



ISSN : 2339 - 1871

BETRIK BESEMAH TEKNOLOGI INFORMASI & KOMPUTER

Editor Office : Pusat Penelitian & Pengabdian Pada Masyarakat
(PPPM) ITPA

Phone : 0857-9716-9578

email : betriktpa@itpa.ac.id

Prediksi Jumlah Titik Ruang Terbuka Hijau (RTH) Menggunakan *Regresi Linier Dan Model Random Forest*

M.Azzuhri Dinata¹, Helda Yenni², Wirta Agustin³, Aguston⁴

Program Studi Teknik Informatika, Universitas Sains Dan Teknologi Indonesia^{1,2,3,4}

Jl. Pendidikan No.17, Sidomulyo Barat, Tampan, Kota Pekanbaru, Riau 28293, Indonesia

Sur-el : dinataazzuhri@gmail.com¹, helayenni@sar.ac.id²

, wirtaagustin@usti.ac.id³, agustin@sar.ac.id⁴

Penulis Korespondensi: M.Azzuhri Dinata, dinataazzuhri@gmail.com

Abstrak: Pembangunan dan pelestarian Ruang Terbuka Hijau (RTH) merupakan bagian penting dalam menjaga keseimbangan lingkungan, khususnya di wilayah Ekoregion Sumatera. Penelitian ini bertujuan untuk memprediksi jumlah titik RTH dengan menggunakan pendekatan regresi linier dan algoritma *Random Forest*. Data yang digunakan mencakup variabel-variabel seperti luas wilayah dan luas hutan dari beberapa provinsi di Sumatera. Evaluasi performa model dilakukan menggunakan metrik MAE, RMSE, dan koefisien determinasi (R^2). Hasil analisis menunjukkan bahwa model *Random Forest* memiliki performa yang lebih unggul dibandingkan regresi linier, dengan nilai MAE sebesar 5.52, RMSE sebesar 5.88, dan R^2 sebesar 0.74. Sementara itu, regresi linier hanya mampu mencapai R^2 sebesar 0.45. Temuan ini menunjukkan bahwa *Random Forest* lebih efektif dalam menangkap pola data *non-linier* dan lebih akurat dalam memprediksi jumlah titik RTH. Penelitian ini memberikan kontribusi dalam pemanfaatan teknologi data *science* untuk mendukung perencanaan lingkungan yang berkelanjutan, serta menjadi dasar dalam pengambilan kebijakan tata ruang berbasis data.

Kata kunci : Ruang Terbuka Hijau, Prediksi, *Random Forest*, *Regresi Linier*, Ekoregion Sumatera, Data Science, Evaluasi Model,

Abstract : The development and preservation of Green Open Space (GOS) is an important part of maintaining environmental balance, especially in the Sumatra Ecoregion. This study aims to predict the number of GOS points using a linear regression approach and the *Random Forest* algorithm. The data used include variables such as area and forest area from several provinces in Sumatra. Model performance evaluation was carried out using MAE, RMSE, and coefficient of determination (R^2) metrics. The analysis results show that the *Random Forest* model has superior performance compared to linear regression, with an MAE value of 5.52, RMSE of 5.88, and R^2 of 0.74. Meanwhile, linear regression was only able to achieve an R^2 of 0.45. These findings indicate that *Random Forest* is more effective in capturing non-linear data patterns and more accurate in predicting the number of GOS points. This study contributes to the use of data science technology to support sustainable environmental planning, as well as becoming a basis for data-based spatial planning policy making.

Keywords: Green Open Space, Prediction, *Random Forest*, Linear Regression, Sumatra Ecoregion, Data Science, Model Evaluation.

Received: 29-07-2025 | Accepted: 08-08-2025 | Published Online: 30-08-2025

All author: M. Azzuhri Dinata, Helda Yenni, Wirta Agustin, Aguston

1. PENDAHULUAN

Peningkatan populasi dan perkembangan urbanisasi yang pesat telah membawa dampak signifikan terhadap tata ruang kota, khususnya dalam hal ketersediaan ruang terbuka hijau (RTH). Ruang terbuka hijau (RTH) merupakan salah satu elemen vital dalam mendukung keseimbangan ekologis, estetika, dan kualitas hidup di kawasan perkotaan. Namun, banyak kota di dunia, termasuk di Indonesia, menghadapi tantangan besar dalam mempertahankan proporsi RTH yang ideal. Berbagai studi menunjukkan bahwa keberadaan RTH yang memadai berperan sebagai paru-paru kota serta sebagai elemen mitigasi bencana lingkungan, seperti banjir dan pencemaran udara[1].

Ekoregion Sumatera suatu kawasan ekologis yang didefinisikan oleh Kementerian Lingkungan Hidup dan Kehutanan (KLHK) sebagai wilayah dengan karakteristik ekologi yang serupa, mencakup berbagai ekosistem alami di Pulau Sumatera. Kawasan ini mencakup hutan hujan tropis, lahan basah, dan ekosistem daratan yang memiliki peran penting dalam menjaga keseimbangan ekologis [2].

