



Prediksi Risiko Demam Berdarah Dengue Menggunakan Algoritma Random Forest Berbasis GIS di Kota Tangerang

Muhammad Fikri*, Jefry Sunupurwa Asri, Agus Herwanto

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Esa Unggul, Jakarta, Indonesia

Email: ^{1,*}muhammadfikri77xc@student.esaunggul.ac.id, ²jefry.sunupurwa@esaunggul.ac.id, ³agus.herwanto@esaunggul.ac.id

Email Penulis Korespondensi: muhammadfikri77xc@student.esaunggul.ac.id

Abstrak—Demam Berdarah Dengue (DBD) masih menjadi salah satu tantangan kesehatan masyarakat di Indonesia karena penyebarannya dipengaruhi oleh berbagai faktor, seperti kondisi iklim, lingkungan, kepadatan penduduk, dan karakteristik spasial wilayah. Identifikasi dini terhadap wilayah berisiko diperlukan untuk mendukung upaya pencegahan dan pengendalian penyakit secara lebih efektif. Penelitian ini bertujuan mengembangkan model prediksi risiko DBD berbasis Geographic Information System (GIS) menggunakan algoritma Random Forest di Kota Tangerang. Penelitian menggunakan data sekunder berupa data historis kasus DBD dari Dinas Kesehatan Kota Tangerang, data curah hujan, suhu, kelembapan, fasilitas kesehatan, dan batas administrasi wilayah selama periode 2023–2025. Data kasus DBD yang digunakan merupakan data kasus yang tercatat pada Dinas Kesehatan Kota Tangerang dan telah melalui pemeriksaan kelengkapan serta konsistensi data. Tahapan penelitian meliputi prapemrosesan data, rekayasa fitur, integrasi data spasial, pembangunan model, dan evaluasi menggunakan *Spatial Cross-Validation* untuk mengurangi potensi bias akibat autokorelasi spasial serta mengukur kemampuan generalisasi model. Hasil evaluasi berbasis waktu menunjukkan nilai Area Under the Curve (AUC) sebesar 0,973 dan F1-Score sebesar 0,889, sedangkan evaluasi menggunakan *Spatial Cross-Validation* menghasilkan AUC sebesar 0,956 dan F1-Score sebesar 0,892. Hasil tersebut menunjukkan bahwa model mempertahankan performa yang baik ketika diuji pada wilayah geografis yang berbeda. Selain itu, dashboard berbasis GIS yang dikembangkan mampu memvisualisasikan distribusi risiko DBD secara interaktif sehingga dapat mendukung identifikasi wilayah prioritas dan pengambilan keputusan berbasis data dalam upaya pencegahan dan pengendalian DBD di Kota Tangerang.

Kata Kunci: Demam Berdarah Dengue; Sistem Informasi Geografis; Prediksi Risiko; Random Forest; Spatial Cross-Validation

Abstract—Dengue Hemorrhagic Fever (DHF) remains a major public health challenge in Indonesia, as its transmission is influenced by various factors, including climate conditions, environmental characteristics, population density, and spatial characteristics. Early identification of high-risk areas is needed to support more effective dengue prevention and control strategies. This study aims to develop a Geographic Information System (GIS)-based dengue risk prediction model using the Random Forest algorithm in Tangerang City. The study utilized secondary data consisting of historical dengue case records obtained from the Tangerang City Health Office, rainfall, temperature, humidity, healthcare facilities, and administrative boundary data for the 2023–2025 period. The dengue case data consisted of recorded cases from the Tangerang City Health Office and underwent completeness and consistency checks. The research process included data preprocessing, feature engineering, spatial data integration, model development, and evaluation using *Spatial Cross-Validation* to reduce potential bias caused by spatial autocorrelation and assess the model's generalization capability. The time-based evaluation resulted in an Area Under the Curve (AUC) of 0.973 and an F1-score of 0.889, while the *Spatial Cross-Validation* evaluation produced an AUC of 0.956 and an F1-score of 0.892. These results indicate that the model maintained good performance when evaluated across different geographic areas. In addition, the developed GIS-based dashboard provides interactive visualization of dengue risk distribution to support the identification of priority areas and data-driven decision-making for dengue prevention and control in Tangerang City.

Keywords: Dengue Hemorrhagic Fever; Geographic Information System; Risk Prediction; Random Forest; Spatial Cross-Validation

1. PENDAHULUAN

Demam Berdarah Dengue (DBD) masih menjadi salah satu penyakit menular yang menjadi tantangan utama kesehatan masyarakat di berbagai negara tropis, termasuk Indonesia. Penyakit ini ditularkan oleh nyamuk *Aedes aegypti* dan penyebarannya dipengaruhi oleh berbagai faktor, seperti kondisi iklim, lingkungan, kepadatan penduduk, serta karakteristik spasial suatu wilayah. Menurut *World Health Organization* (WHO), hampir separuh populasi dunia berisiko terinfeksi virus dengue, dengan jumlah kasus yang terus meningkat dalam beberapa dekade terakhir [1]. Indonesia merupakan salah satu negara dengan beban kasus DBD yang tinggi karena kondisi iklim tropis, pertumbuhan kawasan perkotaan, kepadatan penduduk, serta perubahan lingkungan yang mendukung perkembangan vektor penyakit [2], [3]. Kondisi tersebut menunjukkan perlunya pendekatan prediksi yang mampu mengidentifikasi wilayah berisiko secara lebih dini sebagai dasar penyusunan strategi pencegahan dan pengendalian DBD.

Penyebaran DBD dipengaruhi oleh berbagai faktor lingkungan dan meteorologi, seperti curah hujan, suhu udara, dan kelembapan. Variabel tersebut berperan dalam memengaruhi kondisi lingkungan, perkembangan vektor, dan tingkat penularan virus dengue sehingga sering digunakan dalam pemodelan risiko DBD [4]–[7]. Hubungan antara faktor iklim dan risiko DBD tidak selalu bersifat linear. Perubahan suhu dapat memengaruhi perkembangan vektor dan periode inkubasi ekstrinsik virus dengue di dalam tubuh nyamuk, sedangkan curah hujan dan kelembapan dapat memengaruhi ketersediaan tempat perkembangbiakan serta kelangsungan hidup vektor. Akibatnya, perubahan suatu variabel iklim tidak selalu menghasilkan perubahan risiko DBD dalam proporsi yang sama. Kondisi tersebut menjadi pertimbangan dalam pemilihan Random Forest karena algoritma ini mampu menangkap hubungan non-linear dan interaksi antarvariabel tanpa memerlukan asumsi hubungan linear antara variabel prediktor dan risiko DBD. Yavari Nejad dan Varathan (2021) menunjukkan bahwa faktor iklim dapat digunakan untuk mengidentifikasi faktor risiko dan membangun model prediksi



dengue [4]. Kakarla et al. (2023) serta Ong et al. (2023) juga menunjukkan bahwa data meteorologi dapat dikombinasikan dengan metode *machine learning* untuk memprediksi prevalensi dan transmisi dengue [5], [7].

