

BAB 2 LANDASAN TEORI

Dalam bab ini akan dijelaskan tentang dasar teori yang digunakan sebagai acuan dalam penelitian ini. Bab ini akan menjelaskan tentang emisi karbon dioksida, mobil dan faktor yang mempengaruhi emisi karbon dioksida, *machine learning*, algoritma *Linear Regression*, *preprocessing*, *Root Mean Squared Error*, koefisien determinasi (R^2), dan validasi silang *10-fold*.

2.1 Emisi Karbon Dioksida (CO_2)

Sejak dimulainya revolusi industri, emisi karbon dioksida (CO_2) menjadi permasalahan utama dari dampak pemanasan global dan perubahan iklim [14]. Penggunaan energi, pertumbuhan ekonomi memiliki dampak positif yang signifikan terhadap emisi karbon dioksida (CO_2) yang dimaksudkan adalah menyumbang dampak pemanasan global dan perubahan iklim [11]. Di sektor transportasi, hasil proyeksi dari *International Transport Forum* bahwa emisi karbon dioksida (CO_2) dari sektor transportasi akan meningkat sebesar 60% di tahun 2050 [15].

Sejak tahun 2015, Pemerintahan Kanada resmi menandatangani Perjanjian Paris (*Paris Agreement*) dengan acuan data pada tahun 2005 sebagai target untuk menurunkan emisi karbon dioksida (CO_2). Di tahun 2021, pemerintahan Kanada berkomitmen untuk mengurangi emisi karbon dioksida (CO_2) dengan target sebesar 40-45% dibawah jumlah emisi karbon dioksida (CO_2) di tahun 2005 untuk tahun 2030. Di dalam program Rencana Pengurangan Emisi 2030 (*2030 Emissions Reductions Plan*) yang dirilis pada bulan Maret tahun 2022, pemerintahan Kanada mengumumkan tujuan sementara untuk mengurangi emisi karbon dioksida (CO_2) sebesar 20% dibawah jumlah emisi karbon dioksida (CO_2) pada tahun 2005 pada tahun 2026 [3]. Rencana pemerintahan Kanada dalam mengurangi emisi karbon dioksida (CO_2) mengadopsi dari Protokol Kyoto (*Kyoto Protocol*) yang merupakan perjanjian lingkungan hidup yang berafiliasi dengan *United Nations Framework Convention on Climate Change (UNFCCC)* yang dibuat pada tahun 1997 [16].

2.2 Kendaraan Mobil dan Faktor

Dalam konteks emisi karbon dioksida (CO_2) pada kendaraan mobil, faktor yang mempengaruhi emisi karbon dioksida (CO_2) pada sektor transportasi adalah adanya peningkatan kebutuhan bahan bakar yang terus meningkat, jenis kendaraan yang digunakan, dan konsumsi bahan bakar [4].

Perbedaan bahan bakar yang digunakan pada kendaraan bermotor khususnya mobil merupakan aspek yang krusial terhadap dampak emisi karbon dioksida (CO_2) pada kendaraan bermotor, khususnya mobil. Perbedaan bahan bakar yang digunakan bisa dilihat pada hasil kecepatan dan akselerasi pada kendaraan bermotor, khususnya mobil [17].

2.3 Machine Learning

Machine Learning adalah sebuah paradigma atau kerangka teori komputasi yang memiliki kapasitas untuk menyelesaikan permasalahan yang diberikan yang telah dibangun berdasarkan contoh sebelumnya. Ide dasar dari penemuan *machine learning* adalah penalaran berdasarkan kasus yang dimana proses penalaran masalah dengan merujuk kepada masalah sebelumnya. Contoh sebelumnya yang dimaksud adalah contoh latihan, dan proses untuk menyelesaikan masalah riil setelah mempelajari contoh latihan disebut juga dengan generalisasi [18].

Terdapat tiga model dalam pengaplikasian *machine learning*, model klasifikasi adalah proses mengklasifikasikan objek menjadi satu atau beberapa diantara kategori yang telah ditentukan sebelumnya, model regresi adalah proses yang mengestimasi nilai lanjutan atau nilai yang berdasarkan faktor input saat ini, dan model klusterisasi yang membagi seluruh kelompok menjadi sub kelompok berdasarkan persamaan dari setiap objek [18]. Untuk proses regresi akan menggunakan dua tipe variabel, variabel independen digunakan untuk sebagai data input dengan tujuan untuk memprediksi nilai yang akan dibandingkan dengan nilai dependen atau nilai sebenarnya [9].