Saat ini, kendala utama dalam pengelolaan RTH adalah minimnya metode yang akurat untuk memprediksi perubahan jumlah titik RTH secara efektif. Oleh karena itu, diperlukan pendekatan berbasis teknologi dan analisis data guna memantau serta memprediksi perubahan tersebut. Penelitian ini menawarkan solusi dengan menerapkan metode *Regresi Linier* dan model *Random Forest* untuk memprediksi jumlah titik RTH di masa depan. Kedua metode ini dipilih karena telah terbukti efektif dalam berbagai penelitian terkait prediksi data spasial dan lingkungan[3].

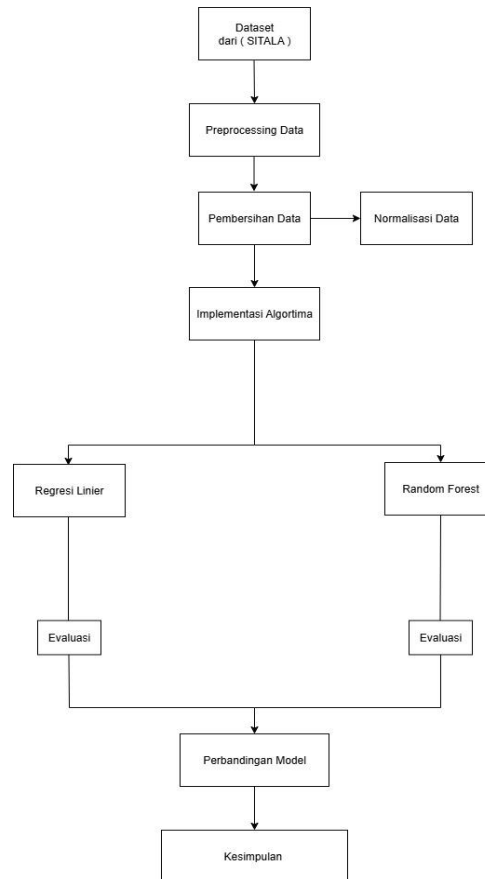
Perencanaan ruang terbuka hijau (RTH) didasarkan pada pertimbangan dapat terwujudnya keseimbangan, keserasian, dan keselamatan bangunan gedung dengan lingkungan di sekitarnya, serta mempertimbangkan terciptanya ruang luar bangunan gedung dan ruang terbuka hijau yang seimbang, serasi, dan selaras dengan lingkungan di sekitarnya. Rencana tata ruang menjadi landasan dalam mengantisipasi pesatnya perkembangan ruang-ruang terbangun, yang harus diikuti dengan kebijakan penyediaan ruang terbuka.[4]

Python adalah bahasa pemrograman *open-source* tingkat tinggi yang mudah dipahami dan dapat digunakan untuk berbagai bidang, termasuk keamanan siber. Python didesain dengan sintaks yang sederhana dan jelas, sehingga mudah dipelajari dan digunakan. Python dikembangkan oleh Guido van Rossum pada tahun 1991. Python memiliki banyak *library* dan *framework* yang memudahkan pengembangan berbagai jenis aplikasi, termasuk web aplikasi. Kelebihan utama Python adalah kemampuannya untuk memprioritaskan keterbacaan kode, sehingga pengembang dapat dengan mudah memahami dan mengelola kode program agar lebih rapih dan efektif. Python juga mendukung beragam paradigma pemrograman seperti pemrograman berorientasi objek, pemrograman fungsional, dan pemrograman prosedural.[5]

Ruang terbuka hijau kawasan perkotaan yang inklusif menjadi salah satu poin dalam tujuan Sustainable Development Goals yang menyebutkan adanya keinginan untuk membangun kota yang inklusif, aman, tangguh, dan berkelanjutan. Ruang terbuka hijau yang inklusif artinya penyediaannya harus berfokus pada aksesibilitas dan kualitasnya.[6]

2. METODOLOGI PENELITIAN

Desain penelitian adalah rancangan umum yang menggambarkan langkah-langkah sistematis yang dilakukan mulai dari perumusan masalah, Datase, penerapan metode, hingga analisis hasil untuk mencapai tujuan penelitian. Adapun penjelasan tahapan penelitian ini dijelaskan pada Gambar 1.



Gambar 1. Metodologi Penelitian

2.1 Dataset

Dataset adalah kumpulan data yang terstruktur dan disusun dalam format tertentu untuk keperluan analisis, pelatihan model, atau pengambilan keputusan. *Dataset* biasanya terdiri dari baris (*record*) dan kolom (atribut/fitur/variabel). Adapun tahapan dataset dalam penelitian ini adalah sebagai berikut :

2.2.1 Sumber Data

Sumber data yang digunakan dalam penelitian ini berasal dari *website* SITALA (<https://ppkl.menlhk.go.id/iklh/dashboardData>). Platform ini menyediakan akses publik terhadap data spasial dan lingkungan yang berkaitan dengan Ruang Terbuka Hijau (RTH), yang selanjutnya dianalisis menggunakan metode *regresi linier* dan model *Random Forest* untuk menghasilkan prediksi jumlah titik RTH di wilayah Ekoregion Sumatera.