Perkembangan *Geographic Information System* (GIS) memberikan kontribusi dalam bidang epidemiologi karena mampu mengintegrasikan data spasial dan nonspasial untuk memetakan distribusi penyakit, mengidentifikasi wilayah berisiko, serta mendukung pengambilan keputusan berbasis lokasi [8], [9]. Nayak et al. (2025) menunjukkan bahwa GIS banyak digunakan dalam surveilans dengue untuk mengintegrasikan informasi epidemiologi dengan faktor geografis dan lingkungan [8]. Sylvestre et al. (2022) melalui kajian sistematis menunjukkan bahwa berbagai metode berbasis data dan *machine learning* telah digunakan untuk prediksi dan surveilans dengue [10]. Anggraeni et al. (2024) juga mengembangkan pendekatan berbasis *machine learning* untuk peramalan kasus DBD dengan mempertimbangkan karakteristik data deret waktu [11]. Selain itu, Chen dan Moraga (2025) menunjukkan bahwa metode *machine learning* dengan kovariat iklim dapat digunakan dalam peramalan dengue [12], sedangkan Satoto et al. (2026) menerapkan Random Forest dengan faktor iklim, lingkungan, dan sosial-demografi untuk prediksi dengue di Indonesia [13]. Penelitian-penelitian tersebut menunjukkan bahwa pendekatan berbasis data dan *machine learning* memiliki potensi dalam prediksi dengue, tetapi fokus dan unit analisisnya berbeda. Oleh karena itu, penelitian ini mengintegrasikan Random Forest dan GIS pada tingkat kelurahan serta menggunakan *Spatial Cross-Validation* untuk mengevaluasi kemampuan generalisasi model antarwilayah.

Aspek spasial menjadi penting dalam penelitian DBD karena wilayah yang berdekatan dapat memiliki kondisi lingkungan dan pola penyebaran penyakit yang serupa. Kondisi tersebut dapat menimbulkan autokorelasi spasial dan memengaruhi hasil evaluasi apabila data pelatihan dan pengujian berasal dari wilayah yang saling berdekatan. Dalam konteks penelitian ini, Kota Tangerang terdiri atas 13 kecamatan dan 104 kelurahan, sehingga keterhubungan geografis antarwilayah, termasuk kelurahan yang berdekatan atau berbatasan, perlu dipertimbangkan dalam proses evaluasi model. Jika data dari wilayah yang memiliki karakteristik serupa masuk secara bersamaan ke data pelatihan dan pengujian, performa model berpotensi terlihat lebih tinggi dibandingkan kemampuan generalisasinya pada wilayah yang berbeda.

Untuk mengatasi permasalahan tersebut, *Spatial Cross-Validation* digunakan sebagai evaluasi tambahan dengan membagi data berdasarkan pengelompokan spasial sehingga kemampuan model dapat diuji pada wilayah geografis yang berbeda. Wang et al. (2023) menunjukkan pentingnya mempertimbangkan hubungan geografis dalam evaluasi model *machine learning* geospasial [14], sedangkan Linnenbrink et al. (2024) menekankan pentingnya mempertimbangkan jarak antarobservasi dalam evaluasi model spasial [15]. Dengan demikian, *Spatial Cross-Validation* dalam penelitian ini digunakan untuk melengkapi evaluasi berbasis waktu dan memberikan gambaran mengenai kemampuan generalisasi Random Forest terhadap karakteristik wilayah yang berbeda.

Berdasarkan penelitian terdahulu, faktor iklim dan metode *machine learning* telah banyak digunakan dalam prediksi DBD [4]-[7], [10]-[13], sedangkan GIS banyak dimanfaatkan untuk pemetaan dan surveilans penyakit [8], [9]. Namun, integrasi Random Forest, GIS, dan *Spatial Cross-Validation* pada tingkat kelurahan, khususnya di Kota Tangerang, masih terbatas. Oleh karena itu, penelitian ini mengintegrasikan data historis kasus DBD, faktor iklim, fasilitas kesehatan, dan informasi spasial dalam model Random Forest, kemudian mengevaluasinya menggunakan *Spatial Cross-Validation*. Hasil prediksi selanjutnya divisualisasikan melalui dashboard GIS interaktif untuk memberikan informasi mengenai wilayah yang perlu mendapatkan perhatian lebih.

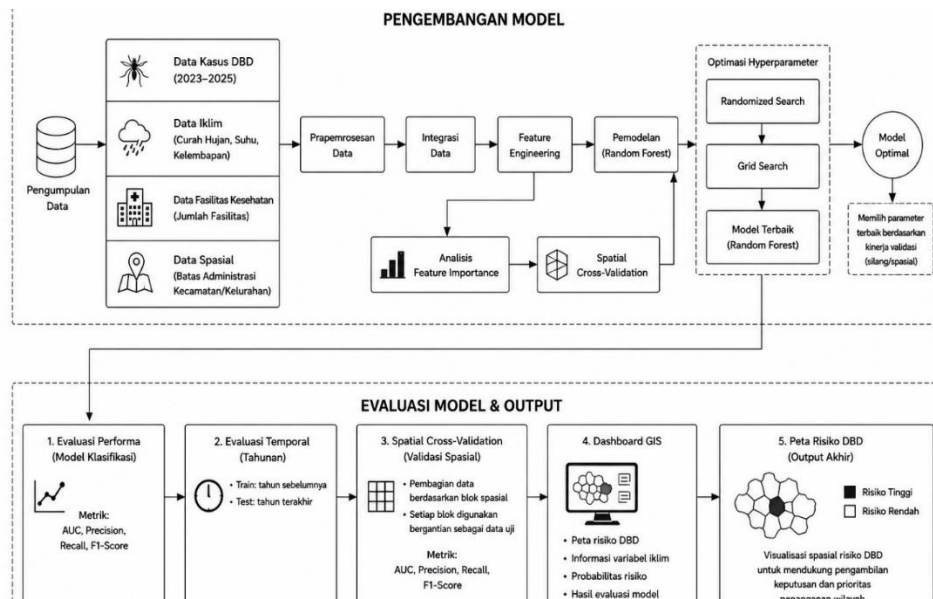
Berdasarkan kondisi tersebut, penelitian ini bertujuan mengembangkan model prediksi risiko Demam Berdarah Dengue berbasis *Geographic Information System* (GIS) menggunakan algoritma Random Forest di Kota Tangerang. Model dibangun menggunakan data historis kasus DBD, curah hujan, suhu, kelembapan, fasilitas kesehatan, serta batas administrasi wilayah selama periode 2023–2025. Tahapan penelitian meliputi prapemrosesan data, integrasi data spasial, rekayasa fitur, pembangunan model, serta evaluasi menggunakan *Spatial Cross-Validation* untuk mengurangi potensi bias akibat autokorelasi spasial dan mengukur kemampuan generalisasi model. Hasil prediksi kemudian divisualisasikan melalui dashboard GIS interaktif untuk mendukung identifikasi wilayah prioritas serta pengambilan keputusan berbasis data dalam upaya pencegahan dan pengendalian DBD di Kota Tangerang.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang saling berkaitan, mulai dari pengumpulan data hingga menghasilkan visualisasi hasil prediksi DBD berbasis Sistem Informasi Geografis (GIS). Secara umum, tahapan penelitian meliputi pengumpulan data, prapemrosesan data, integrasi data, *feature engineering*, pemodelan Random Forest, optimasi *hyperparameter*, analisis *feature importance*, evaluasi model, dan visualisasi hasil. Alur keseluruhan tahapan penelitian ditunjukkan pada Gambar 1.

Gambar 1 menunjukkan bahwa penelitian diawali dengan pengumpulan data yang terdiri atas data kasus DBD periode 2023–2025, data iklim yang meliputi curah hujan, suhu, dan kelembapan, data fasilitas kesehatan, serta data spasial berupa batas administrasi kecamatan dan kelurahan. Data yang diperoleh selanjutnya diproses melalui tahap prapemrosesan data untuk memastikan keseragaman format dan struktur data. Setelah itu, dilakukan integrasi data untuk menggabungkan informasi epidemiologi, iklim, fasilitas kesehatan, dan spasial berdasarkan wilayah dan periode pengamatan.



Gambar 1. Tahapan Penelitian

Tahap selanjutnya adalah *feature engineering* untuk membentuk variabel yang digunakan sebagai masukan model. Fitur yang dibentuk mencakup informasi historis kasus DBD, karakteristik iklim, fasilitas kesehatan, serta fitur turunan dan interaksi antarvariabel. Dataset hasil *feature engineering* kemudian digunakan pada tahap pemodelan Random Forest.

