

BAB I

PENDAHULUAN

1.1. Latar Belakang

Pertumbuhan industri yang signifikan pada era globalisasi saat ini menyebabkan daya saing antar perusahaan di Indonesia menjadi lebih [1]. Data mining merupakan proses pencarian data yang cukup besar dari sekumpulan data yang besar, , sedangkan regresi linier ialah suatu cara statistik yang berfungsi untuk membuat model suatu hubungan linear antar variabel dependen dengan variable independen baik itu satu maupun banyak, sedangkan k-nearest neighbor merupakan algoritma yang berfungsi untuk melakukan pengelompokan dan regresi [2]. Data mining adalah tentang metode untuk penggalian pengetahuan atau menemukan pola dari suatu data [3]. Seiring dengan meningkatnya jumlah kendaraan di dunia, konsumsi bahan bakar ialah salah satu komponen fundamental yang butuh diperhatikan dalam pengelolaan industri otomotif dan transportasi. Kenaikan harga bahan bakar, kebutuhan untuk mengurangi emisi karbon, serta tantangan lingkungan memaksa produsen kendaraan dan perusahaan transportasi untuk mencari solusi yang lebih efisien dalam penggunaan bahan bakar. Prediksi konsumsi bahan bakar yang lebih akurat dapat membantu dalam merancang kebijakan yang lebih efektif dan mengoptimalkan operasi kendaraan untuk mengurangi biaya dan dampak negatif terhadap lingkungan. Namun, dalam praktiknya, prediksi konsumsi bahan bakar sering kali hanya didasarkan pada data historis atau perkiraan kasar, yang tidak selalu mencerminkan kondisi nyata dan variabel dinamis yang mempengaruhi konsumsi. Oleh karena itu, dibutuhkan pendekatan yang lebih canggih untuk memprediksi konsumsi bahan bakar dengan menggunakan metode analisis data yang lebih akurat, seperti data mining. Prediksi merupakan suatu ilmu untuk menduga kejadian yang akan terjadi menggunakan sebuah kumpulan data yang diolah sedemikian rupa dan dapat memberikan proyeksi jawaban yang akan datang dengan menggunakan model yang sistematis. [4].

Metode yang digunakan dalam penelitian ini yaitu metode regresi linier sederhana [5]. Metode Regularized Linear Regression dipilih dalam penelitian ini karena mampu mengatasi masalah multikolinearitas dalam data dan mengurangi overfitting, yang merupakan tantangan umum dalam pemodelan dengan data yang kompleks. Regresi adalah salah satu cara untuk membuat suatu model yang dapat digunakan untuk melakukan prediksi suatu value dari masukan yang dibeikan [6]. Penerapan data model algoritma regresi linier sederhana dapat terlaksana dengan baik dan dapat memberikan pengetahuan baru [7]. Variabel yang digunakan dalam penelitian ini seperti Vehicle Class, Engine Size, Cylinders, Transmissions, dan CO emissions, Kami berharap ini akan memberi kita gambaran lebih lengkap tentang faktor-faktor yang memengaruhi konsumsi bahan bakar kendaraan. Dengan menggunakan teknik data mining dan regresi linier berganda, penelitian ini bertujuan untuk mengembangkan model yang lebih akurat dalam memprediksi konsumsi bahan bakar, yang pada gilirannya dapat digunakan oleh produsen kendaraan, perusahaan transportasi, dan pembuat kebijakan untuk meningkatkan efisiensi bahan bakar, mengurangi biaya operasional, dan mendukung upaya pengurangan emisi gas rumah kaca. Analisis Regresi sederhana memiliki maksud untuk mengetahui apakah suatu variabel dapat mempengaruhi variabel lainya atau tidak.[8].

Melalui penelitian ini, diharapkan program yang dibuat dapat menemukan hasil yang kemudian digunakan sebagai prediksi pembeli [9].

1.2. Rumusan Masalah

Rumusan masalah yang berdasarkan dari latar belakang yang telah diuraikan dalam penelitian ini yaitu :

1. Bagaimana cara memprediksi konsumsi bahan bakar kendaraan menggunakan metode regularized linear regression ?
2. Berapa keakuratan dari program ini untuk menciptakan kendaraan yang dapat mengurangi upaya pengurangan emisi gas rumah kaca?

1.3. Tujuan dan Manfaat

Adapun manfaat dan tujuan dari melakukan penelitian ini yaitu :

1. Menggunakan metode regularized linear regression juga data *Vehicle Class*, *Engine Size*, *Cylinders*, *Transmissions*, dan *CO emissions* dalam penelitian ini.
2. Keakuratan program ini adalah 79% didapatkan pada nilai R-Squares pada program yang diteliti. Program dengan keakuratan yang cukup tinggi ini mendapatkan hasil mobil dengan bahan bakar paling rendah dan juga rendah emisi karbon. Untuk mendapatkan keakuratan yang tinggi, tentunya didukung karena adanya data kelas kendaraan, Ukuran mesin, Silinder, Transmisi dan Emisi karbon yang dihasilkan. Semakin banyak data pendukung yang bisa didapatkan, maka semakin tinggi akurasi yang didapatkan dari Analisa ini.