2.2.2 Metode Pengumpulan Dataset

Metode pengumpulan dataset dalam penelitian ini bertujuan untuk memperoleh data yang relevan dalam rangka memprediksi jumlah titik Ruang Terbuka Hijau (RTH) di wilayah Ekoregion

Sumatera. Data yang dikumpulkan mencakup berbagai variabel lingkungan yang berpengaruh terhadap dinamika RTH. Data yang dipakai atau digunakan dalam penelitian ini diperoleh dari SITALA (Sistem Indeks Kualitas Lingkungan Hidup) yang dimana biasa dilihat di website PUSDAL LH. Dimana data ini berisi informasi tentang Jumlah titik RTH perkab/kota di Sumatera per tahun mulai dari 2022,2023, sampai dengan 2024, dengan berdasarkan data yang berjumlah 454 baris dengan 5 Variabel.

2.2 Preprocessing Data

Preprocessing data merupakan tahapan penting dalam penelitian ini untuk memastikan data yang digunakan dalam pemodelan prediksi jumlah titik Ruang Terbuka Hijau (RTH) di Ekoregion Sumatera berada dalam kondisi bersih, konsisten, dan sesuai untuk dianalisis oleh model prediksi. Adapun tahapan dalam preprocessing data adalah sebagai berikut:

2.2.1 Pembersihan Data

Pada tahap ini dilakukan proses identifikasi dan penanganan terhadap data yang tidak valid, tidak lengkap, mengandung anomali, atau tidak konsisten. Tujuan dari proses ini adalah untuk memastikan data yang digunakan dalam pemodelan prediksi memiliki kualitas yang baik dan bebas dari kesalahan yang dapat mempengaruhi hasil analisis.

2.2.2 Normalisasi dan Transformasi Data

Normalisasi data bertujuan untuk menstandarkan skala variabel numerik agar tidak ada variabel yang mendominasi proses analisis akibat perbedaan skala. Hal ini sangat penting terutama dalam penerapan algoritma regresi linier dan random forest. Transformasi data dilakukan untuk mengubah data mentah ke dalam format, struktur, atau skala yang lebih sesuai dan optimal untuk proses analisis dan pemodelan.

2.3 Prediksi

Prediksi/*Forecasting* adalah salah satu metode untuk memprediksi meramalkan sesuatu yang tidak akan pernah terjadi. Tujuannya untuk mengurangi kesalahan yang mungkin terjadi, Prediksi adalah kegiatan yang struktural yang menggunakan data dari masa lalu dan saat ini untuk membuat prediksi tentang peristiwa yang mungkin terjadi di masa depan. Berdasarkan pengertian tersebut, prediksi dapat disimpulkan sebagai aktivitas peramalan atau perkiraan suatu keadaan di masa depan dengan 3 menggunakan data masa lalu guna mengurangi potensi kesalahan[7].

2.4 Ruang Terbuka Hijau (RTH)

Ruang terbuka hijau (RTH) adalah area yang dipenuhi vegetasi, baik secara alami maupun melalui penanaman, di dalam suatu teritorial yang lebih luas. RTH bisa berupa hamparan lahan atau jalur panjang dan biasanya bebas dari bangunan. RTH berfungsi sebagai ruang yang mendukung fungsi ekologis, seperti meningkatkan kualitas udara, menurunkan suhu, mengurangi polusi, dan meredam kebisingan. Selain itu, RTH juga berperan dalam penggunaan lahan dan pelestarian sumber daya alam, termasuk air dan tanah serta keanekaragaman hayati. Perencanaan RTH di lingkungan perkotaan meliputi teritorial dalam sebuah

wilayah atau daerah yang lebih luas, pada jalur yang memanjang di daerah terbuka tanpa bangunan sesuai peraturan Menteri Dalam Negeri No.14 Tahun 1988[8].

2.5 Pemrograman *Python*

Python merupakan bahasa pemrograman tingkat tinggi bersifat *open-source* yang dirancang agar mudah dipahami dan digunakan dalam berbagai bidang, termasuk dalam keamanan siber. Bahasa ini memiliki sintaks yang ringkas dan jelas, sehingga membuatnya mudah untuk dipelajari dan diimplementasikan. *Python* pertama kali dikembangkan oleh Guido van Rossum pada tahun 1991. Bahasa ini dilengkapi dengan berbagai pustaka (*library*) dan kerangka kerja (*framework*) yang mendukung proses pengembangan berbagai jenis program, termasuk aplikasi *web*. *Python* bersifat interpretatif, yang berarti kode program dapat langsung dijalankan melalui interpreter tanpa perlu proses kompilasi sebelumnya. Kelebihan utama *Python* adalah kemampuannya untuk memprioritaskan keterbacaan kode, sehingga pengembang dapat dengan mudah memahami dan mengelola kode program agar lebih rapih dan efektif. *Python* juga mendukung beragam paradigma pemrograman seperti pemrograman berorientasi objek, pemrograman fungsional, dan pemrograman prosedural[5].