Model Random Forest selanjutnya melalui proses optimasi *hyperparameter* menggunakan *Randomized Search* dan *Grid Search*. Proses tersebut dilakukan untuk memperoleh konfigurasi parameter yang memberikan kinerja validasi terbaik. Model dengan konfigurasi terbaik kemudian digunakan dalam proses analisis dan evaluasi. Pada tahap ini juga dilakukan analisis *feature importance* untuk mengetahui kontribusi relatif variabel yang digunakan dalam pembentukan model.

Tahap evaluasi terdiri atas evaluasi performa model, evaluasi temporal, dan *Spatial Cross-Validation*. Evaluasi performa dilakukan menggunakan metrik *Area Under the Curve* (AUC), *Precision*, *Recall*, dan *F1-Score*. Evaluasi temporal dilakukan dengan menggunakan data tahun sebelumnya sebagai data pelatihan dan tahun terakhir sebagai data pengujian. Sementara itu, *Spatial Cross-Validation* dilakukan dengan membagi data berdasarkan blok spasial dan menggunakan setiap blok secara bergantian sebagai data pengujian. Pendekatan tersebut digunakan untuk menilai kemampuan generalisasi model pada wilayah geografis yang berbeda.

Tahap akhir penelitian adalah implementasi hasil pemodelan ke dalam dashboard GIS. Dashboard digunakan untuk menampilkan hasil prediksi, informasi variabel iklim, probabilitas risiko, dan hasil evaluasi model. Hasil pemodelan selanjutnya divisualisasikan dalam bentuk peta spasial untuk menunjukkan perbedaan hasil prediksi antarwilayah. Visualisasi tersebut digunakan sebagai informasi pendukung untuk mengidentifikasi wilayah yang perlu mendapatkan perhatian dan membantu pengambilan keputusan dalam upaya pencegahan dan pengendalian DBD di Kota Tangerang.

2.2 Pengumpulan dan Integrasi Data

Pengumpulan data pada penelitian ini bertujuan menyediakan informasi yang diperlukan untuk analisis geospasial dan pembangunan model prediksi risiko Demam Berdarah Dengue (DBD) di Kota Tangerang. Data yang digunakan merupakan data sekunder yang terdiri atas data kasus DBD, data iklim, data fasilitas kesehatan, dan data spasial wilayah administrasi. Penggunaan beberapa jenis data tersebut ditujukan untuk merepresentasikan kondisi epidemiologi, lingkungan, fasilitas kesehatan, dan karakteristik geografis wilayah secara terpadu.

Data kasus DBD diperoleh dari Dinas Kesehatan Kota Tangerang dan digunakan sebagai data historis jumlah kasus pada tingkat kelurahan selama periode 2023–2025. Data iklim diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), yang meliputi curah hujan, suhu, dan kelembapan. Data iklim digunakan untuk merepresentasikan kondisi meteorologi yang berkaitan dengan karakteristik lingkungan wilayah. Selain itu, penelitian ini menggunakan data fasilitas kesehatan berupa jumlah fasilitas kesehatan pada tingkat kecamatan sebagai salah satu variabel pendukung dalam pemodelan.

Data spasial yang digunakan berupa batas administrasi kelurahan Kota Tangerang. Data tersebut berfungsi sebagai dasar integrasi informasi tabular dengan lokasi geografis serta digunakan dalam proses visualisasi hasil prediksi melalui dashboard GIS. Integrasi dilakukan berdasarkan identitas wilayah kecamatan dan kelurahan. Karena data kasus DBD tersedia pada tingkat kelurahan, sedangkan data iklim dan fasilitas kesehatan tersedia pada tingkat kecamatan, penyelarasan dilakukan berdasarkan hubungan administratif. Data iklim dan fasilitas kesehatan pada suatu kecamatan digunakan sebagai informasi kontekstual bagi kelurahan yang berada dalam kecamatan tersebut. Dengan pendekatan ini, kelurahan dalam kecamatan yang sama memperoleh nilai iklim dan fasilitas kesehatan yang sama. Perbedaan resolusi spasial tersebut diperlakukan sebagai keterbatasan data dan dipertimbangkan dalam interpretasi hasil penelitian.



Seluruh data yang digunakan telah melalui pemeriksaan awal untuk memastikan kelengkapan, konsistensi, kesesuaian periode pengamatan, dan kesesuaian identitas wilayah sebelum dilakukan proses integrasi. Data iklim yang memiliki resolusi bulanan selanjutnya diselaraskan dengan data kasus DBD yang tersedia secara tahunan, sedangkan data fasilitas kesehatan pada tingkat kecamatan dihubungkan dengan wilayah kelurahan berdasarkan hubungan administratif. Sumber dan karakteristik data yang digunakan dalam penelitian ini disajikan pada Tabel 1.

Tabel 1. Sumber dan Karakteristik Data Penelitian

No.	Jenis Data	Sumber Data	Bentuk Data	Unit Analisis	Keterangan
1	Data kasus DBD	Dinas Kesehatan Kota Tangerang	Tabular	Kelurahan	Jumlah kasus tahun 2023–2025
2	Data iklim	BMKG	Tabular	Kecamatan	Curah hujan, suhu, dan kelembapan per bulan
3	Data fasilitas kesehatan	BPS/Dinas Kesehatan	Tabular	Kecamatan	Jumlah fasilitas kesehatan
4	Data spasial	Data administrasi/GeoJSON	Geospasial	Kelurahan	Batas administrasi wilayah

Sumber dan karakteristik data pada Tabel 1 menunjukkan bahwa penelitian menggabungkan data dengan karakteristik spasial dan temporal yang berbeda. Data kasus DBD dan data spasial memiliki unit analisis kelurahan, sedangkan data iklim dan fasilitas kesehatan tersedia pada tingkat kecamatan. Perbedaan unit analisis tersebut menimbulkan perbedaan resolusi spasial antara variabel prediktor dan data kasus yang perlu diperhatikan dalam proses integrasi. Kondisi ini berkaitan dengan potensi *Modifiable Areal Unit Problem* (MAUP), karena informasi iklim pada tingkat kecamatan tidak secara langsung menggambarkan variasi kondisi pada setiap kelurahan. Dalam penelitian ini, data tingkat kecamatan digunakan sebagai variabel kontekstual berdasarkan hubungan administratif dan tidak dianggap sebagai pengukuran langsung kondisi masing-masing kelurahan.

Pada tahap integrasi, data iklim yang memiliki resolusi bulanan diagregasikan menjadi informasi tahunan agar sesuai dengan periode data kasus DBD. Selanjutnya, data iklim dan fasilitas kesehatan pada tingkat kecamatan dikaitkan dengan data kelurahan berdasarkan hubungan administratif. Dengan demikian, kelurahan yang berada dalam kecamatan yang sama menggunakan informasi iklim dan fasilitas kesehatan pada tingkat kecamatan yang sama. Pendekatan tersebut digunakan secara konsisten untuk menjaga kesesuaian struktur data dan menghindari interpretasi bahwa nilai iklim kecamatan merupakan pengukuran langsung pada tingkat kelurahan. Data tabular yang telah diselaraskan kemudian digabungkan dengan data spasial kelurahan sehingga terbentuk dataset terpadu yang digunakan untuk tahap *feature engineering*, pemodelan Random Forest, evaluasi spasial, dan visualisasi GIS.

