

PENGEMBANGAN MODEL PREDIKSI EMISI KARBON DIOKSIDA (CO₂) DAN REKOMENDASI KENDARAAN RENDAH EMISI DENGAN RANDOM FOREST REGRESSION DAN CONTENT-BASED FILTERING

Kristian Abel Baptista, Eko Purwanto, Dwi Hartanti
Teknik Informatika, Universitas Duta Bangsa Surakarta
Jalan Bhayangkara No. 55-57 Kota Surakarta, Indonesia
220103129@mhs.udb.ac.id

ABSTRAK

Pesatnya pertumbuhan transportasi modern menyebabkan peningkatan emisi karbon dioksida (CO₂) yang berdampak signifikan pada penurunan kualitas udara dan perubahan iklim. Penelitian ini bertujuan untuk mengembangkan sistem yang mampu memprediksi emisi kendaraan secara akurat serta memberikan rekomendasi kendaraan alternatif dengan emisi lebih rendah berdasarkan spesifikasi kendaraan. *Dataset* yang digunakan berasal dari *platform* Kaggle mencakup 833 data kendaraan. Tahapan penelitian meliputi *preprocessing* data dengan mengubah data kategorikal menjadi numerik, kemudian menerapkan algoritma *random forest regression* untuk memodelkan prediksi emisi kendaraan. Selain itu, metode *content-based filtering* digunakan untuk merekomendasikan kendaraan serupa yang memiliki emisi lebih rendah berdasarkan tingkat kemiripan data. Hasil pengujian menunjukkan bahwa model mampu menghasilkan prediksi dengan akurasi sangat tinggi pada konfigurasi 100 *trees*, dengan nilai R² sebesar 0.9973 dan RMSE sebesar 2.9983. Temuan penting dalam penelitian ini menunjukkan bahwa konsumsi bahan bakar (*fuel consumption*) merupakan fitur paling berpengaruh terhadap jumlah emisi yang dihasilkan. Dengan demikian, integrasi kedua metode ini dinilai efektif dalam mendukung upaya pengurangan emisi di sektor transportasi serta berpotensi untuk dikembangkan menggunakan *dataset* yang lebih luas dan beragam.

Kata kunci : prediksi, sistem rekomendasi, random forest regression, content-based filtering, emisi CO₂

1. PENDAHULUAN

Perkembangan transportasi modern yang semakin pesat telah memberikan kemudahan bagi masyarakat dalam melakukan aktivitas sehari-hari. Namun, peningkatan penggunaan kendaraan berbahan bakar fosil juga berdampak pada meningkatnya emisi karbon dioksida (CO₂) yang dapat memengaruhi kualitas udara dan mempercepat perubahan iklim. Peningkatan jumlah kendaraan berbasis bahan bakar fosil menyebabkan konsumsi energi semakin tinggi dan menghasilkan emisi karbon dioksida (CO₂) yang memperburuk kondisi lingkungan [1].

Mengingat emisi karbon yang tinggi berasal dari berbagai aktivitas seperti kendaraan, kebakaran hutan dan kegiatan industri, dengan sektor kendaraan sebagai sumber terbesar [2].

Oleh karena itu, diperlukan analisis emisi untuk memahami kondisi lingkungan serta dampak yang dapat terjadi pada masa mendatang.

Permasalahan muncul dalam upaya menentukan model prediksi dan rekomendasi yang efisien namun tetap akurat dan relevan. Sejumlah penelitian sebelumnya, termasuk yang dilakukan oleh Niken Larasati telah menggunakan metode *neural network* untuk memprediksi emisi kendaraan berdasarkan karakteristik teknisnya karena mampu menghasilkan prediksi yang akurat, meskipun membutuhkan sumber daya komputasi yang besar dan memiliki alur logika yang cukup rumit untuk diurai [3].

Selain itu, tantangan dalam pemodelan data yang kompleks adalah bagaimana memastikan aturan

keputusan dalam variabel tetap stabil agar tidak menghasilkan prediksi yang inkonsisten [4].

Masalah lain juga muncul pada aspek penyediaan rekomendasi, di mana kualitas rekomendasi sangat bergantung pada kelengkapan dan kualitas data fitur yang tersedia agar tidak terjadi kesalahan target rekomendasi [5].

Oleh karena itu, Penelitian ini bertujuan untuk mengembangkan sebuah sistem terintegrasi yang mampu memprediksi emisi karbon dioksida (CO₂) serta merekomendasikan kendaraan dengan emisi yang lebih rendah. Pengembangan sistem ini diharapkan dapat menjadi langkah awal dalam memahami dampak emisi kendaraan terhadap lingkungan dan menyediakan alternatif kendaraan yang lebih ramah lingkungan, berkontribusi pada upaya pengurangan emisi, serta menjadi dasar bagi pengembangan penelitian di bidang transportasi berkelanjutan dan peningkatan kualitas lingkungan di masa mendatang.

Sebagai rencana penyelesaian, penelitian ini menerapkan dua tahapan. Tahap pertama adalah menerapkan *random forest regression* yang digunakan untuk mencapai prediksi secara akurat melalui penggabungan banyak pohon keputusan [6].

Selanjutnya, sistem akan memberikan rekomendasi menggunakan metode *content-based filtering* karena sangat efektif dalam menyarankan kendaraan berdasarkan kemiripan spesifikasi teknisnya. Metode ini sangat praktis karena sistem dapat memberikan saran yang relevan hanya dengan

mengandalkan detail atribut kendaraan tanpa perlu menunggu adanya penilaian (*rating*) dari pengguna [7].

