecology and epidemiology of malaria in the western Kenya highlands: opportunities and challenges for control under climate change risk. *J Neurochem*, 4(1), 1–15. <https://doi.org/10.1016/j.actatropica.2011.10.002>.Progress
- Apriliani. (2021). ANALISIS FAKTOR RISIKO KEJADIAN MALARIA DI INDONESIA (Analisis Data Riskesdas 2018). *Data Riskesdas 2018*, 10–25.
- Arsin, A. A. (2011). Malaria di Indonesia: Tinjauan Aspek Epidemiologi. In *Masagena Press*. <https://doi.org/10.1532/HSF98.S001S119>
- Asante-Okyere, S., Shen, C., Ziggah, Y. Y., Rulegeya, M. M., & Zhu, X. (2018). Investigating the predictive performance of *Gaussian Process* regression in evaluating reservoir porosity and permeability. *Energies*, 11(12). <https://doi.org/10.3390/en11123261>
- Avichena, A., & Anggriyani, R. (2023). Analisis Penyakit Malaria Akibat Infeksi Plasmodium sp. terhadap Darah Manusia. *EKOTONIA: Jurnal Penelitian Biologi, Botani, Zoologi Dan Mikrobiologi*, 8(1), 30–37. <https://doi.org/10.33019/ekotonia.v8i1.4128>
- Blanford, J. I., Blanford, S., Crane, R. G., Mann, M. E., Paaijmans, K. P., Schreiber, K. V., & Thomas, M. B. (2013). Implications of temperature variation for malaria parasite development across Africa. *Scientific Reports*, 3, 1–11. <https://doi.org/10.1038/srep01300>
- Cheng, L., Ramchandran, S., Vatanen, T., Lietzén, N., Lahesmaa, R., Vehtari, A., & Lähdesmäki, H. (2019). An additive *Gaussian Process* regression model for interpretable non-parametric analysis of longitudinal data. *Nature Communications*, 10(1), 1–11. <https://doi.org/10.1038/s41467-019-09785-8>

- Dahliani, E. R., Rahmatan, H., & Djufri. (2020). The correlation between students' interest and learning outcomes in biology. *Journal of Physics: Conference Series*, 1460(1). <https://doi.org/10.1088/1742-6596/1460/1/012072>
- Ding, Y. (2018). Control and Optimization of soft exosuit to improve the efficiency of human walking. *Angewandte Chemie International Edition*, 6(11), 951–952., 2013–2015.
- Dong, H., Mao, L., Qu, K., Zhao, J., Li, F., & Jiang, L. (2022). State of Health Estimation for Li-ion Batteries using Improved *Gaussian Process* Regression and Multiple Health Indicators. *International Journal of Electrochemical Science*, 17, 1–15. <https://doi.org/10.20964/2022.08.34>
- Elyazar, I. R. F., Hay, S. I., & Baird, J. K. (2011). Malaria distribution, prevalence, drug resistance and control in Indonesia. In *Advances in Parasitology* (1st ed., Vol. 74, Issue C). Elsevier Ltd. <https://doi.org/10.1016/B978-0-12-385897-9.00002-1>
- Eric. (2023, May 2). *Understanding Cross-Validation*. Aptech.
- 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), 1–14. <https://doi.org/10.1186/s12936-022-04262-y>
- Fischer, L., Gültekin, N., Kaelin, M. B., Fehr, J., & Schlagenhauf, P. (2020). Rising temperature and its impact on receptivity to malaria transmission in Europe: A systematic review. *Travel Medicine and Infectious Disease*, 36, 101815. <https://doi.org/10.1016/j.tmaid.2020.101815>
- Global Health, D. of P. D. and M. (2023). *Symptoms of Malaria*. https://www.cdc.gov/malaria/about/symptoms_malaria.html
- Health Organization, W. (2018). *Malaria Surveillance, Monitoring & Evaluation: a Reference Manual*.
