

**KINERJA *SUPPORT VECTOR REGRESSION* DAN *RANDOM FOREST*  
*REGRESSION* DALAM PREDIKSI KONSUMSI BAHAN BAKAR  
KENDARAAN**

Oleh

**MUHAYMI NURDIN  
NPM 2017051070**

(Skripsi)



**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS LAMPUNG  
BANDAR LAMPUNG  
2025**

## DAFTAR ISI

	Halaman
DAFTAR ISI.....	II
DAFTAR GAMBAR.....	III
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian.....	5
1.4 Manfaat Penelitian.....	5
1.5 Batasan Masalah.....	6
BAB II.....	7
TINJAUAN PUSTAKA.....	7
2.1 Definisi Konsumsi Bahan Bakar.....	7
2.2 <i>Machine Learning</i> .....	11
2.3 <i>Support Vector Regression (SVR)</i> .....	13
2.4 <i>Random Forest Regression (RFR)</i> .....	19
2.5 Metrik Evaluasi.....	22
2.6 Penelitian Terdahulu.....	24
BAB III.....	26
METODOLOGI PENELITIAN.....	26
3.1 Waktu dan Tempat Penelitian.....	26
3.2 Alur Kerja Penelitian.....	26
3.3 Perangkat Penelitian.....	37
BAB V.....	38
KESIMPULAN DAN SARAN.....	38
5.1 Kesimpulan.....	38
5.2 Saran.....	39
DAFTAR PUSTAKA.....	40

## DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Ilustrasi model machine learning .....	13
Gambar 2. Ilustrasi konsep SVR (Elsa, 2023).....	13
Gambar 3 Kontruksi <i>Random Forest Regression</i> (Zahra, 2023). .....	20
Gambar 4. Alur Kerja Penelitian.....	27
Gambar 5. Deskripsi Dataset <i>Fuel Consumption Ratings</i> . .....	28
Gambar 6. Deskripsi Variasi Dataset <i>Fuel Consumption Ratings</i> . .....	29
Gambar 7. Total Dataset <i>Fuel Consumption Ratings</i> . .....	31

# BAB I

## PENDAHULUAN

### **1.1 Latar Belakang**

Pada era globalisasi, isu lingkungan dan efisiensi energi menjadi perhatian utama. Penggunaan bahan bakar fosil yang berlebihan menyebabkan polusi udara meningkatkan emisi gas rumah kaca, dan memperburuk perubahan iklim. Sektor transportasi merupakan penyumbang emisi gas rumah kaca terbesar kedua di Indonesia sekitar 23%, dengan transportasi darat menyumbang 90% dari emisi sektor ini (Bintang, 2023).

Menurut data Kementerian Energi dan Sumber Daya Mineral (ESDM), 11 juta mobil di jalanan Indonesia menghasilkan lebih dari 35 juta ton emisi CO<sub>2</sub>, sementara truk mengeluarkan lebih dari 50 juta ton (Adi, 2024). Kondisi ini mendorong perlunya upaya serius dalam meningkatkan efisiensi penggunaan bahan bakar kendaraan melalui kebijakan pemerintah dan inovasi teknologi di bidang otomotif. Pemerintah Indonesia telah berupaya mengurangi ketergantungan pada bahan bakar fosil di sektor transportasi dengan mendorong penggunaan bahan bakar nabati serta teknologi yang mendukung efisiensi energi. Di sektor logistik, perusahaan seperti PT Telkomsat mengembangkan sistem prediksi efisiensi bahan bakar untuk mengidentifikasi masalah kebocoran, pencurian bahan bakar oleh supir dan meningkatkan manajemen penggunaannya. Sistem prediksi yang akurat dapat membantu perusahaan dan pemerintah memantau serta mengelola penggunaan bahan bakar dengan lebih baik, sehingga mengurangi biaya operasional dan dampak negatif terhadap lingkungan.

Konsumsi bahan bakar kendaraan merupakan aspek penting dalam industri otomotif yang berkaitan erat dengan efisiensi energi dan dampak lingkungan. Seiring meningkatnya kesadaran global tentang perlunya pengurangan emisi gas rumah kaca, penelitian yang berfokus pada prediksi konsumsi bahan bakar menjadi semakin relevan. Prediksi yang akurat dapat membantu produsen kendaraan merancang mesin yang lebih efisien serta membantu pengguna mengelola konsumsi bahan bakar secara optimal (Alayida et al., 2023).

Secara umum, tingkat konsumsi bahan bakar minyak suatu kendaraan dipengaruhi oleh kondisi lalu lintas, faktor lingkungan, faktor kendaraan dan faktor pengemudi (Pangesty et al., 2021). Oleh karena itu, pengembangan model prediksi konsumsi bahan bakar yang akurat memerlukan algoritma yang mampu menangani kompleksitas data. Dalam beberapa tahun terakhir, metode *machine learning* menjadi pendekatan yang banyak digunakan untuk memprediksi fenomena kompleks, termasuk prediksi konsumsi bahan bakar.