2.3 Prapemrosesan dan *Feature Engineering*

Prapemrosesan dilakukan untuk memastikan data memiliki format dan struktur yang konsisten sebelum digunakan dalam pemodelan. Tahapan ini meliputi penyelarasan identitas wilayah, penyesuaian periode pengamatan, serta pengolahan data iklim berdasarkan wilayah dan tahun. Data iklim bulanan diagregasikan menjadi total curah hujan serta rata-rata suhu dan kelembapan tahunan.

Selanjutnya dilakukan *feature engineering* untuk membentuk variabel masukan model. Fitur yang digunakan meliputi jumlah kasus tahun sebelumnya (kasus_t_1), total curah hujan tahun sebelumnya, total curah hujan tahunan, rata-rata suhu dan kelembapan, variasi suhu, jumlah bulan dengan curah hujan tinggi, jumlah fasilitas kesehatan, tren perubahan kasus, serta interaksi curah hujan dengan kelembapan dan suhu dengan kelembapan. Daftar fitur yang digunakan disajikan pada Tabel 2.

Tabel 2. Fitur yang Digunakan dalam Pemodelan

No.	Fitur	Keterangan
1	kasus_t_1	Kasus DBD tahun sebelumnya
2	total_curah_hujan_tahunan	Total curah hujan tahunan
3	rata_rata_suhu_tahunan	Rata-rata suhu tahunan
4	rata_rata_kelembapan_tahunan	Rata-rata kelembapan tahunan
5	std_suhu_tahunan	Variasi suhu tahunan
6	jumlah_bulan_hujan_tinggi	Jumlah bulan dengan hujan tinggi
7	jumlah_fasilitas_kesehatan	Jumlah fasilitas kesehatan
8	trend_perubahan_kasus	Tren perubahan kasus
9	interaksi_hujan_kelembapan	Interaksi curah hujan dan kelembapan
10	interaksi_suhu_kelembapan	Interaksi suhu dan kelembapan

Tabel 2 menunjukkan fitur yang digunakan untuk merepresentasikan kondisi historis kasus, faktor iklim, fasilitas kesehatan, serta hubungan antarvariabel lingkungan. Fitur-fitur tersebut digunakan sebagai variabel prediktor dalam pemodelan Random Forest, sedangkan hasil prediksi selanjutnya dievaluasi menggunakan pendekatan temporal dan *Spatial Cross-Validation*.



2.4 Pemodelan dan Validasi Prediktif

Pemodelan prediktif dalam penelitian ini bertujuan untuk memperkirakan jumlah kasus dan mengklasifikasikan hasil prediksi Demam Berdarah Dengue (DBD) di Kota Tangerang berdasarkan data historis kasus, kondisi iklim, fasilitas kesehatan, dan fitur hasil *feature engineering*. Model yang digunakan terdiri atas Random Forest Regression untuk memprediksi jumlah kasus dan Random Forest Classification untuk menghasilkan probabilitas serta klasifikasi hasil prediksi. Random Forest dipilih karena mampu menangkap hubungan non-linear dan interaksi antarvariabel tanpa memerlukan asumsi hubungan linear antara variabel prediktor dan hasil prediksi [16], [17].

Random Forest bekerja dengan membangun sejumlah *decision tree* berdasarkan sampel data dan fitur yang dipilih secara acak. Pada Random Forest Regression, prediksi akhir diperoleh melalui rata-rata hasil prediksi seluruh pohon. Mekanisme tersebut dinyatakan pada Persamaan (1).

$$\hat{y} = \frac{1}{B} \sum_{b=1}^B T_b(X) \quad (1)$$

dengan $\hat{y}$ merupakan hasil prediksi jumlah kasus DBD, B merupakan jumlah *decision tree*, $T_b(X)$ merupakan hasil prediksi pohon ke- b , dan X merupakan vektor fitur masukan.

Hasil prediksi digunakan untuk memperkirakan jumlah kasus DBD pada tahun 2026 dan 2027. Karena penelitian ini tidak menggunakan data jumlah penduduk maupun luas wilayah sebagai dasar normalisasi, hasil prediksi ditafsirkan sebagai estimasi jumlah kasus pada masing-masing kelurahan, bukan sebagai angka kejadian DBD yang telah disesuaikan terhadap ukuran populasi atau luas wilayah.

Untuk keperluan klasifikasi hasil prediksi dan visualisasi pada dashboard GIS, penelitian ini menggunakan nilai ambang 60 kasus sebagai *threshold* operasional. Prediksi jumlah kasus ≥ 60 kasus dikategorikan sebagai “Prediksi Tinggi”, sedangkan prediksi jumlah kasus < 60 kasus dikategorikan sebagai “Prediksi Rendah”. *Threshold* tersebut digunakan sebagai batas pengelompokan teknis untuk membedakan hasil prediksi antarwilayah dan bukan merupakan batas epidemiologis atau ukuran risiko yang telah dinormalisasi berdasarkan jumlah penduduk maupun luas wilayah.

Selanjutnya, hasil prediksi dan klasifikasinya diintegrasikan dengan data spasial kelurahan dan divisualisasikan melalui dashboard GIS. Visualisasi tersebut digunakan untuk menunjukkan distribusi estimasi jumlah kasus DBD secara spasial serta membantu mengidentifikasi wilayah dengan hasil prediksi yang lebih tinggi sebagai informasi pendukung dalam penentuan wilayah prioritas.

2.5 Optimasi dan Evaluasi Model

Untuk memperoleh konfigurasi model yang sesuai, dilakukan optimasi *hyperparameter* menggunakan Randomized Search yang dilanjutkan dengan Grid Search. Parameter yang dioptimasi meliputi *n_estimators*, *max_depth*, *min_samples_split*, *min_samples_leaf*, dan *max_features*. Pendekatan optimasi *hyperparameter* pada Random Forest digunakan untuk memperoleh konfigurasi model dengan kinerja validasi yang lebih baik [18].

Validasi dilakukan menggunakan *Spatial Cross-Validation* untuk menguji kemampuan generalisasi model pada wilayah geografis yang berbeda. Data dibagi berdasarkan blok spasial, kemudian setiap blok digunakan secara bergantian sebagai data pengujian, sedangkan blok lainnya digunakan sebagai data pelatihan. Pendekatan ini digunakan untuk mengurangi potensi bias evaluasi akibat autokorelasi spasial, terutama karena kelurahan yang berdekatan dapat memiliki karakteristik lingkungan dan pola kasus yang relatif serupa. Dengan demikian, kinerja model dapat dievaluasi pada wilayah yang berbeda dari data yang digunakan dalam proses pelatihan [19].

Random Forest Regression dievaluasi berdasarkan kemampuan model dalam menghasilkan estimasi jumlah kasus DBD, sedangkan Random Forest Classification dievaluasi menggunakan *Area Under the Curve* (AUC), *Precision*, *Recall*, dan *F1-Score*. Pada model klasifikasi, kelas yang digunakan terdiri atas “Prediksi Tinggi” dan “Prediksi Rendah” sesuai dengan *threshold* operasional yang telah ditetapkan pada tahap pemodelan. AUC digunakan untuk mengukur kemampuan model dalam membedakan kedua kelas, *Precision* digunakan untuk mengukur ketepatan prediksi kelas “Prediksi Tinggi”, *Recall* digunakan untuk mengukur kemampuan model dalam mengidentifikasi kelas tersebut, sedangkan *F1-Score* digunakan untuk mengukur keseimbangan antara *Precision* dan *Recall*.

Hasil evaluasi *Spatial Cross-Validation* digunakan untuk menilai kemampuan generalisasi model terhadap wilayah geografis yang berbeda. Model dengan konfigurasi terbaik selanjutnya digunakan untuk menghasilkan prediksi jumlah kasus dan klasifikasi hasil prediksi yang kemudian divisualisasikan melalui dashboard GIS.