- Health Organization, W. (2022). *SDG Target 3.3 Communicable diseases*. https://www.who.int/data/gho/data/themes/topics/sdg-target-3_3-communicable-diseases
- Health Organization, W. (2022). *World malaria report 2022*.
- Irwan. (2016). Epidemiologi Penyakit Menular. In *Pengaruh Kualitas Pelayanan... Jurnal EMBA* (Vol. 109, Issue 1).
- Islamiyati, A. (2009). *Pemilihan Knot Optimal dalam Estimator Spline Terbobot Pada Regresi Nonparametrik Heteroskedastik Data Longitudinal*. November, 1–11.

- Kesehatan, K. (2022). *Wilayah-wilayah Endemis Malaria Tinggi di Indonesia*.
<https://p2pm.kemkes.go.id/publikasi/artikel/wilayah-wilayah-endemis-malaria-tinggi-di-indonesia>
- KHAYRIAH, F. (2020). *PERSENTASE MALARIA DI PUSKESMAS TARUSAN TAHUN 2016 Oleh : FILZA KHAYRIAH*.
- Makinde, O. S., Abiodun, G. J., & Ojo, O. T. (2021). Modelling of malaria incidence in Akure, Nigeria: negative binomial approach. *GeoJournal*, 86(3), 1327–1336.
<https://doi.org/10.1007/s10708-019-10134-x>
- 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>
- Naveed, M. H., Hashmi, U. S., Tajved, N., Sultan, N., & Imran, A. (n.d.). *Assessing Deep Generative Models on Time Series Network Data*.
<https://doi.org/10.1109/ACCESS.2022.DOI>
- Nkiruka, O., Prasad, R., & Clement, O. (2021). Prediction of malaria incidence using climate variability and machine learning. *Informatics in Medicine Unlocked*, 22, 100508.
<https://doi.org/10.1016/j.imu.2020.100508>
- Nugraha, W., & Sasongko, A. (2022). Hyperparameter Tuning on Classification Algorithm with Grid Search. *SISTEMASI*, 11(2), 391. <https://doi.org/10.32520/stmsi.v11i2.1750>
- Oh, E., & Lee, H. (2020). An imbalanced data handling framework for industrial big data using a *Gaussian Process* regression-based generative adversarial network. *Symmetry*, 12(4). <https://doi.org/10.3390/SYM12040669>
- Okiring, J., Routledge, I., Epstein, A., Namuganga, J. F., Kamya, E. V., Obeng-Amoako, G. O., Sebuguzi, C. M., Rutazaana, D., Kalyango, J. N., Kamya, M. R., Dorsey, G., Wesonga, R., Kiwuwa, S. M., & Nankabirwa, J. I. (2021). Associations between environmental covariates and temporal changes in malaria incidence in high transmission settings of Uganda: a distributed lag nonlinear analysis. *BMC Public Health*, 21(1). <https://doi.org/10.1186/s12889-021-11949-5>
- Penyakit, D. J. P. dan P. (2022). *Mengenal Malaria: Penyakit Mematikan Dunia*.
<https://p2pm.kemkes.go.id/publikasi/artikel/mengenal-malaria-penyakit-mematikan-dunia>
- Putra, T. R. I. (2011). Malaria dan Permasalahannya. *The FASEB Journal*, 11(S1), 103–114.
https://doi.org/10.1096/fasebj.2018.32.1_supplement.819.17

- Putri, I. M. E. (2018). *Regresi Nonparametrik Spline Truncated Untuk Memodelkan Persentase Peserta Kb Baru Pasca Persalinan Dan Pasca Keguguran Di Jawa Timur Tahun 2016 (Tugas Akhir)*. 1–96.