Dua algoritma yang populer digunakan dalam prediksi variabel kontinu adalah *Support Vector Regression (SVR)* dan *Random Forest Regression (RFR)*. *SVR* merupakan pengembangan dari algoritma *Support Vector Machine* yang dirancang untuk menangani masalah regresi dengan mencari hyperplane terbaik yang meminimalkan kesalahan prediksi (Smola & Schölkopf, 2004). Di sisi lain, *RFR* merupakan metode ensemble yang menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting (Breiman, 2001). Kedua algoritma ini dipilih karena memiliki keunggulan masing-masing dalam memodelkan data non-linear dan mengelola variabel yang kompleks.

Dataset yang digunakan dalam penelitian ini adalah *Fuel Consumption Ratings* yang diambil dari sumber resmi pemerintah Kanada. Dataset ini mencakup informasi mendalam tentang konsumsi bahan bakar dari berbagai jenis kendaraan, baik itu kendaraan pribadi maupun komersial. Setiap entri dalam dataset mencakup sejumlah variabel yang menggambarkan berbagai faktor yang memengaruhi efisiensi bahan bakar, seperti jenis mesin, kapasitas mesin, jenis bahan bakar, transmisi, berat kendaraan, serta emisi gas buang.

Selain itu, dataset ini juga mencakup informasi teknis mengenai kendaraan, seperti tahun pembuatan, model, dan pabrikan, yang memungkinkan analisis perbandingan antar kendaraan yang memiliki spesifikasi berbeda. Data yang komprehensif ini memberikan gambaran yang jelas mengenai faktor-faktor yang berkontribusi terhadap variasi konsumsi bahan bakar di antara kendaraan yang berbeda, serta memberikan dasar yang kuat untuk membangun model prediksi yang akurat.

Pemilihan dataset ini didasarkan pada relevansinya dengan tujuan penelitian untuk mengembangkan model prediksi konsumsi bahan bakar kendaraan. Meskipun berasal dari Kanada, dataset ini memiliki potensi yang besar untuk diadaptasi dan diterapkan pada konteks yang lebih luas, termasuk di Indonesia, dengan mempertimbangkan kesamaan dalam karakteristik kendaraan dan faktor-faktor yang mempengaruhi efisiensi bahan bakar, seperti jenis bahan bakar yang digunakan dan spesifikasi teknis kendaraan. Selain itu, adaptasi model prediksi ini diharapkan dapat memberikan manfaat dalam perencanaan kebijakan transportasi yang lebih efisien dan ramah lingkungan.

Penelitian yang membandingkan kinerja *SVR* dan *RFR* dalam berbagai konteks telah menunjukkan hasil yang beragam. Dalam prediksi risiko kredit, *SVR* dengan kernel *RBF* juga menunjukkan kinerja yang lebih baik dibandingkan *RFR*. Penelitian menunjukkan bahwa *SVR* dengan kernel *RBF* memiliki *Mean Absolute Percentage Error (MAPE)* sebesar 11,63% dan *Mean Squared Error (MSE)* sebesar 0,2486, mengungguli *RFR* yang cenderung mengalami *overfitting* dengan performa buruk pada data pengujian (Tualeka, 2024). Sebaliknya, dalam prediksi laju penguapan, *RFR* menunjukkan kinerja yang lebih baik dibandingkan *SVR*. Penelitian menunjukkan bahwa *RFR* memiliki koefisien determinasi *R Square (R<sup>2</sup>)* sebesar 0,81 dan *RMSE* sebesar 0,53, sementara *SVR* memiliki *R<sup>2</sup>* sebesar 0,56 dan *RMSE* sebesar 0,81, menunjukkan bahwa *RFR* lebih akurat dalam konteks ini (Penalun et al., 2023). Namun, penelitian khusus yang membandingkan kinerja *SVR* dan *RFR* dalam konteks prediksi konsumsi bahan bakar kendaraan masih terbatas.

Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma *SVR* dan *RFR* dalam memprediksi konsumsi bahan bakar kendaraan. Hasil dari penelitian ini diharapkan dapat memberikan kontribusi dalam pemilihan metode prediksi yang lebih efektif dan akurat, serta memberikan wawasan bagi pengembangan teknologi kendaraan yang lebih efisien dan ramah lingkungan. Selain itu, penelitian ini juga mempertimbangkan aplikasi praktis sistem prediksi di industri logistik, khususnya dalam mengidentifikasi kebocoran bahan bakar, pencurian bahan bakar oleh supir dan meningkatkan efisiensi operasional.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana kinerja *Support Vector Regression (SVR)* dan *Random Forest Regression (RFR)* dalam memprediksi efisiensi bahan bakar kendaraan menggunakan dataset "*Fuel Consumption Ratings*"?
2. Sejauh mana akurasi prediksi yang dihasilkan oleh *SVR* dan *RFR* dalam hal *Mean Absolute Percentage Error (MAPE)*, *Mean Squared Error (MSE)*, *Root Mean Square Error (RMSE)*, dan koefisien determinasi *R Square (R<sup>2</sup>)*?
3. Apa saja kelebihan dan kekurangan dari masing-masing algoritma *SVR* dan *RFR* dalam memodelkan prediksi konsumsi bahan bakar kendaraan menggunakan dataset "*Fuel Consumption Ratings*"?