1.4. Batasan Masalah

1. Data yang digunakan menggunakan beberapa brand mobil, diantaranya Suzuki, Honda, Audi, Nissan, Ford, GMC, Mazda, Toyota serta brand lainnya.
2. Data yang digunakan melibatkan 36 Brand dengan total 639 Data.
3. Analisis dilakukan dengan menggunakan metode Regularized Linear Regression.

1.5. Keterbaruan Penelitian

Berikut ini terdahulu yang berkaitan dengan implementasi metode Regularized Linear Regression, antara lain:

1. Judul penelitian “Prediksi Pengisian BBM HSD Dengan Metode Multiple Linear Regression” yang diteliti oleh Latifah dan Muhammad Deni pada tahun 2022 bertujuan untuk membuat sistem pencatatan dan prediksi BBM High Speed Diesel secara otomatis. Penelitian tersebut menghasilkan nilai MSE sebesar 7.874 dan MAPE sebesar 5%.

Perbedaan dengan penelitian ini : Penelitian kami menggunakan metode Regularized Linear Regression yang mampu mengatasi masalah multikolinearitas, menghasilkan nilai MSE lebih rendah yaitu 2.01, serta memberikan keakuratan prediksi yang lebih tinggi. Selain itu, penelitian kami juga memperhitungkan variabel tambahan seperti Vehicle Class, Engine Size, Cylinders, Transmissions, dan CO emissions untuk meningkatkan ketepatan model

2. Penelitian yang berjudul “Prediksi Penyakit Diabetes Untuk Pencegahan Dini Dengan Metode Regresi Linear” yang dilakukan oleh Amat Basri, Umi Atmaja, Dewi Marini Umi Atmaja Arif Rahman, dan Niko Suwaryo Amat Basri pada tahun 2023 menggunakan aplikasi RapidMiner Studio dan menghasilkan nilai RMSE 0.00 +/- 0.000.

Perbedaan dengan penelitian ini: Penelitian tersebut tidak menampilkan nilai akurasi secara eksplisit, sedangkan dalam penelitian kami ditampilkan secara lengkap nilai R squared, MSE, dan tingkat akurasi sebesar 79%. Fokus penelitian kami juga berbeda, yakni pada konsumsi bahan bakar kendaraan dan pengurangan emisi, bukan pada bidang kesehatan.

3. Penelitian yang berjudul “Metode Regresi Linier Berganda Untuk Prediksi Pemakaian BBM PT. Kalonica Bara Kusuma” yang dilakukan oleh Wawan Joko Pranoto dan Augie Sugiarto Nunka pada tahun 2024 yang bertujuan untuk mengoptimalkan penggunaan BBM di sektor pertambangan. Namun, tidak dicantumkan nilai akurasi dari model prediksinya.

Perbedaan dengan penelitian ini: Penelitian kami menyajikan hasil akurasi model secara kuantitatif (nilai R-squared), serta menggunakan pendekatan regularisasi dalam regresi linier yang tidak digunakan pada penelitian tersebut, sehingga meningkatkan generalisasi model pada data yang kompleks

4. Penelitian yang berjudul tentang “Penerapan Metode Regresi Linear Dalam Peramalan Penjualan Kue Di Toko Karya Bahari Samarinda Berbasis Website” yang dilakukan oleh Yosep Agus Pranoto, Franciscus Xaverius Ariwibisono, dan Welly Sebastian pada tahun 2024, hanya menampilkan nilai MAPE sebesar 2%, tanpa menampilkan metrik lain seperti MAE, MSE, RMSE, dan R-squared.

Perbedaan dengan penelitian ini: Penelitian kami menampilkan semua metrik evaluasi model secara lengkap, termasuk MSE dan R-squared, yang dapat memberikan gambaran lebih menyeluruh mengenai performa model. Selain itu, konteks aplikasinya juga berbeda, yakni pada efisiensi bahan bakar kendaraan.

5. Penelitian yang berjudul “Penerapan Metode Regresi Linear Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet” yang dilakukan oleh Ela Srihartini, Zaenal Muttaqin, dan Harsiti, menyatakan tingkat akurasi sebesar 98,505% tanpa menunjukkan nilai R-squared, serta hanya menggunakan metode penghitungan berdasarkan pengurangan dari 100% terhadap nilai MAP.

Perbedaan dengan penelitian ini : Penelitian kami menyajikan pendekatan evaluasi model yang lebih komprehensif, termasuk dengan menunjukkan nilai R-squared dan MSE. Selain itu, penggunaan metode Regularized Linear Regression membuat model kami lebih tahan terhadap overfitting, dibanding metode regresi sederhana yang digunakan oleh penelitian tersebut.