Dengan pengelolaan parameter yang tepat pada *random forest regression* dan perbandingan fitur yang mendalam pada *content-based filtering*, sistem ini akan menjadi solusi yang seimbang antara akurasi prediksi dan relevansi rekomendasi.

2. TINJAUAN PUSTAKA

Pada bab ini disajikan kajian literatur yang meliputi tiga bagian utama, yaitu landasan teori sebagai dasar dalam memahami penelitian, tinjauan terhadap penelitian terdahulu yang relevan dengan topik pembahasan, serta analisis *gap* penelitian untuk mengidentifikasi ruang pengembangan dan kebaruan yang diangkat dalam penelitian ini. Adapun pembahasan bab ini dijelaskan sebagai berikut:

2.1. Emisi CO₂ dan Emisi Pada Kendaraan

Emisi karbon dioksida (CO₂) merupakan salah satu gas rumah kaca yang berperan dalam peningkatan suhu global karena kemampuannya menjebak panas di atmosfer [8].

Pada sektor transportasi, terutama CO₂ dapat dihasilkan dari proses pembakaran bahan bakar kendaraan. Jumlah emisi yang dihasilkan tidak hanya dipengaruhi oleh jenis bahan bakar, tetapi juga oleh karakteristik mesin, teknologi pembakaran, tingkat konsumsi bahan bakar, serta pola penggunaan kendaraan. Dampaknya tidak terbatas pada perubahan iklim, tetapi juga memengaruhi kesehatan manusia serta kondisi lingkungan secara umum [9].

Oleh karena itu, pengelolaan dan pengungkapan emisi CO₂ menjadi langkah penting dalam upaya mengendalikan jumlah emisi yang dihasilkan [10].

Estimasi emisi yang akurat dapat menjadi dasar dalam perumusan strategi pengurangan emisi, termasuk peningkatan efisiensi kendaraan dan penyusunan kebijakan transportasi yang lebih ramah lingkungan.

2.2. Machine Learning dan Pemodelan Prediktif

Machine Learning merupakan metode komputasi yang memungkinkan sistem mempelajari pola dari data dan menghasilkan prediksi secara otomatis, menyerupai proses belajar manusia dari pengalaman [4].

Dalam penelitian prediksi emisi, pendekatan ini digunakan untuk menganalisis hubungan antara karakteristik kendaraan seperti kapasitas mesin, jenis bahan bakar, dan tingkat konsumsi bahan bakar dengan estimasi emisi CO₂ yang dihasilkan.

Dalam konteks kecerdasan buatan, *machine learning* dikembangkan dengan landasan algoritma, matematika, dan statistika yang mendukung proses analisis data secara sistematis [11].

Pendekatan ini efektif dalam menangani hubungan antarvariabel yang kompleks dan tidak selalu menunjukkan pola yang tetap, sehingga mampu

menghasilkan prediksi yang lebih adaptif dan mendekati kondisi nyata.

2.3. Random Forest Regression

Random Forest Regression merupakan suatu algoritma pembelajaran terawasi yang menggunakan data berlabel sebagai acuan pelatihan, termasuk dalam keluarga *decision tree* dan bekerja dengan mengombinasikan sejumlah pohon keputusan untuk menghasilkan prediksi yang lebih konsisten dan akurat [12].

Model ini dibangun dengan membuat sejumlah pohon regresi secara bersamaan, masing-masing dilatih menggunakan sampel data yang dipilih secara acak, kemudian hasil akhirnya diperoleh dari rata-rata seluruh nilai prediksi setiap pohon [13].

Random Forest Regression sering digunakan untuk mendukung metode *content-based filtering* dalam meningkatkan kualitas rekomendasi, karena setiap pohon belajar dari kombinasi fitur yang berbeda sehingga mampu memperkuat akurasi prediksi.

Random Forest Regression menggabungkan sejumlah *decision tree* dalam satu model untuk meningkatkan akurasi prediksi serta mengurangi risiko *overfitting* yang sering terjadi pada *decision tree* tunggal [6].

Dalam konteks prediksi emisi karbon dioksida (CO₂), algoritma ini dipilih karena mampu mengolah banyak fitur sekaligus, menghasilkan estimasi yang konsisten, serta menyediakan informasi mengenai tingkat kepentingan variabel (*feature importance*) untuk mendukung interpretasi model.

2.4. Sistem Rekomendasi

Sistem rekomendasi merupakan sistem yang memanfaatkan data historis untuk menyajikan informasi dan saran yang relevan guna memberikan referensi pilihan bagi pengguna secara lebih efektif [14].

Sistem ini bekerja dengan menganalisis pola data melalui proses penyaringan (*filtering*) dan pengurutan (*ranking*) guna menghasilkan rekomendasi yang sesuai dengan kebutuhan pengguna. Hal ini bertujuan untuk mengurangi kompleksitas informasi dan meningkatkan efisiensi.

Dalam bidang transportasi, sistem rekomendasi dapat digunakan untuk menyarankan kendaraan dengan tingkat emisi yang lebih rendah berdasarkan data teknis kendaraan dan estimasi emisi yang dihasilkan. Dengan pendekatan tersebut, sistem mampu membantu pengguna dalam memilih kendaraan yang lebih efisien dan ramah lingkungan [3].

2.5. Content-Based Filtering

Content-Based Filtering merupakan metode rekomendasi yang memberikan saran *item* dengan mencocokkan atribut atau karakteristik *item* terhadap preferensi dan riwayat minat pengguna agar hasilnya lebih relevan [15].

