

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Perubahan iklim global merupakan tantangan utama yang dihadapi dunia saat ini, dengan emisi karbon dioksida ( $\text{CO}_2$ ) sebagai kontributor signifikan terhadap pemanasan global. Sektor transportasi merupakan penyumbang penting emisi  $\text{CO}_2$ , khususnya di negara maju seperti Kanada. Berdasarkan laporan resmi *Environment and Climate Change Canada* tahun 2023, sektor transportasi menyumbang sekitar 23% dari total emisi gas rumah kaca nasional, yaitu sebesar 157 megaton  $\text{CO}_2$  ekuivalen [1]. Kendaraan berbahan bakar fosil di Kanada menghasilkan rata-rata emisi  $\text{CO}_2$  mencapai 206 g  $\text{CO}_2/\text{km}$ , termasuk yang tertinggi di antara negara-negara maju [2]. Pemerintah Kanada telah menetapkan target untuk mengurangi emisi gas rumah kaca sebesar 40-45% di bawah level 2005 pada tahun 2030 dan mencapai net-zero pada 2050 [3].

Untuk mewujudkan target tersebut, diperlukan pemahaman menyeluruh tentang faktor-faktor yang mempengaruhi emisi  $\text{CO}_2$  kendaraan dan kemampuan untuk memprediksinya dengan akurat. Kemampuan memantau dan memprediksi emisi  $\text{CO}_2$  kendaraan memungkinkan peneliti, produsen kendaraan, dan pembuat kebijakan memahami sumber serta pola emisi  $\text{CO}_2$ , sehingga dapat merumuskan strategi pengurangan emisi yang efektif [4].

*Machine learning* merupakan pendekatan modern yang dapat digunakan untuk memprediksi dan menganalisis emisi  $\text{CO}_2$  kendaraan berdasarkan karakteristiknya. Teknologi ini mampu memproses data dalam jumlah besar dengan efisien dan mengidentifikasi pola kompleks. Algoritma *machine learning* memiliki keunggulan dalam menangkap hubungan kompleks [5], yang memungkinkan pemahaman lebih jelas tentang faktor-faktor yang mempengaruhi emisi  $\text{CO}_2$  kendaraan. Model *machine learning* juga dapat diterapkan pada berbagai tipe data dan beradaptasi dengan karakteristik *dataset* yang berbeda [6]. Namun, model ini juga memiliki keterbatasan seperti potensi *overfitting* dan keterbatasan interpretasi untuk beberapa algoritma kompleks [7].

Dalam penelitian ini, tujuh algoritma *machine learning* akan diimplementasikan *Linear Regression*, *Ridge Regression*, *Random Forest*,

*Gradient Boosting*, *Support Vector Regression (SVR)*, *K-Nearest Neighbor (KNN)*, dan *Neural Network*. Algoritma-algoritma ini akan digunakan untuk memprediksi emisi CO<sub>2</sub> kendaraan menggunakan *dataset CO<sub>2</sub> Emissions by Vehicles* yang tersedia di Kaggle dan dikompilasi dari *Open Data* Pemerintah Kanada. Meskipun algoritma ini memiliki potensi besar, beberapa tantangan perlu diatasi, seperti *overfitting* dan multikolinearitas antara fitur-fitur seperti konsumsi bahan bakar dan ukuran mesin yang dapat mempengaruhi akurasi prediksi [8].

Untuk mengatasi tantangan tersebut, beberapa strategi akan diterapkan: (1) teknik *preprocessing data* seperti *scaling* dan *encoding* kategorikal untuk menstandarisasi input, (2) *hyperparameter tuning* pada model terbaik untuk mengoptimalkan parameter, (3) evaluasi model menggunakan metrik seperti *R<sup>2</sup>*, *MAE*, *RMSE*, dan *MAPE*, serta (4) *feature engineering* untuk menciptakan variabel baru yang lebih informatif.

Penelitian ini bertujuan untuk mengembangkan model prediksi emisi CO<sub>2</sub> kendaraan yang akurat berdasarkan karakteristiknya, mengidentifikasi faktor-faktor yang paling berpengaruh, serta mengimplementasikan model terbaik dalam aplikasi web menggunakan *Streamlit* sebagai alat bantu bagi konsumen, produsen kendaraan, dan pembuat kebijakan dalam upaya mengurangi emisi karbon dari sektor transportasi di Kanada.

## 1.2 Perumusan masalah

Berdasarkan latar belakang yang telah diuraikan, penelitian ini berupaya menjawab beberapa pertanyaan kunci sebagai berikut:

1. Bagaimana performa berbagai algoritma *machine learning* yaitu *Linear Regression*, *Ridge Regression*, *Random Forest*, *Gradient Boosting*, *Support Vector Regression*, *K-Nearest Neighbors*, dan *Neural Network* dalam memprediksi emisi CO<sub>2</sub> kendaraan di Kanada, dan algoritma mana yang memberikan hasil prediksi paling akurat?
2. Faktor-faktor atau fitur apa saja yang memiliki pengaruh paling signifikan terhadap emisi CO<sub>2</sub> kendaraan berdasarkan analisis model *machine learning*, dan bagaimana pemahaman tentang faktor-faktor ini dapat dimanfaatkan untuk pengembangan kebijakan transportasi yang lebih efektif di Kanada?