2.6 Implementasi dan Evaluasi Sistem

Implementasi penelitian diwujudkan dalam bentuk dashboard berbasis web menggunakan Streamlit yang menghubungkan hasil pemodelan Random Forest dengan data spasial Kota Tangerang. Streamlit digunakan sebagai antarmuka untuk menyajikan hasil prediksi dan informasi risiko DBD, sedangkan Folium digunakan untuk menampilkan peta interaktif berbasis Geographic Information System (GIS). Melalui pemetaan tersebut, distribusi risiko dapat dilihat berdasarkan wilayah sehingga pola persebaran dan lokasi yang perlu mendapat perhatian dapat dikenali dengan lebih mudah [20].

Dashboard menampilkan distribusi hasil prediksi, kategori prediksi, estimasi jumlah kasus, dan *feature importance*. Hasil prediksi kemudian dikaitkan dengan batas administrasi kelurahan sehingga setiap informasi dapat

ditampilkan sesuai dengan lokasi geografisnya. Pendekatan ini memungkinkan hasil pemodelan yang sebelumnya berupa data numerik untuk disajikan dalam bentuk spasial yang lebih mudah dipahami.

Evaluasi sistem dilakukan dengan memeriksa fungsi utama dashboard, meliputi pemuatan data, visualisasi peta menggunakan Folium, penyajian hasil prediksi, kategori prediksi, dan informasi *feature importance*. Hasil yang ditampilkan juga diperiksa kesesuaiannya dengan keluaran model dan data spasial. Dengan demikian, dashboard yang dikembangkan dapat digunakan sebagai alat bantu untuk mengenali wilayah dengan hasil prediksi yang lebih tinggi serta mendukung pengambilan keputusan berbasis data dalam upaya pencegahan dan pengendalian DBD di Kota Tangerang.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penerapan Metode

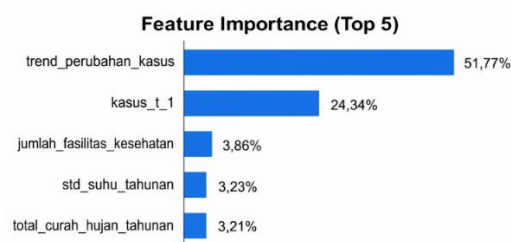
Penerapan metode Random Forest dilakukan untuk menghasilkan estimasi jumlah kasus DBD dan klasifikasi hasil prediksi di Kota Tangerang. Sebelum digunakan dalam proses prediksi, dilakukan optimasi *hyperparameter* untuk memperoleh konfigurasi model yang sesuai dengan karakteristik data penelitian. Hasil konfigurasi *hyperparameter* terbaik disajikan pada Tabel 3.

Tabel 3. Konfigurasi *Hyperparameter* Terbaik

<i>Hyperparameter</i>	Nilai Optimal
<i>n_estimators</i>	700
<i>max_depth</i>	10
<i>min_samples_split</i>	4
<i>min_samples_leaf</i>	1
<i>max_features</i>	<i>sqrt</i>
<i>class_weight</i>	<i>balanced</i>

Berdasarkan Tabel 3, model menggunakan 700 *decision tree* dengan kedalaman maksimum (*max_depth*) sebesar 10. Parameter *min_samples_split* sebesar 4 dan *min_samples_leaf* sebesar 1 digunakan untuk mengatur proses pembentukan dan pemisahan *node*. Parameter *max_features* menggunakan *sqrt*, sehingga sebagian fitur dipilih secara acak pada setiap proses pemisahan. Sementara itu, parameter *class_weight=balanced* digunakan untuk mengantisipasi ketidakseimbangan jumlah observasi antara kelas “Prediksi Tinggi” dan “Prediksi Rendah”. Pengaturan ini memberikan bobot yang lebih besar pada kelas dengan jumlah observasi lebih sedikit sehingga proses pembelajaran model tidak terlalu didominasi oleh kelas mayoritas.

Konfigurasi tersebut selanjutnya digunakan dalam proses pemodelan untuk menghasilkan prediksi jumlah kasus DBD. Hasil prediksi kemudian diintegrasikan dengan batas administrasi kelurahan Kota Tangerang untuk divisualisasikan secara spasial melalui dashboard GIS. Selain menghasilkan prediksi, dilakukan analisis *feature importance* untuk mengetahui kontribusi relatif setiap fitur terhadap pembentukan model. Lima fitur dengan kontribusi terbesar ditampilkan pada Gambar 2.



Gambar 2. Lima Fitur dengan *Feature Importance* Tertinggi pada Model Random Forest

Berdasarkan Gambar 2, fitur *trend_perubahan_kasus* memiliki kontribusi relatif terbesar sebesar 51,77%, diikuti oleh *kasus_t_1* sebesar 24,34%. Selanjutnya, *jumlah_fasilitas_kesehatan* memberikan kontribusi sebesar 3,86%, *std_suhu_tahunan* sebesar 3,23%, dan *total_curah_hujan_tahunan* sebesar 3,21%. Hasil tersebut menunjukkan bahwa fitur yang merepresentasikan informasi historis kasus memiliki kontribusi relatif lebih besar dibandingkan fitur iklim dan fasilitas kesehatan dalam pembentukan prediksi model.

Besarnya kontribusi fitur historis tersebut perlu diperhatikan dalam interpretasi model. *Feature importance* menunjukkan kontribusi relatif fitur dalam proses pembentukan keputusan Random Forest dan tidak dapat diartikan sebagai hubungan sebab-akibat. Oleh karena itu, pembentukan *trend_perubahan_kasus* dan *kasus_t_1* perlu dipastikan hanya menggunakan informasi dari periode sebelumnya agar tidak terjadi penggunaan informasi dari periode target (*look-ahead bias*).

Hasil prediksi selanjutnya digunakan untuk menghasilkan klasifikasi “Prediksi Tinggi” dan “Prediksi Rendah” berdasarkan *threshold* operasional yang telah ditetapkan pada tahap pemodelan. Hasil tersebut kemudian divisualisasikan dalam bentuk peta untuk menunjukkan distribusi prediksi DBD antarwilayah kelurahan di Kota Tangerang.

3.2 Implementasi dan Pengujian

Pengujian dilakukan untuk mengevaluasi kinerja Random Forest Classification dalam mengklasifikasikan hasil prediksi DBD. Evaluasi utama dilakukan berdasarkan pembagian data secara temporal, yaitu data pada tahun-tahun sebelum tahun pengujian digunakan sebagai data pelatihan, sedangkan tahun terakhir digunakan sebagai data pengujian. Selain itu, *Spatial Cross-Validation* digunakan sebagai evaluasi tambahan untuk mengetahui kemampuan model dalam melakukan generalisasi pada wilayah geografis yang berbeda. Kinerja model dievaluasi menggunakan *Area Under the Curve* (AUC), *Precision*, *Recall*, dan *F1-Score*. AUC digunakan untuk mengukur kemampuan model dalam membedakan kelas “Prediksi Tinggi” dan “Prediksi Rendah”, sedangkan *Precision* dan *Recall* digunakan untuk melihat ketepatan dan kemampuan model dalam mengenali kelas positif. *F1-Score* digunakan untuk menggambarkan keseimbangan antara *Precision* dan *Recall*. Pada *Spatial Cross-Validation*, data dibagi ke dalam beberapa blok spasial dan setiap blok digunakan secara bergantian sebagai data pengujian, sedangkan blok lainnya digunakan sebagai data pelatihan. Nilai setiap metrik kemudian dirata-ratakan dari *fold* yang dapat dievaluasi.