Prosesnya dilakukan dengan menghitung tingkat kemiripan data untuk menentukan opsi yang relevan. Berbeda dari pendekatan berbasis kolaboratif, metode ini tidak bergantung pada perilaku pengguna lain melainkan sepenuhnya menyesuaikan rekomendasi berdasarkan karakteristik dan minat pengguna itu sendiri [7].

Dalam pendekatan ini, setiap *item* diberi atribut atau label yang mempresentasikan karakteristik dan isi utamanya agar dapat dikenali serta dibandingkan berdasarkan kesamaan fitur yang dimiliki [5].

Ketika dikombinasikan dengan prediksi emisi menggunakan model *random forest regression*, sistem dapat secara konsisten merekomendasikan kendaraan dengan emisi yang lebih rendah tanpa mengurangi relevansi terhadap preferensi pengguna.

2.6. Tinjauan Penelitian Terkait

Terkait prediksi emisi, penelitian oleh Niken Larasati dalam menerapkan algoritma *neural network* untuk mengestimasi emisi kendaraan dan memberikan rekomendasi alternatif berdasarkan batasan harian dengan akurasi cukup tinggi [3].

Meskipun memiliki fokus yang serupa, penelitian ini menggunakan *random forest regression* untuk mempermudah interpretasi hasil prediksi. Selain itu, penggunaan *content-based filtering* dalam penelitian ini difokuskan pada kecocokan karakteristik teknis kendaraan agar rekomendasi yang dihasilkan lebih spesifik dan relevan bagi pengguna.

Dalam bidang sistem rekomendasi, penelitian oleh Oktavianda, Lusiana Efrizoni, Eiva Fatdha, dan Hadi Asnal dalam mengembangkan sistem rekomendasi *video game* berbasis usia pada *platform* Steam dengan memanfaatkan metode *content-based filtering* dan *cosine similarity* yang memperoleh nilai *precision* 0.98 dan *recall* 1.00 serta *error* yang tergolong rendah berdasarkan MAE, MSE, dan RMSE [16].

Hasil penelitian referensi tersebut menunjukkan bahwa sistem mampu menyajikan rekomendasi yang akurat dan sesuai dengan kategori usia pengguna. Berbeda dengan penelitian referensi tersebut yang hanya memberikan rekomendasi secara akurat dan relevan. Pada penelitian ini memanfaatkan *random forest regression* untuk memperkirakan emisi karbon dioksida (CO₂) serta mengintegrasikannya dengan *content-based filtering* digunakan untuk merekomendasikan kendaraan dengan tingkat emisi lebih rendah secara lebih terukur dan berbasis hasil prediksi.

Mengenai efektivitas algoritma, penelitian oleh Siska Septiyanah dan Ghita Athalina yang menunjukkan *random forest regression* lebih unggul dibanding regresi linear dalam memprediksi harga laptop dengan mempertimbangkan fitur teknis [6].

Penelitian ini memanfaatkan keunggulan algoritma yang sama untuk memprediksi emisi karbon dioksida (CO₂) pada kendaraan berdasarkan fitur teknis. Meski memiliki bidang yang berbeda, kedua

penelitian menekankan kemampuan *random forest regression* dalam mengelola variabel numerik dan kategorikal untuk menghasilkan prediksi akurat.

Penelitian referensi oleh Muhammad Fiarus Ramadhani yang menunjukkan efektivitas metode *content-based filtering* berbasis TF-IDF dalam mengidentifikasi *developer* yang sesuai berdasarkan kesamaan pola kerja historis mereka dengan akurasi *hit rate* sebesar 84.4% [17].

Sementara itu, penelitian ini diterapkan pada bidang yang berbeda yakni prediksi emisi karbon dioksida (CO₂) dan rekomendasi kendaraan dengan emisi yang lebih rendah. Perbedaan utamanya terletak pada integrasi model, penelitian ini menggabungkan *random forest regression* untuk menghitung perkiraan emisi secara akurat serta *content-based filtering* digunakan untuk memberikan rekomendasi berdasarkan atribut teknis kendaraan, bukan mengandalkan analisis teks aktivitas harian seperti pada penelitian referensi tersebut

Begitu pula dengan penelitian oleh Nailatul Azizah dan Anief Fauzan Rozi yang menunjukkan bahwa metode *content-based filtering* efektif dalam memberikan rekomendasi produk yang akurat dengan tingkat kemiripan (*similarity*) mencapai 0.722 [7].

Meskipun terdapat kesamaan dalam penggunaan metode rekomendasi, penelitian ini berfokus pada bidang yang berbeda, yakni prediksi emisi karbon dioksida (CO₂) serta merekomendasikan kendaraan rendah emisi. Perbedaan utamanya terletak pada integrasi *random forest regression* untuk menghitung estimasi emisi secara terukur berdasarkan atribut teknis kendaraan, sehingga tidak hanya mengandalkan analisis teks deskripsi produk seperti pada penelitian referensi tersebut.

2.7. Gap Penelitian

Berdasarkan analisis terhadap penelitian sebelumnya, sebagian besar penelitian cenderung masih berfokus pada satu aspek tertentu, seperti prediksi menggunakan algoritma *machine learning* atau sistem rekomendasi berbasis *content-based filtering*. Meskipun masing-masing metode telah mencapai tingkat akurasi yang memuaskan, tetapi belum mengintegrasikan kedua pendekatan tersebut dalam satu sistem yang saling mendukung. Selain itu, penerapan metode rekomendasi umumnya masih terbatas pada analisis teks, kategori pengguna, atau pola historis tertentu.