2.6 Algoritma *Regresi Linier*

Regresi linier sederhana merupakan salah satu metode yang digunakan untuk memodelkan hubungan antara satu variabel bebas (*independen*) dengan satu variabel terikat (*dependen*). Dalam proses regresi, variabel *independen* digunakan untuk memprediksi atau menjelaskan variabel *dependen*. Hubungan yang terbentuk dalam regresi linier bersifat linier, artinya jika terjadi perubahan pada variabel *X*, maka perubahan tersebut akan memengaruhi variabel *Y* secara konstan atau tetap. Sebaliknya, jika hubungan yang terjadi bersifat non-linier, maka perubahan pada variabel *X* tidak akan diikuti oleh perubahan yang sebanding pada variabel *Y*. [9].

Persamaan *regresi linier* adalah :

$$y = \alpha + bX \quad (1)$$

Dimana :

- a. y = variabel *dependen*
- b. X = variabel *independen*
- c. α = *intercept*
- d. b = koefisien regresi

2.6.1 Prinsip Kerja

Tahapan Implementasi Algoritma *Regresi Linier* Implementasi algoritma *Regresi Linier* dalam penelitian ini dilakukan melalui beberapa langkah berikut:

- a. Identifikasi Variabel — Menentukan variabel *dependen* (Jumlah titik RTH) dan variabel *independen* (Luas wilayah).
- b. Pembentukan Model — Menyusun model *regresi linier* sederhana berdasarkan rumus yang telah ditentukan.

- c. Estimasi Koefisien — Menghitung nilai koefisien regresi melalui bahasa pemrograman *Python* pada platform Google Colab.

Regresi linier dipilih karena kesederhanaannya dalam menginterpretasikan hubungan antar variabel dan kemampuannya dalam memprediksi tren jangka panjang.

2.7 Algoritma *Random Forest*

Algoritma *Random Forest* dikembangkan oleh Leo Breiman. Metode ini terdiri dari sekumpulan pohon keputusan untuk regresi maupun klasifikasi, yang dibentuk tanpa proses pemangkasan, dengan pengambilan sampel data secara acak. Hasil prediksi diperoleh dengan menggabungkan output dari seluruh pohon keputusan yang ada. *Random Forest* dikenal memiliki beberapa keunggulan, di antaranya mampu mendeteksi kesalahan secara lebih baik, memberikan hasil klasifikasi yang akurat, efektif meskipun jumlah data latih sedikit, serta dapat mengatasi permasalahan data yang hilang melalui pendekatan estimasi yang efisien. [10]

Algoritma utama dalam *Random Forest* meliputi:

1. Pengambilan Sampel Acak: Data latih dipecah menjadi beberapa subset secara acak.
2. Pembangunan Pohon Keputusan: Setiap subset digunakan untuk membangun pohon keputusan secara independen.
3. Penggabungan Hasil: Prediksi akhir dihasilkan dari agregasi output seluruh pohon keputusan melalui voting atau averaging.

Tahapan Metode *Random Forest*:

1. Bangun Model *Random Forest*
Tentukan parameter seperti jumlah pohon (*n_estimators*), kedalaman pohon (*max_depth*), dll.
2. Latih Model (*Training*)
Model akan membuat banyak pohon dari data yang berbeda-beda (secara acak) dan masing-masing pohon belajar membuat prediksi.
3. Uji Model (*Testing*)
Hasil prediksi dari semua pohon akan dirata-ratakan untuk menentukan hasil akhir.
4. Evaluasi Model
Gunakan metrik evaluasi yang sama seperti regresi linier:
 - a. R^2 Score
 - b. MAE
 - c. RMSE
 - d. Visualisasi Hasil (Opsional)

Bisa dalam bentuk:

- a. Grafik *feature importance* (fitur mana yang paling berpengaruh)
- b. Perbandingan hasil prediksi dan nilai sebenarnya.

Metode ini sangat cocok digunakan dalam analisis perubahan lingkungan yang kompleks, seperti prediksi jumlah titik RTH di ekoregion Sumatera.

2.8 Evaluasi Model

Evaluasi model dilakukan untuk menilai performa dua algoritma yang digunakan, yaitu *regresi Linier* dan *random forest*, dalam memprediksi jumlah titik ruang terbuka hijau (RTH). Evaluasi dilakukan berdasarkan tiga indikator utama:

1. MAE (*Mean Absolute Error*)

Mengukur rata-rata selisih absolut antara nilai asli dan nilai prediksi.

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

- a. y_i : nilai aktual
- b. \hat{y}_i : nilai prediksi
- c. n : Jumlah data

Semakin kecil MAE, semakin baik modelnya.

Tabel 1. Model Metrik MAE

Rasio MAE terhadap rata-rata Y	Kategori
< 10%	Sangat Rendah (akurat)
10–20%	Rendah
20–30%	Sedang
> 30%	Tinggi (besar)

2. RMSE (*Root Mean Squared Error*)

Mengukur besar akar dari rata-rata kuadrat selisih antara nilai asli dan prediksi.