3. Sejauh mana proses *hyperparameter tuning* dapat meningkatkan akurasi model prediksi emisi CO<sub>2</sub>, dan apa parameter optimal yang dapat diimplementasikan untuk model terbaik?
4. Bagaimana model prediksi emisi CO<sub>2</sub> berbasis *machine learning* dapat diimplementasikan dalam bentuk aplikasi yang dapat diakses dan dimanfaatkan oleh pemangku kepentingan di sektor transportasi Kanada?

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah di atas, tujuan penelitian ini adalah:

1. Menganalisis dan membandingkan performa tujuh algoritma *machine learning* dalam memprediksi emisi CO<sub>2</sub> kendaraan menggunakan dataset *CO<sub>2</sub> Emissions by Vehicles*, serta mengidentifikasi algoritma dengan tingkat akurasi prediksi tertinggi.
2. Mengidentifikasi faktor-faktor kendaraan yang memiliki pengaruh paling signifikan terhadap emisi CO<sub>2</sub> melalui analisis *feature importance*, yang dapat memberikan wawasan berharga untuk pengembangan kebijakan transportasi yang lebih efektif di Kanada.
3. Meningkatkan akurasi model terbaik melalui proses *hyperparameter tuning*, serta menentukan kombinasi parameter optimal yang menghasilkan prediksi paling akurat.
4. Mengimplementasikan model *machine learning* terbaik dalam bentuk aplikasi web interaktif menggunakan Streamlit, yang dapat diakses dan dimanfaatkan oleh berbagai pemangku kepentingan untuk memprediksi emisi CO<sub>2</sub> kendaraan.

### 1.4 Batasan Masalah

Batasan yang digunakan dalam penelitian ini agar pembahasan yang dilakukan tidak meluas yaitu:

1. Penelitian ini akan menggunakan dataset emisi CO<sub>2</sub> *Emissions by Vehicles* dari Kaggle yang dikompilasi dari *Open Data* Pemerintah Kanada,
2. Penelitian ini memanfaatkan model regresi untuk memprediksi emisi CO<sub>2</sub> berdasarkan karakteristik kendaraan, tanpa mempertimbangkan faktor eksternal seperti kondisi jalan atau cuaca.

3. Penelitian ini membandingkan 7 algoritma *machine learning* *Linear Regression*, *Ridge Regression*, *Random Forest*, *Gradient Boosting*, *SVR*, *KNN*, dan *Neural Network* dan memilih model terbaik berdasarkan metrik evaluasi.
4. Model terpilih akan dioptimalkan dengan *hyperparameter tuning* untuk meningkatkan performa prediksi.
5. Deployment model menggunakan *Streamlit* untuk membuat aplikasi prediksi emisi CO<sub>2</sub> kendaraan yang interaktif.
6. Penelitian tidak menggunakan data emisi kendaraan dari Indonesia karena:
  - a. Perbedaan Regulasi Emisi:

Negara maju seperti Singapura, Uni Eropa, Jepang, dan Kanada menerapkan regulasi ketat terkait standar emisi dan usia kendaraan. Sebagai contoh, Singapura membatasi usia kendaraan maksimal 10 tahun melalui sistem *Certificate of Entitlement (COE)* dan mewajibkan standar emisi Euro VI sejak 2018 [9] [10]. Di Indonesia, penerapan standar emisi Euro 4 masih belum menyeluruh, dan belum ada pembatasan usia kendaraan secara nasional [11].

- b. Ketersediaan & Kualitas Data:

Data emisi di negara-negara seperti Kanada dan Amerika Serikat dicatat secara sistematis dan terbuka oleh lembaga pemerintah seperti *Environment and Climate Change Canada* dan *US EPA*, dengan parameter lengkap seperti konsumsi BBM, ukuran mesin, dan emisi CO<sub>2</sub> dalam satuan *g/km* [1]. Sementara itu, di Indonesia, data uji emisi bersifat terbatas, insidental, tidak terpusat, dan belum tersedia secara publik dalam format yang terstandarisasi [12].

- c. Konsistensi Pemodelan:

Dataset dari Kanada menggunakan satuan standar internasional seperti *g/km* dan *L/100 km*, yang kompatibel dengan pemodelan regresi. Sementara di Indonesia, data emisi sering menggunakan format seperti *%CO* atau *ppm HC*, yang memerlukan *preprocessing* tambahan dan konversi unit agar bisa digunakan untuk pemodelan yang sama [1].

### 1.5 Manfaat Penelitian

Berdasarkan latar belakang yang telah disampaikan maka dapat manfaat dari penelitian ini adalah:

1. Model prediksi ini membantu pemerintah dan masyarakat mengambil langkah tepat dalam mengurangi emisi karbon dan dampaknya terhadap iklim.
2. Data prediksi emisi mendukung perencanaan kebijakan lingkungan yang lebih akurat dan strategis.
3. Proyek ini membuktikan bahwa teknologi seperti *machine learning* dan visualisasi data bisa dimanfaatkan untuk kepentingan lingkungan dan mendorong inovasi hijau.