- Rasmussen, C. E., & Williams, C. K. I. (n.d.). *Gaussian Processes for Machine Learning*. www.GaussianProcess.org/gpml
- Roser, M. (2022). *Malaria*. <https://ourworldindata.org/malaria-introduction>
- 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), 1–11. <https://doi.org/10.1186/s12936-018-2220-x>
- Santosh, T., Ramesh, D., & Reddy, D. (2020). LSTM based prediction of malaria abundances using big data. *Computers in Biology and Medicine*, 124(June), 103859. <https://doi.org/10.1016/j.combiomed.2020.103859>
- Sulasmis, S., Setyaningtyas, D. E., Rosanji, A., & Rahayu, N. (2019). Pengaruh curah hujan, kelembaban, dan temperatur terhadap prevalensi Malaria di Kabupaten Tanah Bumbu Kalimantan Selatan. *Journal of Health Epidemiology and Communicable Diseases*, 3(1), 22–27. <https://doi.org/10.22435/jhecdis.v3i1.1794>
- Sun, A. Y., Wang, D., & Xu, X. (2014). Monthly streamflow forecasting using *Gaussian Process* regression. *Journal of Hydrology*, 511(January), 72–81. <https://doi.org/10.1016/j.jhydrol.2014.01.023>
- Suresh, S. D., Qasim, A., Lal, B., Imran, S. M., & Foo, K. S. (2021). Application of *Gaussian Process* Regression (GPR) in Gas Hydrate Mitigation. *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, 88(2), 27–37. <https://doi.org/10.37934/arfmts.88.2.2737>
- Taki, M., Rohani, A., Soheili-Fard, F., & Abdeslahi, A. (2016). Assessment of energy consumption and modeling of output energy for wheat production by neural network (MLP and RBF) and *Gaussian Process* regression (GPR) models. *Journal of Cleaner Production*, 172, 3028–3041. <https://doi.org/10.1016/j.jclepro.2017.11.107>
- Tazi, K., Lin, J. A., Viljoen, R., Gardner, A., John, S., Ge, H., & Turner, R. E. (2023). *Beyond Intuition, a Framework for Applying GPs to Real-World Data. 1*. <http://arxiv.org/abs/2307.03093>
- Vicalina, A. (2023). *PREDIKSI KEJADIAN MALARIA DENGAN PEMANFAATAN DATA METEOROLOGI MENGGUNAKAN GATED RECURRENT UNIT*.

- Vidya, G. S., & Hari, V. S. (2023). Prediction of Bus Passenger Traffic using *Gaussian Process Regression*. *Journal of Signal Processing Systems*, 95(2–3), 281–292. <https://doi.org/10.1007/s11265-022-01774-3>
- Wang, M., Wang, H., Wang, J., Liu, H., Lu, R., Duan, T., Gong, X., Feng, S., Liu, Y., Cui, Z., Li, C., & Ma, J. (2019). A novel model for malaria prediction based on ensemble algorithms. *PLoS ONE*, 14(12). <https://doi.org/10.1371/journal.pone.0226910>
- Widi, S. (2022). *Kasus Malaria Indonesia Melonjak 36,29% pada 2022*. <https://dataindonesia.id/kesehatan/detail/kasus-malaria-indonesia-melonjak-3629-pada-2022>
- Yadav, S. S., Kadam, V. J., Jadhav, S. M., Jagtap, S., & Pathak, P. R. (2021). Machine learning based malaria prediction using clinical findings. *2021 International Conference on Emerging Smart Computing and Informatics, ESCI 2021, Lmm*, 216–222. <https://doi.org/10.1109/ESCI50559.2021.9396850>
- Yoon, J., Jarrett, D., & Van Der Schaar, M. (n.d.). *Time-series Generative Adversarial Networks*.
- Zhao, J., Guo, H., Han, M., Tang, H., & Li, X. (2019). *Gaussian Process regression for prediction of sulfate content in lakes of China*. *Journal of Engineering and Technological Sciences*, 51(2), 198–215. <https://doi.org/10.5614/j.eng.technol.sci.2019.51.2.4>