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

RMSE lebih sensitif terhadap *outlier* (nilai yang sangat jauh dari rata-rata)

Tabel 2. Model Metrik RMSE

Rasio RMSE terhadap rata-rata Y	Kategori
< 10%	Sangat Rendah
10–20%	Rendah
20–30%	Sedang
> 30%	Tinggi

3. R^2 (Kofisien determinasi)

Mengukur seberapa baik model menjelaskan variasi dari data aktual

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

\bar{y} : rata-rata nilai aktual

R^2 bernilai antara 0-1. Semakin mendekati 1, semakin baik model menjelaskan data.

Tabel 3. Model Metrik R^2

Nilai R^2	Kategori	Interpretasi Singkat
> 0.90	Sangat Tinggi	Hampir semua variasi dijelaskan
0.75 – 0.90	Tinggi	Mayoritas variasi dijelaskan
0.50 – 0.75	Sedang	Sekitar setengah variasi dijelaskan
0.25 – 0.50	Rendah	Sebagian kecil variasi dijelaskan
< 0.25	Sangat Rendah	Hampir tidak dijelaskan oleh model

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Data yang dikumpulkan mencakup berbagai variabel lingkungan yang berpengaruh terhadap dinamika RTH . Data yang dipakai atau digunakan dalam penelitian ini diperoleh dari SITALA (Sitem Indeks Kualitas Lingkungan Hidup) yang dimana biasa dilihat di website PUSDAL LH. Dimana data ini berisi informasi tentang Jumlah titik RTH perkab/kota di Sumatera per tahun mulai dari 2022, 2023, sampai dengan 2024, dengan berdasarkan data yang berjumlah 454 baris dengan 5 Variabel. Dataset yang disajikan pada gambar 3.

	TAHUN	PROVINSI	KAB/KOTA	JUMLAH TITIK RTH (Y)	LUAS WILAYAH (ha) (X)
0	2024	RIAU	KAB.BENGKALIS	27	'861635.784522764
1	2024	RIAU	KAB.INDRAGIRI HILIR	29	'1352126.28358973
2	2024	RIAU	KAB.INDRAGIRI HULU	16	'787184.992123735
3	2024	RIAU	KAB.KAMPAR	17	'1035280.25193582
4	2024	RIAU	KAB.KUANTAN SINGINGI	77	'545786.23555258
5	2024	RIAU	KAB.KEP MERANTI	0	'362355.847170659
6	2024	RIAU	KAB.PELALAWAN	2	'1326211.32028602
7	2024	RIAU	KAB.ROKAN HILIR	35	'906845.847863578
8	2024	RIAU	KAB.ROKAN HULU	19	'765814.908414686
9	2024	RIAU	KAB.SIAK	15	'780553.764826508
10	2024	RIAU	KOTA DUMAI	43	'205961.275314242
11	2024	RIAU	KOTA PEKANBARU	66	'63833.1321304324

Gambar 2. Tampil Awal Dataset

Pada gambar 3 menampilkan dataset berisi informasi Jumlah Titik RTH perkab/kota yang ada di wilayah ekoregion Sumatera yang terdiri dari fitur, yaitu *Tahun*, *Provinsi*, *Kab/Kota*, *Jumlah Titik RTH*, *Luas Wilayah(ha)*. Berikut penjelasan masing-masing fitur:

1. *Tahun* : Tahun pengambilan data atau waktu data dicatat. Dalam konteks ini adalah tahun 2024-2022. Fitur ini bisa digunakan untuk analisis tren jika ada data tahunan.
2. *Provinsi* : Nama provinsi dari wilayah administratif.
3. *Kab/kota* : Nama wilayah administratif

4. *Jumlah Titik RTH* : Jumlah titik lokasi Ruang Terbuka Hijau (RTH) yang tercatat di wilayah tersebut. Ini adalah target dalam model prediksi.
5. *Luas Wilayah(ha)*: Total luas wilayah daratan kabupaten/kota dalam satuan hektar (ha). Ini digunakan sebagai fitur dalam prediksi jumlah RTH. Biasanya semakin besar wilayah, potensi titik RTH juga meningkat.

3.2 Preprocessing Data

Preprocessing data adalah tahap penting dalam pengolahan data untuk memastikan data yang digunakan bersih, konsisten, dan sesuai untuk dianalisis oleh model prediksi. Berikut tahapan-tahapan preprocessing yang dilakukan:

3.2.1 Pembersihan Data

Pembersihan data adalah langkah awal yang penting dalam proses analisis data untuk memastikan bahwa data yang digunakan bebas dari kesalahan, inkonsistensi, atau nilai yang tidak valid. Tujuan utama dari pembersihan data adalah menghasilkan data yang berkualitas tinggi agar model prediksi yang dibangun dapat bekerja secara optimal.