Di sisi lain, terdapat penelitian yang menunjukkan bahwa *random forest regression* mampu menghasilkan prediksi yang akurat dalam mengelola data numerik dan kategorikal. Namun, penerapan algoritma tersebut masih digunakan pada bidang selain prediksi emisi kendaraan, sehingga pemanfaatannya dalam memperkirakan emisi karbon dioksida (CO₂) berdasarkan atribut teknis kendaraan masih belum banyak dibahas. Pada penelitian sebelumnya pula masih belum menitikberatkan pada kemudahan

interpretasi hasil prediksi serta rekomendasi kendaraan rendah emisi secara lebih terukur.

Berdasarkan *gap* tersebut, penelitian ini mengisi celah tersebut dengan menggabungkan *random forest regression* dan *content-based filtering* dalam satu sistem terintegrasi. Pendekatan ini memungkinkan sistem untuk terlebih dahulu dalam menghitung estimasi emisi karbon dioksida (CO₂) berdasarkan karakteristik teknis kendaraan, kemudian menggunakan hasil tersebut sebagai dasar untuk memberikan rekomendasi kendaraan alternatif dengan emisi yang lebih rendah. Dengan demikian, solusi yang dihasilkan pada penelitian ini menjadi lebih lengkap dan terstruktur karena menggabungkan prediksi dan rekomendasi sesuai dengan relevansi kebutuhan pengguna untuk mendukung efisiensi energi secara nyata.

3. METODE PENELITIAN



Gambar 1. Diagram alur tahapan penelitian

Gambar di atas menunjukkan diagram alur tahapan penelitian yang digunakan dalam pengembangan model untuk memprediksi emisi karbon dioksida (CO₂) dan memberikan rekomendasi kendaraan dengan emisi lebih rendah. Berikut adalah penjelasan mendalam mengenai setiap tahapannya:

3.1. Pengumpulan Data

Pada tahap awal penelitian ini, *dataset* yang digunakan bersumber dari Kaggle yang mencakup sekitar 833 data spesifikasi kendaraan serta emisi karbon dioksida (CO₂) yang dihasilkan, guna memastikan model memiliki cakupan informasi yang representatif. Data tersebut tersedia dan dapat diakses secara publik melalui tautan berikut [18]: (<https://www.kaggle.com/datasets/mcpenguin/canadian-fuel-consumption-ratings>).

3.2. Preprocessing Data

Sebelum data diolah lebih lanjut, serangkaian tindakan pembersihan dan persiapan dilakukan untuk meningkatkan kualitas *input*. Proses ini mencakup:

- Menentukan fitur yang relevan, yaitu *make*, *model*, *vehicle class*, *engine size* (L), *cylinders*, *transmission*, *fuel type*, dan *fuel consumption* (comb) (L/100km), dengan CO₂ *emissions* (g/km) sebagai variabel target utama.

- Pengambilan kategori unik dari atribut *transmission* dan *fuel type* untuk menyesuaikan parameter *input* sistem.
- Transformasi data kateogikal menggunakan *one-hot encoding* untuk mengubah data menjadi bentuk numerik.
- Membagi data menjadi 80% data latih (*training*) dan 20% data uji (*testing*).

3.3. Pemodelan Prediksi Emisi

Pada fase ini, dilakukan proses pengembangan model untuk memprediksi emisi karbon dioksida (CO₂) menggunakan algoritma *random forest regression*. Tahapan ini, meliputi:

- Model yang dilatih menggunakan algoritma *random forest regression* dengan mekanisme *ensemble*, untuk menghasilkan prediksi emisi karbon dioksida (CO₂) dalam satuan (g/km). Nilai pemodelan prediksi tersebut diperoleh dengan menghitung rata-rata hasil dari seluruh pohon keputusan yang terbentuk, dengan formulasi sebagai berikut [19]:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T h_t(x) \quad (1)$$

Keterangan:

$\hat{y}$ = Prediksi akhir.

T = Jumlah pohon (*tree*).

$h_t(x)$ = Prediksi dari pohon ke-t.

$\sum_{t=1}^T$ = Penjumlahan semua prediksi.

$\frac{1}{T}$ = Rata-rata aritmetika.

- Evaluasi model digunakan untuk mengukur kinerja dan akurasi model yang dikembangkan menggunakan tiga evaluasi model, yang mencakup sebagai berikut [19]:

- Mean Absolute Error (MAE)**

MAE digunakan untuk menghitung rata-rata selisih kesalahan antara nilai prediksi dengan nilai aktual, yang dirumuskan sebagai berikut:

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

Keterangan:

n = Jumlah data.

y_i = Nilai aktual ke-i.

$\hat{y}_i$ = Nilai prediksi ke-i.

$|y_i - \hat{y}_i|$ = Besarnya selisih absolut antara data aktual dan data hasil prediksi.

- Root Mean Square Error (RMSE)**

RMSE berfungsi mengukur besarnya penyimpangan nilai prediksi terhadap nilai aktual, yang dirumuskan sebagai berikut:

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

Keterangan:

n = Jumlah data.

y_i = Nilai aktual ke-i.

$\hat{y}_i$ = Nilai prediksi ke-i.

$(y_i - \hat{y}_i)^2$ = Besarnya selisih absolut antara data aktual dan data hasil prediksi.

- **R-Squared (R^2)**

R^2 digunakan untuk menunjukkan seberapa akurat model dalam mencocokkan variasi data, menggunakan perhitungan di bawah ini:

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} \quad (4)$$

Keterangan:

$SS_{res} = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ = Residual sum of squares (galat model).

