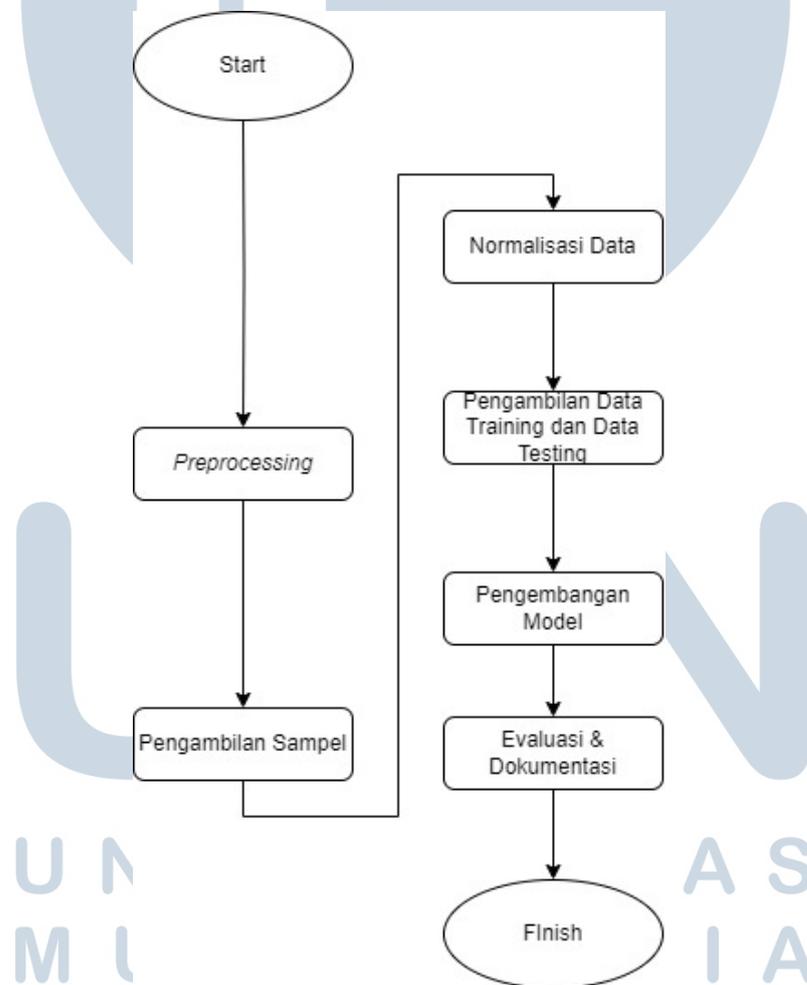


BAB 3 METODOLOGI PENELITIAN

Untuk mencapai tujuan dari yang telah disebutkan, perlunya metodologi sebagai langkah dalam penelitian. Berikut merupakan langkah yang diambil dalam menjalankan penelitian. Literatur yang diambil berfokus pada cara kerja algoritma *Linear Regression*, pengambilan dataset, preprocessing data dan hal lain yang mendukung dalam penelitian. Proses pada pembuatan model terlampir pada Gambar 3.1 dalam bentuk *flowchart*.



Gambar 3.1. Flowchart proses pembuatan model

3.1 Studi Literatur

Studi literatur dilakukan dengan mencari dan membaca hasil penelitian serta paper terdahulu, serta mencari teori yang dibutuhkan dalam penelitian. Pencarian hasil penelitian berfokus pada *preprocessing*, pengambilan sampel, normalisasi data, dan pengembangan model algoritma *Linear Regression*.

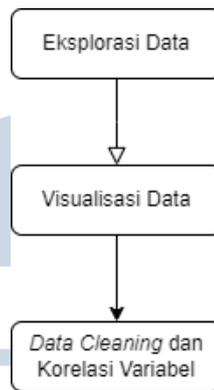
3.2 Pencarian Dataset

Data yang digunakan pada penelitian ini diambil dari Kaggle, data yang digunakan berisikan detail hasil emisi karbon dioksida (CO_2) dari beberapa kendaraan yang berbeda, data tersebut diambil dari data terbuka resmi situs pemerintahan Kanada.

3.3 *Preprocessing*

Tahap *preprocessing* dilakukan dengan tujuan untuk menelusuri, mengidentifikasi, dan memilah data atau variabel yang nantinya akan digunakan dalam pemodelan nanti, tahap-tahap pada *preprocessing* pada penelitian pada Gambar 3.2 adalah sebagai berikut:

- Eksplorasi Data : Eksplorasi data dilakukan dengan tujuan untuk melakukan identifikasi variabel, tipe data, pada data di dalam variabel tersebut.
- Visualisasi Data : Visualisasi data dilakukan dengan maksud untuk memberikan penjelasan secara mendetail terkait variabel yang ada pada dataset.
- *Data Cleaning* dan Korelasi Variabel: Teknik ini bertujuan untuk menghapus beberapa data yang tidak mempunyai nilai pembanding atau memiliki data tunggal pada variabel yang ada, menghilangkan variabel yang tidak relevan dalam pemodelan nanti, menghapus data *outliers* yang mempengaruhi nilai akurasi dalam pemodelan tersebut, dan mengetahui nilai korelasi antar variabel pada dataset tersebut.