2.4 Algoritma Linear Regression

Algoritma *linear regression* adalah metodologi yang sederhana dan populer dalam mengukur hubungan diantara prediktor yang berkelanjutan dan variabel yang menjadi respon [19]. Algoritma *linear regression* adalah metode statistikal yang

berguna untuk memprediksi hasil dari variabel yang bersifat dependen berdasarkan set variabel penjelasan yang memiliki persamaan adalah sebagai berikut :

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon \quad (2.1)$$

Variabel y_i merepresentasikan hasil prediksi yang disebut juga variabel terikat, variabel β_0 merupakan titik potong y ketika nilai y untuk variabel $x_1, x_2,$ dan x_p sama dengan nol, β_1 dan β_2 adalah nilai koefisien regresi yang merupakan parameter dari persamaan *Linear Regression*, mewakili perubahan nilai y secara relatif terhadap variabel independen x_1 dan x_2 . β_p adalah nilai kemiringan dari setiap variabel independen dan ε adalah nilai kesalahan atau disebut juga nilai residual:

Tabel 2.1. Deskripsi Variabel

Nama	Deskripsi
Ukuran Mesin	Ukuran dari mesin mobil dengan satuan liter (L)
Silinder	Jumlah silinder yang ada pada mobil
Konsumsi Bahan Bakar	Jumlah konsumsi bahan bakar berdasarkan gabungan dari hasil konsumsi bahan bakar di jalan kota dan jalan tol, dengan satuan liter per 100 kilometer (L/100km)
Emisi Karbon Dioksida (CO2)	Hasil emisi karbon yang dihasilkan dari knalpot dari gabungan jalan kota dan jalan tol, dengan satuan gram per kilometer (g/km)

2.5 Preprocessing

Preprocessing adalah sebuah langkah awal yang bertujuan untuk mengolah data mentah yang memiliki data yang tidak terstruktur menjadi sebuah data yang mudah dipahami, *preprocessing* juga bertujuan untuk memastikan bahwa data tersebut bisa digunakan di proses validasi dan reliabilitas [20]. *Preprocessing* memiliki 4 tahap dalam implementasi adalah sebagai berikut :

1. *Data Cleaning* : *Data Cleaning* merupakan proses identifikasi dan perbaikan masalah dalam data. Misalnya, kesalahan input, nilai hilang, dan duplikasi. Tujuan utama dari tahap ini adalah untuk menciptakan data set yang konsisten dan akurat.
2. *Data Integration* : *Data Integration* adalah tahapan dari *preprocessing* yang melibatkan penggabungan data dari berbagai sumber menjadi satu dataset

yang kohesif. Tahapan ini penting dilakukan ketika informasi dikumpulkan dari berbagai database atau saat bekerja dengan dataset besar yang tersebar.

3. *Data Transformation* : *Data Transformation* adalah tahap proses mengubah data ke dalam format atau struktur yang lebih sesuai untuk analisis. Tahap ini dapat mencakup normalisasi yang merupakan menyesuaikan skala nilai data ke rentang tertentu, seperti 0-1, untuk memudahkan perbandingan dan standarisasi yang mengubah data sehingga memiliki rata-rata 0 dan standar deviasi 1. Ini membantu beberapa algoritma pembelajaran mesin bekerja lebih efektif.
4. *Data Reduction* : *Data Reduction* merupakan tahapan pengurangan data untuk mengurangi volume data tanpa menghilangkan informasi penting. Hal ini bisa mempercepat pemrosesan dan analisis. Beberapa strategi dalam *Data Reduction* mencakup pengurangan dimensi seperti Analisis Komponen Utama (PCA) digunakan untuk mengurangi jumlah variabel dalam dataset dan *sampling data* untuk memilih subset dari data yang mewakili keseluruhan dataset. Tujuannya adalah untuk mengurangi ukuran data sambil tetap mempertahankan karakteristik statistik penting.

2.6 Root Mean Square Error (RMSE)

Root Mean Square Error atau disingkat RMSE adalah standar metrik yang digunakan untuk evaluasi suatu model, RMSE merupakan akar kuadrat dari MSE (*Mean Square Error*) yang menghasilkan metrik dengan unit yang sama, atau dengan kata lain memudahkan dalam merepresentasikan nilai kesalahan untuk distribusi kesalahan secara normal [21]. Rumus dari RMSE adalah sebagai berikut :

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.2)$$

2.7 Koefisien Determinasi (R^2)

Koefisien determinasi atau dilambangkan sebagai R^2 adalah metrik evaluasi model yang bertujuan untuk menghitung seberapa besar variabel dependen yang digunakan ditentukan oleh variabel independen. Koefisien determinasi memiliki

batas bawah 0 yang menandakan bahwa model yang dibangun tidak mampu memprediksi, dan memiliki batas atas 1 yang menandakan bahwa model yang dibuat mampu memprediksi nilai dengan sempurna [22]. Rumus dari koefisien determinasi adalah sebagai berikut :

$$R^2 = 1 - \frac{\sum_{i=1}^m (X_i - Y_i)^2}{\sum_{i=1}^m (\bar{Y} - Y_i)^2} \quad (2.3)$$

2.8 Validasi Silang 10-Fold

Validasi silang atau *cross-validation* adalah teknik yang digunakan dalam evaluasi model prediktif untuk menilai kinerja dan kemampuan generalisasi dalam suatu model dengan tujuan meningkatkan efektivitas dari data latihan untuk model. Cara kerja dari validasi silang adalah data input akan dibagi menjadi 10 total subkelompok. Subkelompok yang tersisa akan digunakan sebagai data uji, dan sembilan subkelompok digunakan sebagai data latih untuk setiap putaran pelatihan model. Proses ini dilakukan sebanyak 10 kali sehingga setiap *fold* menjadi data satu kali [23].