$SS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$ = Total sum of squares (variansi total).

$\hat{y}_i$ = Nilai prediksi model.

$\bar{y}_i$ = Rata-rata observasi.

n = Jumlah sampel.

- c. Visualisasi model digunakan untuk menyajikan performa model melalui *actual vs predicted*, *feature importance*, *residual plot*, dan interpretasi model untuk memahami logika keputusan sistem.

3.4. Persiapan Sistem Rekomendasi

Setelah model prediksi terbentuk, selanjutnya merancang mekanisme pencarian kesamaan antar kendaraan agar sistem dapat memberikan rekomendasi yang relevan. Langkah-langkahnya, meliputi:

- a. Normalisasi data dalam melakukan standarisasi data dengan menggunakan *standard scaler* agar skala fitur seragam.
- b. Mengukur jarak antar data dengan Euclidean *Distance* yang memerhatikan perbedaan skala nilai pada tiap fiturnya, melalui rumus perhitungan sebagai berikut [17]:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

Keterangan:

$d(x, y)$ = Jarak Euclidean.

x_i = Nilai pada titik 1.

y_i = Nilai pada titik 2.

- c. Mengonversi nilai jarak tersebut menjadi tingkat kemiripan (*similarity*), dengan formulasi sebagai berikut [20]:

$$\text{Sim}(x, y) = \frac{1}{1 + \sqrt{d(x, y)}} \quad (6)$$

Keterangan:

$\text{Sim}(x, y)$ = Nilai kesamaan antara pengguna x dan pengguna y.

$d(x, y)$ = Jarak Euclidean antara pengguna x dan pengguna y.

3.5. Implementasi Sistem

Pada tahap akhir adalah mengintegrasikan seluruh komponen menjadi sebuah sistem pemodelan. Alur kerjanya adalah sebagai berikut:

- a. *Input* pengguna: Pengguna memasukkan spesifikasi kendaraan seperti *engine size* (L), *cylinders*, *fuel consumption* (comb) (L/100km), *fuel type*, dan *transmission*.
- b. Proses sistem: Melakukan validasi *input*, memprediksi emisi dengan menggunakan *random forest regression* yang telah dilatih, serta menghitung kemiripan karakteristik kendaraan

- c. *Filtering*: Sistem secara otomatis melakukan filter kendaraan yang memiliki emisi lebih rendah dari hasil prediksi emisi karbon dioksida (CO_2) serta menyesuaikan dengan tipe *transmission* dan *fuel type* yang serupa.
- d. *Ranking*: Menentukan urutan hasil akhir melalui skor berbobot yang menyeimbangkan kemiripan karakteristik dengan efisiensi emisi. Proses ini memungkinkan sistem mampu merekomendasikan *top 10* kendaraan yang tidak hanya relevan secara spesifikasi, tetapi juga lebih ramah lingkungan.
- e. Implementasi program: Seluruh rangkaian kode program dan sistem pemodelan ini telah disusun secara sistematis di Google Colab dan dapat diakses melalui tautan berikut ini: (<https://github.com/Kriss245/Prediksi-Emisi-CO2-dan-Rekomendasi-Kendaraan-Rendah-Emisi-di-Google-Colab>).

4. HASIL DAN PEMBAHASAN

Bagian ini menguraikan hasil pengujian model dan perbandingan data aktual dengan prediksi untuk melihat tingkat akurasi. Selain itu, mencakup identifikasi fitur-fitur yang paling memengaruhi hasil serta analisis terhadap *error* guna memahami karakteristik performa model secara mendalam.

Selanjutnya, diuraikan pula implementasi ke dalam sistem beserta analisis hasilnya secara menyeluruh. Penjelasan detail mengenai masing-masing tahapan tersebut akan dipaparkan pada sub-bab sebagai berikut:

4.1. Hasil Evaluasi Model

Untuk menentukan konfigurasi model yang paling optimal dalam memprediksi data, dilakukan pengujian terhadap beberapa variasi jumlah estimator (*tree*). Perbandingan performa model tersebut disajikan pada tabel di bawah ini:

Tabel 1. Hasil evaluasi model

Jumlah Estimator (Tree)	MAE	RMSE	R^2
50	1.471700	3.044800	0.997200
100	1.442000	2.998300	0.997300
200	1.432200	3.003000	0.997300

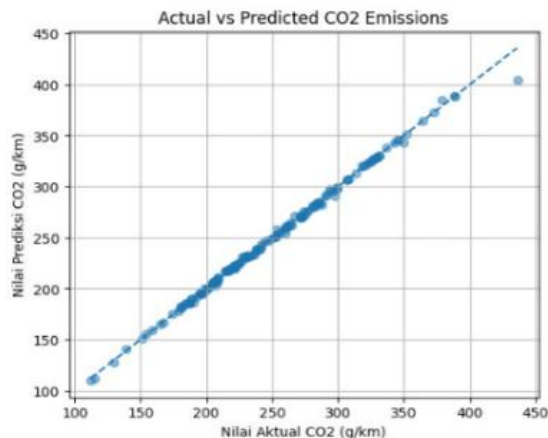
Berdasarkan hasil pengujian, model *random forest regression* memiliki performa yang sangat akurat dengan nilai R^2 yang konsisten di atas 0.99. Hal ini mengindikasikan bahwa model mampu memprediksi target dengan sangat presisi dan memiliki tingkat akurasi prediksi yang hampir sempurna.