Tabel 4. Hasil Evaluasi Kinerja Random Forest

Metode Evaluasi	AUC	Precision	Recall	F1-Score
Evaluasi Utama Berbasis Waktu	0,973	0,846	0,936	0,889
Spatial Cross-Validation	0,956	0,873	0,920	0,892

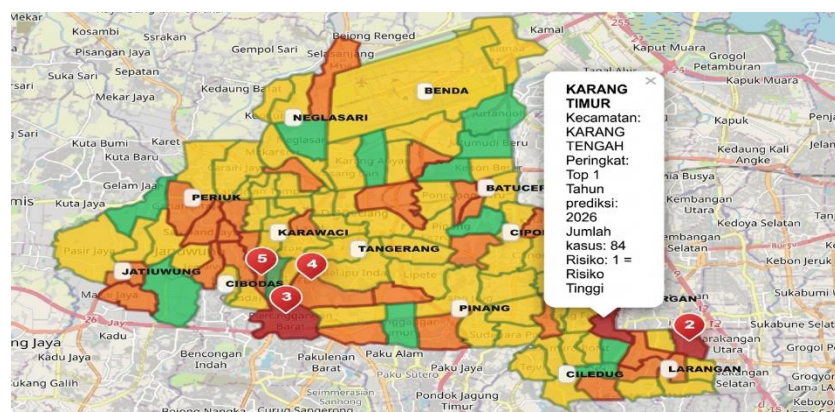
Berdasarkan Tabel 4, evaluasi utama berbasis waktu menghasilkan AUC sebesar 0,973, *Precision* sebesar 0,846, *Recall* sebesar 0,936, dan *F1-Score* sebesar 0,889. Hasil tersebut menunjukkan bahwa model mampu membedakan kelas “Prediksi Tinggi” dan “Prediksi Rendah” dengan kinerja yang tinggi pada data pengujian berbasis waktu.

Pada evaluasi menggunakan *Spatial Cross-Validation*, diperoleh AUC sebesar 0,956, *Precision* sebesar 0,873, *Recall* sebesar 0,920, dan *F1-Score* sebesar 0,892. Dibandingkan dengan evaluasi utama, nilai AUC dan *Recall* mengalami sedikit penurunan, sedangkan *Precision* dan *F1-Score* mengalami peningkatan. Perbedaan tersebut menunjukkan bahwa kinerja model relatif konsisten ketika pengujian mempertimbangkan pemisahan berdasarkan blok geografis. Nilai AUC yang tinggi pada kedua skenario perlu diinterpretasikan secara hati-hati. Salah satu hal yang perlu diperhatikan adalah kontribusi fitur historis, khususnya *trend_perubahan_kasus* dan *kasus_t_1*, yang memiliki kontribusi relatif besar pada *feature importance*. Oleh karena itu, hasil kinerja model tidak hanya dipertimbangkan berdasarkan nilai metrik, tetapi juga dikaitkan dengan proses pembentukan fitur dan pemisahan data temporal untuk memastikan bahwa informasi dari periode target tidak digunakan dalam proses pelatihan.

Secara keseluruhan, hasil pengujian menunjukkan bahwa Random Forest menghasilkan kinerja klasifikasi yang tinggi pada evaluasi berbasis waktu maupun *Spatial Cross-Validation*. Hasil *Spatial Cross-Validation* juga memberikan informasi tambahan mengenai kemampuan model ketika diuji pada blok wilayah yang berbeda. Dengan demikian, hasil evaluasi dapat digunakan sebagai dasar untuk melihat konsistensi kinerja model sebelum hasil prediksi divisualisasikan melalui dashboard GIS.

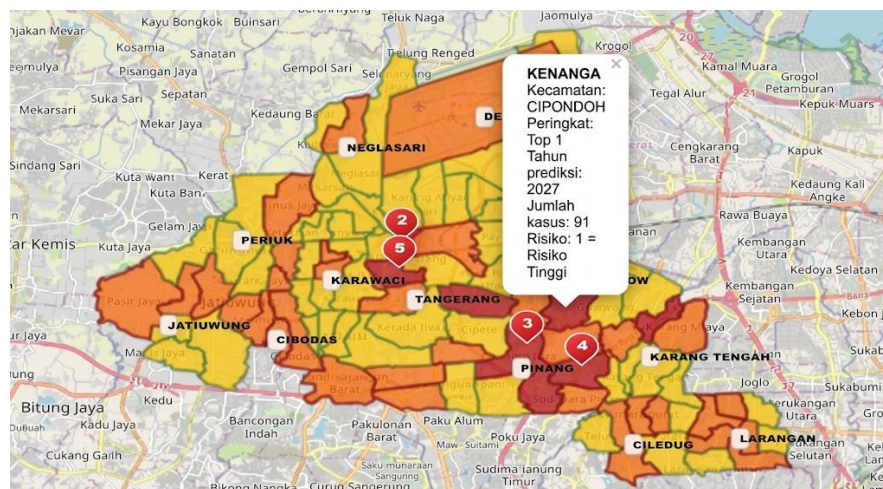
3.3 Hasil Prediksi dan Pemetaan Risiko DBD

Model Random Forest yang telah dioptimasi digunakan untuk menghasilkan prediksi jumlah kasus Demam Berdarah Dengue (DBD) di Kota Tangerang untuk tahun 2026 dan 2027. Hasil prediksi kemudian diintegrasikan dengan batas administrasi kelurahan untuk memvisualisasikan distribusi hasil prediksi secara spasial. Dalam penelitian ini, prediksi jumlah kasus ≥ 60 dikategorikan sebagai “Prediksi Tinggi”, sedangkan prediksi < 60 dikategorikan sebagai “Prediksi Rendah”. Batas tersebut merupakan *threshold* operasional yang digunakan untuk memudahkan interpretasi hasil prediksi pada dashboard GIS. Hasil pemetaan prediksi tahun 2026 ditunjukkan pada Gambar 3. Peta memperlihatkan adanya variasi nilai prediksi antarwilayah kelurahan di Kota Tangerang. Beberapa wilayah memiliki nilai prediksi yang lebih tinggi dan masuk dalam kategori “Prediksi Tinggi”. Nilai prediksi tertinggi pada tahun 2026 terdapat pada Kelurahan Karang Timur, Kecamatan Karang Tengah, dengan estimasi sebesar 84 kasus.



Gambar 3. Peta Prediksi Risiko DBD Kota Tangerang Tahun 2026

Pemetaan tahun 2027 menunjukkan adanya perubahan distribusi nilai prediksi dibandingkan tahun 2026. Nilai prediksi tertinggi pada tahun 2027 terdapat pada Kelurahan Kenanga, Kecamatan Cipondoh, dengan estimasi sebesar 91 kasus. Nilai tersebut termasuk dalam kategori “Prediksi Tinggi” berdasarkan *threshold* yang digunakan.



Gambar 4. Peta Prediksi Risiko DBD Kota Tangerang Tahun 2027

Perbedaan hasil prediksi pada kedua periode menunjukkan bahwa wilayah dengan nilai prediksi tertinggi dapat berubah dari satu tahun ke tahun berikutnya. Untuk memberikan gambaran lebih rinci mengenai wilayah dengan nilai prediksi tertinggi, sepuluh wilayah dengan prediksi kasus tertinggi pada tahun 2026 dan 2027 disajikan pada Tabel 5.