	TAHUN	PROVINSI	KAB/KOTA	JUMLAH TITIK RTH (Y) \
0	2024	RIAU	KAB.BENGKALIS	27
1	2024	RIAU	KAB.INDRAGIRI HILIR	29
2	2024	RIAU	KAB.INDRAGIRI HULU	16
3	2024	RIAU	KAB.KAMPAR	17
4	2024	RIAU	KAB.KUANTAN SINGINGI	77
5	2024	RIAU	KAB.KEP MERANTI	0
6	2024	RIAU	KAB.PELALAWAN	2
7	2024	RIAU	KAB.ROKAN HILIR	35
8	2024	RIAU	KAB.ROKAN HULU	19
9	2024	RIAU	KAB.SIAK	15
10	2024	RIAU	KOTA DUMAI	43
11	2024	RIAU	KOTA PEKANBARU	66

	LUAS WILAYAH (ha) (X)
0	8.616358e+05
1	1.352126e+06
2	7.871850e+05
3	1.035280e+06
4	5.457862e+05
5	3.623558e+05
6	1.326211e+06
7	9.068458e+05
8	7.658149e+05
9	7.805538e+05
10	2.059613e+05
11	6.383313e+04

Gambar 3. Pembersihan Data Pada Jumlah Titik RTH

Gambar tersebut menunjukkan dataset mentah yang masih mengandung kolom kosong dan kemungkinan nilai yang tidak lengkap. Oleh karena itu, proses pembersihan data sangat penting sebelum dilakukan analisis atau implementasi model seperti *regresi linier* atau *random forest*.

3.2.2 Normalisasi dan Transformasi data

Normalisasi data bertujuan untuk menstandarkan skala variabel numerik agar tidak ada variabel yang mendominasi proses analisis akibat perbedaan skala. Hal ini sangat penting terutama dalam penerapan algoritma regresi linier dan random forest.

Transformasi data dilakukan untuk mengubah data mentah ke dalam format, struktur, atau skala yang lebih sesuai dan optimal untuk proses analisis dan pemodelan.

3.3 Implementasi Algoritma

Tahapan implementasi algoritma merupakan proses penerapan metode prediksi terhadap dataset yang telah melalui tahap *preprocessing*. Pada penelitian ini digunakan dua algoritma yaitu Regresi Linier dan Random Forest untuk memprediksi jumlah titik Ruang Terbuka Hijau (RTH) di wilayah Ekoregion Sumatera. Tujuan implementasi adalah untuk mengevaluasi kinerja masing-masing algoritma dalam menghasilkan prediksi yang akurat.

3.3.1 Implementasi Algoritma Regresi Linier

Proses implementasi algoritma Regresi Linear dilakukan untuk memprediksi Jumlah Titik Ruang Terbuka Hijau (RTH) berdasarkan beberapa fitur utama, yaitu Tahun, Jumlah Titik RTH, Jumlah Wilayah. Model dibangun menggunakan 454 baris data, dengan pemrograman Python melalui platform Google Colab. Data dibagi menjadi 80% sebagai data latih dan 20% sebagai data uji (train-test split) guna melatih dan mengevaluasi kinerja model secara terpisah. Hasil dari pelatihan model ini menghasilkan persamaan Regresi Linear yang ditampilkan pada tabel berikut ini :

Tabel 4. Hasil Persamaan Regresi Linear

Regresi Linear	
<i>Koefisien</i>	[-2.9447807e-05]
<i>Intercept</i>	52.381489398140346

Dimana :

X :Luas Wilayah (ha)

Y : Jumlah Titik RTH

Untuk mengukur kinerja model dalam memprediksi Jumlah Titik RTH, digunakan sejumlah metrik evaluasi yaitu: Koefisien Determinasi (R^2), *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Squared Error* (MSE). Seluruh hasil evaluasi ditampilkan pada tabel berikut :

Tabel 5. Evaluasi Model Regresi Linear

Regresi Linear	
MAE	5.87
RMSE	8.52
R^2	0.45

1. $R^2 = 0.45$

- a. Ini adalah koefisien determinasi, menunjukkan seberapa besar proporsi variasi dalam data target yang dapat dijelaskan oleh model.
- b. Nilai 0.45 berarti 45% variasi dalam data target bisa dijelaskan oleh fitur input (variabel prediktor).
- c. Sisanya (55%) tidak dapat dijelaskan oleh model — mungkin karena variabel lain yang belum dimasukkan atau noise.

2. MAE (*Mean Absolute Error*) = 5.87
 - a. Ini menunjukkan rata-rata kesalahan absolut antara nilai prediksi dan nilai aktual.
 - b. Artinya, rata-rata prediksi model meleset sebesar 5.87 satuan dari nilai sebenarnya.
 - c. Semakin kecil nilainya, semakin baik modelnya.
3. RMSE (*Root Mean Squared Error*) = 8.52
 - a. Ini adalah akar dari rata-rata kuadrat selisih antara nilai prediksi dan nilai aktual.
 - b. RMSE lebih sensitif terhadap kesalahan yang besar.
 - c. Nilai 8.52 berarti prediksi model memiliki deviasi sebesar 8.52 satuan secara umum.