Di antara berbagai skenario tersebut, konfigurasi dengan 100 estimator (*tree*) dipilih sebagai konfigurasi paling optimal karena berhasil mencapai nilai RMSE terendah sebesar 2.9983 serta stabilitas skor R^2 sebesar 0.9973. Pemilihan ini juga mempertimbangkan efisiensi komputasi, mengingat penambahan jumlah pohon hingga 200 tidak lagi memberikan peningkatan

akurasi yang signifikan namun justru memperberat beban pemrosesan data. Dengan demikian, model dengan 100 pohon dianggap paling ideal karena berhasil menyeimbangkan antara tingkat kesalahan yang rendah dan stabilitas prediksi.

4.2. Perbandingan Data Aktual dan Prediksi

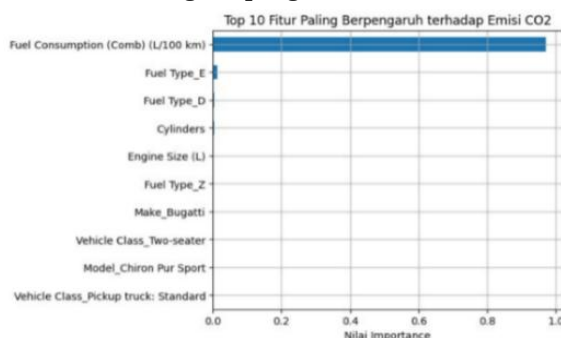
Guna melihat sejauh mana performa model dalam memperkirakan tingkat emisi, visualisasi berikut menyajikan perbandingan antara data aktual dan hasil prediksi yang dapat diamati melalui grafik di bawah ini:



Gambar 2. Perbandingan data aktual dan prediksi

Grafik di atas menunjukkan bahwa sejauh mana hasil prediksi model mendekati data aktual. Secara visual, mayoritas titik data terkonsentrasi rapat pada garis diagonal, yang mengindikasikan hasil prediksi mendekati nilai aslinya. Kedekatan titik terhadap garis mencerminkan minimalnya selisih (*error*) prediksi, sementara sebaran data yang konsisten di sepanjang jalur membuktikan bahwa model mampu bekerja dengan stabil dan memiliki tingkat kesalahan yang sangat rendah.

4.3. Fitur Paling Berpengaruh



Gambar 3. Fitur paling berpengaruh

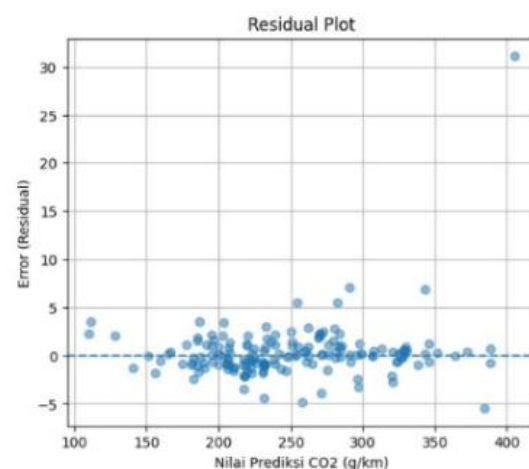
Berdasarkan hasil pemodelan yang telah dilakukan, diperoleh rincian mengenai variabel yang paling memengaruhi tingkat emisi karbon dioksida (CO_2) dapat dijabarkan sebagai berikut:

Visualisasi pada grafik di atas, menunjukkan bahwa fitur *fuel consumption* (comb) (L/100km)

memiliki dominasi yang sangat kuat dalam memprediksi emisi karbon dioksida (CO_2). Hal ini terlihat dari nilai *importance*, fitur tersebut berada di posisi teratas yang memiliki pengaruh terbesar dibandingkan dengan variabel lainnya. Sesuai dengan prinsip nilai *importance*, yang di mana semakin tinggi posisi dan nilai pada suatu variabel tersebut, maka akan semakin besar pula kontribusinya terhadap akurasi prediksi model. Meskipun terdapat variabel lain seperti spesifikasi mesin atau jenis bahan bakar kendaraan tersebut dalam proses pemodelan ini.

4.4. Analisis Error

Untuk mengevaluasi ketepatan model, visualisasi berikut menggambarkan sebaran titik *error* untuk membantu mengenali adanya penyimpangan data sebelum berdampak pada keakuratan sistem yang dapat diamati melalui grafik di bawah ini:



Gambar 4. Analisis error (*residual plot*)

Berdasarkan grafik di atas, menunjukkan bahwa analisis *error* digunakan untuk melihat kualitas model regresi dalam memprediksi emisi. Sebagian besar titik data menyebar secara acak di sekitar garis nol yang menandakan model memiliki tingkat akurasi yang konsisten di berbagai rentang nilai prediksi.

Namun, terdapat satu titik yang menyimpang cukup jauh berada di bagian kanan atas. Data tersebut perlu diperhatikan karena kemungkinan merupakan data yang menyimpang mampu memengaruhi kinerja model dan performa sistem secara keseluruhan.

4.5. Implementasi Sistem

Implementasi sistem prediksi emisi karbon dioksida (CO_2) kendaraan yang dilengkapi dengan fitur rekomendasi kendaraan alternatif rendah emisi, ditunjukkan pada gambar berikut ini:

Berdasarkan gambar 5 menunjukkan bahwa pengguna terlebih dahulu memasukkan spesifikasi kendaraan seperti ukuran mesin, jumlah silinder, konsumsi bahan bakar, jenis bahan bakar, dan transmisi. Setelah itu, proses prediksi akan dilakukan dan sistem akan menampilkan hasil prediksi emisi karbon dioksida (CO_2) kendaraan.