Gambar 3.2. Tahap *Preprocessing*

3.4 Pengambilan Sampel

Pengambilan sampel yang dilakukan pada penelitian ini akan menggunakan *random sampling* dengan formula Cochran yang merupakan rumus untuk menghitung ukuran sampel ideal sesuai dengan tingkat presisi yang diinginkan [24]. Rumus Cochran untuk melakukan pengambilan sampel adalah sebagai berikut:

$$n_0 = \frac{z^2 \cdot p \cdot (1 - p)}{e^2} \quad (3.1)$$

p : ukuran populasi

e : *margin error*

z : nilai z yang ditentukan dengan *confidence level*

3.5 Normalisasi Data

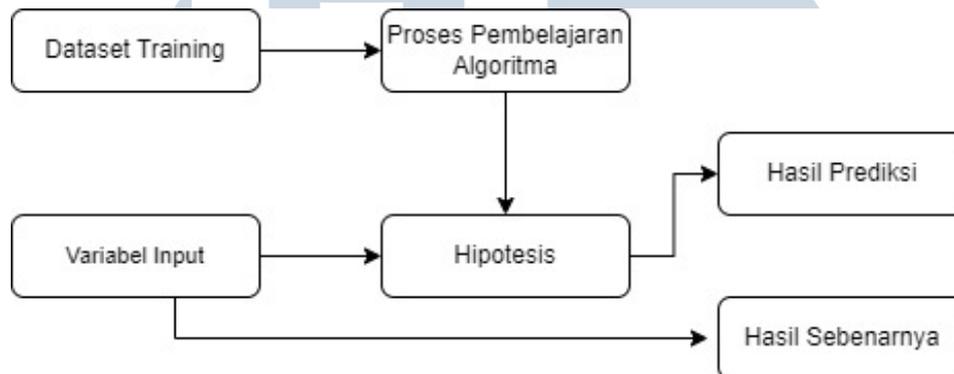
Tahap normalisasi ini dilakukan dengan tujuan untuk meminimalisir kesalahan atau inkonsistensi data yang berakibat pada redundansi dan standarisasi data [25]. Normalisasi data yang dilakukan pada umumnya adalah mengubah tipe data pada variabel, menghapus data duplikat, dan lain-lainnya.

3.6 *Splitting Data Training dan Data Testing*

Pada penelitian ini, dataset yang ada akan dibagi menjadi dua bagian, yaitu data *training* yang digunakan untuk pembelajaran bagi model yang telah dibangun, dan data *testing* digunakan untuk mengetes model yang telah dibangun dari hasil pembelajaran sebelumnya. Pada umumnya, pembagian data *training* dan data

testing beragam, yaitu 80:20, 70:30, 60:40, dan bahkan 50:50. Namun aturan praktis yang biasa digunakan dalam *data splitting* adalah dengan rasio 80:20 [26].

3.7 Pengembangan Model Algoritma *Linear Regression*



Gambar 3.3. Flowchart proses pembuatan model

Berdasarkan proses pada *flowchart* di Gambar 3.3, setelah melakukan semua tahapan sebelumnya, pengembangan model algoritma *Linear Regression* dilakukan dengan memasukkan data *training* yang digunakan sebagai pembelajaran model tersebut, dan hasilnya merupakan hipotesis atau hasil sementara dari prediksi dan akhirnya akan mengeluarkan hasil dari prediksi dari model yang sudah dibangun. Lalu melakukan perbandingan antara nilai dari hasil prediksi dan nilai yang sebenarnya.

3.8 Evaluasi dan Dokumentasi

Setelah melakukan pemodelan dan melakukan prediksi menggunakan algoritma *linear regression*, evaluasi dilakukan dengan mengecek nilai *Root Mean Squared Error* (RMSE) untuk data *training* dan data *testing* dengan tujuan untuk mengetahui nilai rata-rata kesalahan dari hasil latih dan juga hasil uji, koefisien determinasi (R^2) untuk data *training* yang bertujuan untuk penilaian hasil pemodelan yang dilakukan dengan skala 0 sampai 1, koefisien determinasi (R^2) dari validasi silang *10-fold* untuk evaluasi kembali dengan 10 subset data yang berbeda, dan rata-rata perbedaan antara nilai prediksi dan nilai aktual. Dari evaluasi yang telah dilakukan akan didokumentasikan dalam bentuk laporan penelitian.