3.3.2 Impelmentasi Algoritma *Random Forest*

Proses implementasi algoritma *Random Forest* dilakukan untuk memprediksi Jumlah Titik Ruang Terbuka Hijau (RTH) berdasarkan beberapa fitur utama, yaitu Tahun, Luas Wilayah. Model dibangun menggunakan 454 baris data, dengan pemrograman Python melalui platform Google Colab. Data dibagi menjadi 80% sebagai data latih dan 20% sebagai data uji (*train-test split*) guna melatih dan mengevaluasi kinerja model secara terpisah. Algoritma *Random Forest* dipilih karena kemampuannya dalam menangani data dengan hubungan non-linier dan memberikan hasil prediksi yang lebih stabil dibandingkan metode *regresi linier*. Untuk mengukur kinerja model dalam memprediksi Jumlah Titik RTH, digunakan sejumlah metrik evaluasi yaitu: Koefisien Determinasi (R^2), *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan *Mean Squared Error* (MSE) [11]. Seluruh hasil evaluasi ditampilkan pada tabel sebagai berikut:

Tabel 6. Evaluasi Model *Random Forest*

<i>Random Forest</i>	
MAE	5.52
RMSE	5.88
R^2	0.74

1. $R^2 = 0.74$
 - a. Mengukur seberapa baik variansi dalam data target bisa dijelaskan oleh model.
 - b. Nilai 1 menunjukkan prediksi sempurna, nilai 0 berarti model tidak lebih baik dari rata-rata.
 - c. 0.74 artinya model mampu menjelaskan 74% variabilitas dalam data target.
2. MAE (*Mean Absolute Error*) = 5.52
 - a. Rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya.
 - b. Interpretasi: Prediksi model meleset sekitar 5.52 unit dari nilai sebenarnya, secara rata-rata.
 - c. Metrik ini tidak peka terhadap outlier.
3. RMSE (*Root Mean Squared Error*) = 5.88
 - a. Akar dari rata-rata kuadrat kesalahan prediksi.
 - b. Lebih sensitif terhadap nilai ekstrem dibandingkan MAE.

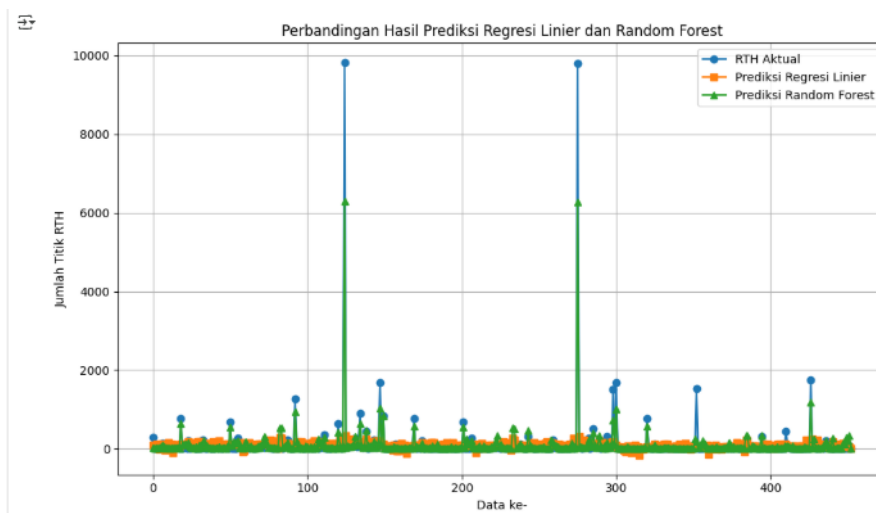
c. Dalam hal ini, error prediksi model sekitar 5.88 unit secara rata-rata.

3.3.3 Hasil Perbandingan Evaluasi Model

Tabel 7. Perbandingan Evaluasi Model

Metrik Evaluasi	Regresi linier	Kategori	Random forest	Kategori
MAE	5.87	Error sedang	5.52	Baik
RMSE	8.52	Error cukup tinggi	5.88	Baik
R^2	0.45	Kurang baik	0.74	Cukup

Berdasarkan hasil evaluasi model menggunakan tiga metrik utama, yaitu MAE, RMSE, dan R^2 , diperoleh bahwa algoritma *Random Forest* memiliki kinerja yang jauh lebih baik dibandingkan *Regresi Linier* dalam memprediksi jumlah titik Ruang Terbuka Hijau (RTH). Secara keseluruhan, model *Random Forest* menunjukkan performa yang lebih baik dibandingkan *Regresi Linier* di semua metrik evaluasi. *Random Forest* memberikan prediksi yang lebih akurat (MAE dan RMSE lebih rendah) dan pemahaman yang lebih baik terhadap data (R^2 lebih tinggi). Oleh karena itu, *Random Forest* lebih direkomendasikan untuk digunakan dalam kasus ini. Visualisasi Perbandingan Hasil *Regresi Linier* dan *Random forest* terdapat pada Gambar 8