Input Prediksi CO2

Engine (L)

Cylinders

Fuel Consumption

Filter

Fuel Type

Transmission

Prediksi & Rekomendasi

===== HASIL PREDIKSI =====

CO2 Emission: 199.37 g/km

===== TOP 10 REKOMENDASI KENDARAAN =====

No.	ID	Make	Model	Vehicle Class	Fuel Type	Transmission	CO2 Emissions (g/km)	Tingkat Rekomendasi
1	657	MINI	Cooper 3 Door	Subcompact	Z	M6	177	Rendah
2	659	MINI	Cooper 5 Door	Subcompact	Z	M6	177	Sedang
3	662	MINI	Cooper Convertible	Minicompact	Z	M6	177	Rendah
4	396	Honda	Civic Sedan Si	Mid-size	Z	M6	180	Tinggi
5	2	Acura	Integra A-SPEC	Full-size	Z	M6	181	Tinggi
6	601	Mazda	MX-5 (SiL)	Two-seater	Z	M6	189	Sedang

===== KETERANGAN WARNA EMISI CO2 =====

- Tingkat emisi rendah
- Tingkat emisi sedang
- Tingkat emisi tinggi

===== KETERANGAN TINGKAT REKOMENDASI =====

- Tinggi : Sangat direkomendasikan
- Sedang : Cukup direkomendasikan
- Rendah : Kurang direkomendasikan

Gambar 5. Implementasi sistem

Selanjutnya, sistem akan memberikan *top 10* rekomendasi kendaraan alternatif yang memiliki emisi lebih rendah dibandingkan hasil prediksi emisi sebelumnya. Sistem juga akan menampilkan tingkat rekomendasi berdasarkan rendahnya emisi pada kendaraan serta kemiripan spesifikasi kendaraan dengan *input* dari pengguna.

4.6. Analisis Hasil

Berdasarkan hasil pengujian, model *random forest regression* mampu menghasilkan performa prediksi yang sangat baik dengan nilai R^2 mencapai 0.9973. Konfigurasi dengan estimator 100 pohon dipilih sebagai parameter optimal karena mampu menghasilkan tingkat kesalahan (*error*) paling rendah, yakni nilai RMSE sebesar 2.9983. Pilihan ini diambil untuk menjaga efisiensi beban kerja sistem tanpa mengorbankan kualitas prediksi, mengingat penambahan jumlah pohon lebih dari itu tidak lagi memberikan dampak peningkatan performa yang signifikan.

Secara visual, kualitas prediksi ini dibuktikan melalui grafik perbandingan data aktual dan hasil prediksi yang di mana mayoritas titik data berada dekat dengan garis diagonal menandakan bahwa hasil prediksi model sangat mendekati nilai sebenarnya. Hal ini menunjukkan bahwa model memiliki tingkat kesalahan yang rendah dan mampu bekerja secara konsisten dalam menunjukkan sebagian besar *residual* tersebar di sekitar garis nol sehingga model dinilai cukup stabil pada berbagai rentang data meskipun masih terdapat beberapa data menyimpang yang berpotensi memengaruhi performa model.

Pada tahap implementasi, sistem mampu memprediksi emisi karbon dioksida (CO_2) berdasarkan spesifikasi kendaraan yang dimasukkan pengguna, kemudian memberikan *top 10* rekomendasi kendaraan alternatif dengan emisi lebih rendah dibandingkan hasil prediksi sebelumnya serta menampilkan tingkat rekomendasi berdasarkan

rendahnya emisi dan kemiripan spesifikasi kendaraan, sehingga hasil rekomendasi menjadi lebih relevan

5. KESIMPULAN DAN SARAN

Berdasarkan penelitian yang telah dilakukan, sistem prediksi emisi karbon dioksida (CO_2) kendaraan berbasis *random forest regression* berhasil dikembangkan dan mampu memberikan hasil prediksi emisi dengan baik berdasarkan spesifikasi kendaraan yang dimasukkan oleh pengguna. Sistem juga mampu memberikan rekomendasi kendaraan alternatif dengan emisi lebih rendah serta tingkat rekomendasi yang relevan, sehingga dapat membantu dalam memperoleh informasi kendaraan yang lebih ramah lingkungan dan mudah dipahami. Untuk pengembangan selanjutnya, sistem disarankan menggunakan *dataset* kendaraan yang lebih beragam, menambahkan parameter kendaraan, serta membandingkan performa metode dan algoritma *machine learning* lainnya agar hasil prediksi dan rekomendasi menjadi lebih optimal. Selain itu, tampilan antarmuka dan fitur visualisasi juga dapat dikembangkan agar sistem lebih interaktif, informatif, dan mudah digunakan.