Tabel 5. Sepuluh Wilayah dengan Prediksi Kasus DBD Tertinggi Tahun 2026 dan 2027

Peringkat	Kecamatan 2026	Kelurahan 2026	Prediksi 2026	Kecamatan 2027	Kelurahan 2027	Prediksi 2027
1	Karang Tengah	Karang Timur	84	Cipondoh	Kenanga	91
2	Larangan	Kreo	83	Tangerang	Sukasih	88
3	Cibodas	Panunggangan Barat	80	Pinang	Kunciran Jaya	87
4	Tangerang	Cikokol	74	Pinang	Pinang	85
5	Karawaci	Karawaci Baru	73	Tangerang	Sukasari	84
6	Karawaci	Koang Jaya	70	Cipondoh	Poris Indah	82
7	Karawaci	Cimone	70	Karang Tengah	Parung Jaya	81
8	Neglasari	Selapajang Jaya	69	Pinang	Kunciran Indah	80
9	Periuk	Gebang Raya	69	Pinang	Panunggangan Utara	78
10	Larangan	Gaga	68	Ciledug	Sudimara Barat	77

Berdasarkan Tabel 5, seluruh wilayah yang termasuk dalam sepuluh prediksi tertinggi pada kedua periode memiliki nilai prediksi ≥ 60 kasus, sehingga termasuk dalam kategori “Prediksi Tinggi”. Nilai prediksi tertinggi meningkat dari 84 kasus pada tahun 2026 menjadi 91 kasus pada tahun 2027. Lokasi dengan nilai prediksi tertinggi juga mengalami perubahan, yaitu dari Kelurahan Karang Timur di Kecamatan Karang Tengah pada tahun 2026 menjadi Kelurahan Kenanga di Kecamatan Cipondoh pada tahun 2027.

Perubahan tersebut menunjukkan adanya perbedaan hasil prediksi antarwilayah dan antarperiode. Namun, perubahan nilai prediksi tidak secara langsung dapat diartikan sebagai peningkatan kasus aktual maupun sebagai akibat faktor sosial-ekonomi atau kebijakan tertentu, karena penelitian ini tidak memasukkan variabel sosial-ekonomi dan kebijakan mitigasi lokal sebagai variabel prediktor. Hasil pemetaan lebih tepat digunakan sebagai informasi pendukung untuk mengidentifikasi wilayah yang memiliki nilai prediksi relatif tinggi.

Dengan demikian, integrasi Random Forest dengan GIS memberikan gambaran spasial mengenai distribusi hasil prediksi DBD pada tingkat kelurahan. Informasi tersebut dapat dimanfaatkan sebagai bahan pendukung dalam menentukan wilayah yang perlu mendapatkan perhatian lebih dalam kegiatan pemantauan dan pencegahan DBD di Kota Tangerang.

3.4 Pembahasan

Hasil penelitian menunjukkan bahwa model Random Forest mampu menghasilkan prediksi risiko DBD dengan kinerja klasifikasi yang baik. Berdasarkan evaluasi utama berbasis waktu, model memperoleh nilai AUC sebesar 0,973, F1-Score



sebesar 0,889, Precision sebesar 0,846, dan Recall sebesar 0,936. Hasil tersebut menunjukkan bahwa model memiliki kemampuan yang baik dalam membedakan kelas “Prediksi Tinggi” dan “Prediksi Rendah”. Nilai Recall yang relatif tinggi juga menunjukkan bahwa sebagian besar wilayah dengan prediksi tinggi dapat dikenali oleh model.

Hasil Spatial Cross-Validation menghasilkan AUC sebesar 0,956, F1-Score sebesar 0,892, Precision sebesar 0,873, dan Recall sebesar 0,920. Perbedaan yang relatif kecil dibandingkan evaluasi berbasis waktu menunjukkan bahwa performa model tetap konsisten ketika pengujian dilakukan pada wilayah geografis yang berbeda. Hal ini menunjukkan bahwa penerapan Spatial Cross-Validation memberikan gambaran tambahan mengenai kemampuan model dalam menghadapi variasi karakteristik antarwilayah.

Hasil prediksi juga menunjukkan adanya perubahan wilayah dengan nilai prediksi tertinggi antara tahun 2026 dan 2027. Pada tahun 2026, nilai prediksi tertinggi terdapat di Kelurahan Karang Timur dengan estimasi 84 kasus, sedangkan pada tahun 2027 terdapat di Kelurahan Kenanga dengan estimasi 91 kasus. Perbedaan tersebut menunjukkan bahwa wilayah dengan nilai prediksi tertinggi dapat berubah antarperiode sehingga analisis risiko DBD perlu mempertimbangkan dimensi spasial dan temporal secara bersamaan.

Karang Timur dan Kenanga merupakan kelurahan yang berada pada kawasan perkotaan Kota Tangerang dengan karakteristik permukiman yang relatif padat dan aktivitas masyarakat yang tinggi. Kondisi tersebut berpotensi meningkatkan intensitas interaksi manusia dengan lingkungan yang mendukung keberadaan vektor DBD, terutama apabila dipengaruhi oleh faktor iklim seperti curah hujan, suhu, dan kelembapan yang digunakan dalam model. Selain itu, perubahan kondisi lingkungan dari waktu ke waktu dapat memengaruhi dinamika prediksi pada kedua wilayah tersebut. Namun demikian, penelitian ini tidak menggunakan variabel sosial-ekonomi maupun data kebijakan mitigasi lokal sebagai variabel prediktor, sehingga perubahan nilai prediksi pada Karang Timur dan Kenanga tidak dapat disimpulkan sebagai akibat langsung dari faktor-faktor tersebut. Interpretasi ini lebih tepat dipandang sebagai indikasi wilayah yang memerlukan pemantauan lebih lanjut melalui data kepadatan penduduk, kondisi permukiman, serta program pengendalian DBD yang tersedia di tingkat kelurahan.

Integrasi hasil prediksi dengan GIS memberikan informasi yang lebih mudah diinterpretasikan karena nilai prediksi dapat dikaitkan langsung dengan lokasi geografis. Peta yang dihasilkan membantu mengidentifikasi wilayah dengan nilai prediksi yang relatif tinggi sehingga dapat digunakan sebagai informasi pendukung dalam penentuan prioritas pemantauan dan upaya pencegahan DBD di Kota Tangerang. Dengan demikian, kombinasi Random Forest, Spatial Cross-Validation, dan GIS dapat digunakan sebagai pendekatan berbasis data untuk mendukung analisis serta pemetaan prediksi DBD pada tingkat kelurahan.

Hasil penelitian ini sejalan dengan Yavari Nejad dan Varathan (2021) yang menunjukkan bahwa faktor iklim dapat digunakan sebagai prediktor dalam pemodelan DBD [4]. Kakarla et al. (2023) juga menunjukkan bahwa variabel cuaca seperti curah hujan, suhu, dan kelembapan dapat digunakan dalam prediksi dengue berbasis machine learning [5]. Selain itu, Ong et al. (2023) menunjukkan bahwa metode machine learning berbasis ensemble dapat digunakan untuk memprediksi transmisi dengue [7]. Temuan tersebut mendukung hasil penelitian ini yang memanfaatkan faktor iklim dan data historis dalam model Random Forest.

Dari aspek spasial, penelitian ini sejalan dengan Nayak et al. (2025) yang menunjukkan bahwa GIS dapat dimanfaatkan untuk mendukung pemetaan dan surveilans DBD [8]. Perbedaannya, penelitian ini mengintegrasikan Random Forest, Spatial Cross-Validation, dan GIS pada tingkat kelurahan. Dengan demikian, penelitian ini tidak hanya menghasilkan prediksi, tetapi juga memberikan informasi spasial mengenai wilayah dengan prediksi relatif tinggi yang dapat digunakan sebagai pendukung penentuan prioritas pemantauan DBD di Kota Tangerang.