Gambar 8. Visualisasi Perbandingan Hasil Regresi Linier dan Random forest

Visualisasi grafik perbandingan hasil prediksi antara model *Regresi Linier* dan *Random Forest* menunjukkan bahwa model *Regresi Linier* menghasilkan prediksi yang cenderung datar dan tidak mengikuti pola variasi nilai aktual jumlah titik RTH, terutama pada data dengan nilai ekstrem atau outlier, sehingga model ini kurang mampu menangkap kompleksitas data dan menghasilkan prediksi yang akurat. Sebaliknya, model *Random Forest* menghasilkan prediksi yang lebih fleksibel dan mampu mengikuti pola data aktual dengan lebih baik, termasuk pada titik-titik puncak seperti pada data ke-100 dan ke-300. Hal ini menunjukkan bahwa *Random Forest* memiliki kemampuan yang lebih baik dalam memodelkan hubungan non-linier antara fitur dan target, serta mampu memberikan hasil prediksi yang lebih stabil dan mendekati nilai sebenarnya

dibandingkan *Regresi Linier*. Oleh karena itu, grafik ini secara visual mendukung hasil evaluasi kuantitatif sebelumnya bahwa *Random Forest* merupakan algoritma yang lebih akurat dan layak digunakan untuk memprediksi jumlah titik Ruang Terbuka Hijau (RTH).

4. KESIMPULAN

Penelitian ini membandingkan dua metode prediksi yaitu Regresi Linier dan Random Forest. Dengan data jumlah titik Ruang Terbuka Hijau (RTH) di wilayah Ekoregion Sumatera. Hasil evaluasi menunjukkan bahwa Regresi Linier memiliki performa yang kurang optimal dengan MAE 5.87, RMSE 8.52, dan R^2 0.45, yang berarti model hanya mampu menjelaskan 45% variasi data. Sebaliknya, Random Forest menunjukkan performa lebih baik dengan MAE 5.52, RMSE 5.88, dan R^2 0.74, menandakan bahwa model ini mampu menjelaskan 74% variasi data secara lebih akurat dan stabil. Visualisasi hasil juga memperlihatkan bahwa Random Forest lebih mampu menyesuaikan pola data aktual dan menangani nilai-nilai ekstrem. Oleh karena itu, Random Forest direkomendasikan sebagai metode yang lebih unggul dalam memprediksi jumlah titik RTH di wilayah ini.

DAFTAR RUJUKAN

- [1] U. Simalungun, "Zonasi Ruang Terbuka Hijau dalam Mendukung Pengelolaan Lingkungan Perkotaan yang Berkelanjutan," vol. 4, no. 2, pp. 2257–2264, 2025.
- [2] Z. Abdi, D. Tampung, and S. Batanghari, "Kajian Daya Tampung Beban Pencemaran Sungai Batanghari Pada Penggal Gasiang – Sungai Langkok Sumatera Barat," *Maj. Geogr. Indones.*, vol. 25, no. 1, pp. 70–94, 2016.
- [3] D. Pramesti and Wiga Maulana Baihaqi, "Perbandingan Prediksi Jumlah Transaksi Ojek Online Menggunakan Regresi Linier Dan Random Forest," *Gener. J.*, vol. 7, no. 3, pp. 21–30, 2023, doi: 10.29407/gj.v7i3.20676.
- [4] S. Samsudi, "Ruang Terbuka Hijau Kebutuhan Tata Ruang Perkotaan Kota Surakarta," *J. Rural Dev.*, vol. 1, no. 1, pp. 11–19, 2019.
- [5] H. Albani Bardian and I. Sutanto, "Pengembangan Aplikasi Vulnerability Scanner Untuk Mendeteksi Celah Keamanan Siber Pada Website," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 9, no. 3, pp. 4404–4411, 2025, doi: 10.36040/jati.v9i3.13656.
- [6] M. Aror, "Menciptakan Ruang Terbuka Hijau dengan Persepsi Keamanan sebagai Elemen Inklusif," pp. 19–32, 2024.
- [7] K. SaThierbach *et al.*, "Faika A Usman," vol. 3, no. 1, pp. 1–15, 2015,
- [8] W. A. Karuru and S. Suryani, "Kajian Elemen Ruang Terbuka Hijau Pada Perumahan Green Mansion, Kabupaten Sidoarjo, Jawa Timur," *Arsitekno*, vol. 12, no. 1, pp. 55–62, 2025, doi: 10.29103/arj.v12i1.16703.
- [9] M. J. Budiman and Fanny Jouke Doringin, "Jurnal Ilmu Komputer," *Biomaterials*, vol. 07, no. 12, pp. 85–90, 2023.
- [10] N. A. Prakoso Indaryono, "Analisa Perbandingan Algoritma Random Forest Dan Naïve Bayes Untuk Klasifikasi Curah Hujan Berdasarkan Iklim Di Indonesia," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 1, pp. 158–167, 2024, doi: 10.29100/jupi.v9i1.4421.
- [11] L. Trihardianingsih and H. Permatasari, "Prediksi Area Kebakaran Hutan Menggunakan Algoritma Random Forest," pp. 37–41, 2024.