DAFTAR PUSTAKA

- [1] D. P. Kurniawan, G. K. Jarek, and J. K. Putra, "EFISIENSI ENERGI DAN EMISI KARBON TRANSPORTASI: STUDI KASUS RUTE JAKARTA-BANDUNG," *Jurnal Inteltek Insan Cendikia*, vol. 2, no. 9, pp. 16576–16587, 2025.
- [2] A. Setiyono, A. Dharmawan, and Z. Furqon, "Prediksi Faktor Penyebab Emisi CO2 pada Kendaraan Menggunakan Gradient Boosting Regression," *Jurnal Fokus Elektroda: Energi Listrik, Telekomunikasi, Komputer, Elektronika dan Kendali*, vol. 9, no. 4, pp. 243–247, 2024.
- [3] N. Larasati, "Prediksi Emisi Karbon Kendaraan Pribadi dan Rekomendasi Kendaraan Alternatif Menggunakan Machine Learning dengan Model Neural Network," *Proceedings of Life and Applied Sciences*, vol. 1, 2023.
- [4] F. E. Penalun, A. Hermawan, and D. Avianto, "Perbandingan random forest regression dan support vector regression pada prediksi laju penguapan," *Jurnal Fasikom*, vol. 13, no. 02, pp. 104–111, 2023.
- [5] K. Crismastiana, M. Raihan, D. Z. P. Raisya, and A. H. Wahyu, "Sistem Rekomendasi Pekerjaan di bidang IT Menggunakan Algoritma Content-Based Filtering," *JOURNAL OF CREATIVE STUDENT RESEARCH Yupeidumenu: Politeknik Pratama Purwokerto*, vol. 1, no. 6, pp. 78–88, 2023.
- [6] S. Septiyanah and G. Athalina, "Implementasi Algoritma Random Forest Regression Dalam Prediksi Harga Laptop," *Jurnal Mnemonic*, vol. 8, no. 1, pp. 70–73, 2025.
- [7] N. Azizah and A. F. Rozi, "Sistem Rekomendasi Produk Somethinc Menggunakan Metode Content-based Filtering," *Jurnal Teknologi Dan*

- Sistem Informasi Bisnis*, vol. 6, no. 3, pp. 461–468, 2024.
- [8] H. A. Nurdin, N. M. Saptarini, and B. M. Bactiar, “Evaluasi Emisi Karbon Dioksida (CO₂) Proses Granulasi pada Produksi Tablet Y di Pabrik X,” *Majalah Farmasetika*, vol. 11, no. 1, 2026.
- [9] A. Suryandaru, A. T. Hidayat, and M. I. Apriadi, “Penanganan emisi gas karbon di Kota Cilegon menggunakan pendekatan sistem dinamis,” *JENIUS: Jurnal Terapan Teknik Industri*, vol. 5, no. 1, pp. 150–155, 2024.
- [10] F. Afnilia and C. D. Astuti, “Pengaruh volume emisi karbon, pengungkapan emisi karbon, dan tata kelola perusahaan terhadap nilai perusahaan,” *Jurnal Ekonomi Trisakti*, vol. 3, no. 2, pp. 3795–3804, 2023.
- [11] A. R. Andhika and E. Bayunanda, “STUDI LITERATURE REVIEW PEMANFAATAN MACHINE LEARNING DALAM MULTISEKTOR,” *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 03 Agustus, pp. 6865–6880, 2025.
- [12] A. M. Bekhit, M. Sobh, M. Abdel Zaher, T. Abdel Fattah, and A. I. Diab, “Predicting terrestrial heat flow in Egypt using random forest regression: a machine learning approach,” *Geothermal Energy*, vol. 13, no. 1, p. 18, 2025.
- [13] S. V. Hutagalung, Y. Yennimar, E. R. Rumapea, M. J. G. Hia, T. Sembiring, and D. R. Manday, “Comparison of support vector regression and random forest regression algorithms on gold price predictions,” *Jurnal Sistem Informasi dan Ilmu Komputer*, vol. 7, no. 1, pp. 255–262, 2023.
- [14] A. Simangunsong, R. M. Simanjorang, F. Amalia, and P. Khairunnisa, “Implementasi Sistem Rekomendasi dengan Collaborative Filtering dalam Pemilihan Produk Skincare,” *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika dan Komputer)*, vol. 24, no. 1, pp. 56–63, 2025.
- [15] I. I. Permana and M. Miftahudin, “Penerapan Metode Content-Based Filtering untuk Rekomendasi pada Resep Obat Berdasarkan Diagnosa Pasien,” *Jurnal SAINTEKOM (Sains dan Teknologi Komputasi)*, vol. 1, no. 1, pp. 52–62, 2025.
- [16] O. Oktavianda, L. Efrizoni, E. Fatdha, and H. Asnal, “SISTEM REKOMENDASI VIDEO GAME BERBASIS USIA SEBAGAI ALAT PENGAWASAN ORANG TUA DI PLATFORM STEAM MENGGUNAKAN CONTENT-BASED FILTERING,” *Jurnal Sistem Informasi dan Informatika*, vol. 8, no. 2, pp. 371–383, 2025.
- [17] M. F. Ramadhani, “Penerapan metode content-based filtering dalam rekomendasi software developer,” in *Prosiding Seminar Nasional Penelitian dan Pengabdian Kepada Masyarakat LPPM Universitas’ Aisyiyah Yogyakarta*, 2026.
- [18] M. Chan, “Dataset Kaggle: Canadian Fuel Consumption Ratings.” Accessed: Feb. 21, 2026. [Online]. Available: <https://www.kaggle.com/datasets/mcpenguin/canadian-fuel-consumption-ratings>
- [19] S. A. N. Cahyo and M. Y. T. Sulistyono, “Comparison of Multiple Linear Regression and Random Forest Methods for Predicting National Rice Production in Indonesia,” *Journal of Applied Informatics and Computing*, vol. 9, no. 6, pp. 3479–3489, 2025.
- [20] Y. Muflihah and M. W. Wardhani, “Implementasi Hybrid Filtering dalam Sistem Rekomendasi Berbasis Web untuk Persewaan Tenda Pesta,” *Jurnal Rekayasa Sistem Informasi dan Teknologi*, vol. 1, no. 4, pp. 361–372, 2024