4. KESIMPULAN

Penelitian ini menghasilkan model prediksi risiko Demam Berdarah Dengue (DBD) di Kota Tangerang dengan mengintegrasikan data historis kasus DBD, iklim, fasilitas kesehatan, dan informasi spasial menggunakan Random Forest serta *Spatial Cross-Validation*. Model menghasilkan kinerja yang baik pada evaluasi berbasis waktu maupun spasial, dengan hasil prediksi tertinggi sebesar 84 kasus di Kelurahan Karang Timur pada tahun 2026 dan 91 kasus di Kelurahan Kenanga pada tahun 2027. Perubahan lokasi prediksi tertinggi tersebut menunjukkan bahwa wilayah prioritas dapat berbeda antarperiode dan tidak dapat langsung diartikan sebagai akibat perubahan urbanisasi maupun fluktuasi iklim mikro karena penelitian ini belum memasukkan variabel tersebut secara khusus. Oleh karena itu, hasil prediksi lebih tepat digunakan sebagai indikator awal untuk menentukan wilayah yang perlu dipantau lebih lanjut. Secara praktis, dashboard GIS yang dikembangkan dapat menjadi informasi pendukung bagi Dinas Kesehatan Kota Tangerang dalam menentukan prioritas surveilans, pemantauan lingkungan, dan perencanaan tindakan preventif pengendalian DBD pada wilayah yang memiliki prediksi kasus tinggi. Penelitian selanjutnya disarankan memperpanjang periode pengamatan, menambahkan variabel kepadatan penduduk, kondisi lingkungan, sosial-ekonomi, serta informasi program pengendalian DBD di tingkat kelurahan agar faktor yang berkaitan dengan perubahan risiko dapat dianalisis secara lebih mendalam dan mendukung perumusan kebijakan kesehatan yang lebih tepat sasaran.

REFERENCES

- [1] W. H. Organization, “Dengue and Severe Dengue,” 2025. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>



- [2] K. K. R. Indonesia, "Profil Kesehatan Indonesia 2024," 2025. [Online]. Available: <https://kemkes.go.id/id/profil-kesehatan-indonesia-2024>
- [3] T. Nakase, M. Giovanetti, U. Obolski, and J. Lourenço, "Population at Risk of Dengue Virus Transmission Has Increased Due to Coupled Climate Factors and Population Growth," *Commun. Earth Environ.*, vol. 5, no. 1, p. 475, 2024, doi: 10.1038/s43247-024-01639-6.
- [4] F. Yavari Nejad and K. D. Varathan, "Identification of Significant Climatic Risk Factors and Machine Learning Models in Dengue Outbreak Prediction," *BMC Med. Inform. Decis. Mak.*, vol. 21, p. 141, 2021, doi: 10.1186/s12911-021-01493-y.
- [5] S. G. Kakarla, P. K. Kondeti, and H. Vavilala, "Weather Integrated Multiple Machine Learning Models for Prediction of Dengue Prevalence in India," *Int. J. Biometeorol.*, vol. 67, 2023, doi: 10.1007/s00484-022-02405-z.
- [6] H. Akbar, "The Effect of Climate Variability on Dengue Disease in Endemic Areas: A Study of Bandung City, Indonesia," *J. Public Heal. Dev.*, vol. 23, no. 3, pp. 276–291, 2025, doi: 10.55131/jphd/2025/230321.
- [7] S. Q. Ong, N. Ahmad, and F. Rahman, "Predicting Dengue Transmission Rates by Comparing Different Machine Learning Models with Vector Indices and Meteorological Data," *Sci. Rep.*, vol. 13, 2023, doi: 10.1038/s41598-023-46342-2.
- [8] P. P. Nayak, B. Pai, and S. Govindan, "Leveraging Geographic Information System for Dengue Surveillance: A Scoping Review," *Trop. Med. Health*, vol. 53, p. 102, 2025, doi: 10.1186/s41182-025-00783-9.
- [9] B. Haryanto, F. Dwirahmadi, T. Nurlambang, and A. Asyary, "Spatial Analysis on Dengue Fever Vulnerability in the Provinces of South Sulawesi and East Nusa Tenggara in Indonesia," *Ann. Glob. Heal.*, vol. 91, no. 1, pp. 86–98, 2025, doi: 10.5334/aogh.4915.
- [10] E. Sylvestre, C. Joachim, and E. Cecilia-Joseph, "Data-Driven Methods for Dengue Prediction and Surveillance Using Real-World and Big Data: A Systematic Review," *PLoS Negl. Trop. Dis.*, vol. 16, no. 1, p. e0010056, 2022, doi: 10.1371/journal.pntd.0010056.
- [11] W. Anggraeni, E. M. Yuniarno, and R. F. Rachmadi, "A Hybrid EMD-GRNN-PSO in Intermittent Time-Series Data for Dengue Fever Forecasting," *Expert Syst. Appl.*, vol. 237, p. 121438, 2024, doi: 10.1016/j.eswa.2023.121438.
- [12] X. Chen and P. Moraga, "Assessing Dengue Forecasting Methods: A Comparative Study of Statistical Models and Machine Learning Techniques in Rio de Janeiro, Brazil," *Trop. Med. Health*, vol. 53, p. 23, 2025, doi: 10.1186/s41182-025-00723-7.
- [13] T. B. T. Satoto, N. A. Pascawati, R. Frutos, Erizal, T. A. Garjito, and M. F. Salim, "AI-based prediction of dengue incidence using climatic, environmental, and socio-demographic factors: an ensemble random forest approach with agile system development," *BMC Infect. Dis.*, vol. 26, no. 1, p. 1199, 2026, doi: 10.1186/s12879-026-13270-1.
- [14] Y. Wang, M. Khodadadzadeh, and R. Zurita-Milla, "Spatial+: A New Cross-Validation Method to Evaluate Geospatial Machine Learning Models," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 121, p. 103364, 2023, doi: 10.1016/j.jag.2023.103364.
- [15] J. Linnenbrink, C. Milà, M. Ludwig, and H. Meyer, "kNNDM CV: k-fold Nearest-Neighbour Distance Matching Cross-Validation for Map Accuracy Estimation," *Geosci. Model Dev.*, vol. 17, no. 15, pp. 5897–5912, 2024, doi: 10.5194/gmd-17-5897-2024.
- [16] H. Hayatunnufus, S. Surbakti, and T. H. Febriana, "Prediction of Dengue Fever in Coastal Areas of North Sumatera (Kuala Namu and Belawan) With Random Forest and Support Vector Machine (SVM) Methods," *Data Sci. J. Comput. Appl. Informatics*, vol. 7, no. 2, pp. 103–110, 2023, doi: 10.32734/jocai.v7.i2-14355.
- [17] A. Kukkar, Y. Kumar, J. K. Sandhu, M. Kaur, T. S. Walia, and M. Amoon, "DengueFog: A Fog Computing-Enabled Weighted Random Forest-Based Smart Health Monitoring System for Automatic Dengue Prediction," *Diagnostics*, vol. 14, no. 6, p. 624, 2024, doi: 10.3390/diagnostics14060624.
- [18] R. Muzayanah, D. A. A. Pertiwi, M. Ali, and M. A. Muslim, "Comparison of GridSearchCV and Bayesian Hyperparameter Optimization in Random Forest Algorithm for Diabetes Prediction," *J. Soft Comput. Explor.*, vol. 5, no. 1, pp. 86–91, 2024, doi: 10.52465/joscex.v5i1.308.
- [19] J. Deppner and M. Cajias, "Accounting for Spatial Autocorrelation in Algorithm-Driven Hedonic Models: A Spatial Cross-Validation Approach," *J. Real Estate Financ. Econ.*, vol. 68, pp. 235–273, 2024, doi: 10.1007/s11146-022-09915-y.
- [20] N. Yuniari and I. Subrata, "Analisis Spasial Kasus Demam Berdarah Dengue dengan Sistem Informasi Geografis (SIG) di Kota Denpasar Tahun 2020-2022," *Arch. Community Heal.*, vol. 11, no. 2, p. 450, 2024, doi: 10.24843/ACH.2024.v11.i02.p13.