Rumusan masalah ini difokuskan pada perbandingan kinerja dua algoritma *machine learning*, yaitu *SVR* dan *RFR*, dalam memprediksi konsumsi bahan bakar kendaraan berdasarkan dataset yang relevan dan metrik evaluasi yang terukur.

### 1.3 Tujuan Penelitian

Berdasarkan rumusan masalah, tujuan dari penelitian ini adalah:

1. Menganalisis kinerja algoritma *Support Vector Regression (SVR)* dan *Random Forest Regression (RFR)* dalam memprediksi efisiensi bahan bakar kendaraan dengan menggunakan dataset "*Fuel Consumption Ratings*".
2. Menilai akurasi prediksi yang dihasilkan oleh algoritma *SVR* dan *RFR* berdasarkan metrik evaluasi seperti *Mean Absolute Percentage Error (MAPE)*, *Mean Squared Error (MSE)*, *Root Mean Square Error (RMSE)*, dan koefisien determinasi *R Square (R<sup>2</sup>)*.
3. Mengidentifikasi kelebihan dan kekurangan masing-masing algoritma dalam memodelkan prediksi konsumsi bahan bakar kendaraan menggunakan dataset "*Fuel Consumption Ratings*".

### 1.4 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Peneliti: Memberikan pemahaman mendalam mengenai kinerja dan karakteristik algoritma *SVR* dan *RFR* dalam konteks prediksi konsumsi bahan bakar kendaraan, serta kontribusinya dalam pengembangan model prediksi berbasis *machine learning*.
2. Bagi Pemerintah: Menyediakan informasi yang dapat digunakan untuk merumuskan kebijakan transportasi yang lebih efisien, ramah lingkungan, dan mendukung pengurangan emisi gas rumah kaca.
3. Bagi Industri Otomotif: Membantu pengembangan kendaraan yang lebih efisien bahan bakarnya dengan merancang model prediksi yang lebih akurat dan dapat diadaptasi untuk berbagai jenis kendaraan.
4. Bagi Industri Logistik: Menyediakan alternatif metode prediksi yang lebih baik untuk meningkatkan efisiensi operasional dan mengelola penggunaan bahan bakar secara lebih optimal.

5. Bagi Masyarakat: Berkontribusi dalam upaya global mengurangi polusi udara, mengurangi ketergantungan pada bahan bakar fosil, dan mengurangi dampak perubahan iklim.

### **1.5 Batasan Masalah**

Batasan masalah dalam penelitian ini adalah:

1. Dataset yang digunakan adalah "*Fuel Consumption Ratings*" yang berasal dari sumber resmi pemerintah Kanada, yang mencakup berbagai jenis kendaraan.
2. Variabel yang dianalisis terbatas pada faktor-faktor teknis kendaraan seperti jenis mesin, kapasitas mesin, jenis bahan bakar, transmisi, berat kendaraan, dan emisi gas buang, serta tidak mencakup faktor eksternal seperti kondisi lalu lintas dan cuaca.
3. Penelitian ini akan membandingkan kinerja algoritma *SVR* dan *RFR* hanya berdasarkan pada metrik evaluasi yang meliputi *MAE*, *RMSE*, dan  $R^2$ , tanpa membahas implementasi praktis di dunia industri.
4. Model prediksi yang dikembangkan hanya berfokus pada konsumsi bahan bakar kendaraan dan tidak mengelola faktor-faktor lain yang mempengaruhi efisiensi bahan bakar, seperti perilaku pengemudi atau kondisi jalan sekaligus di fokuskan untuk prediksi kendaraan kendaraan di Indonesia terutama pada sektor logistik menyesuaikan system prediksi yang sedang dikembangkan PT Telkomsat.
5. Model ini dirancang khusus untuk memprediksi konsumsi bahan bakar pada kendaraan di Indonesia, terutama di sektor logistik, dengan menyesuaikan sistem prediksi yang sedang dikembangkan oleh PT Telkomsat.

## BAB II

### TINJAUAN PUSTAKA

#### **2.1 Definisi Konsumsi Bahan Bakar**

Konsumsi bahan bakar kendaraan merujuk pada jumlah bahan bakar yang digunakan untuk menempuh jarak tertentu, yang biasanya diukur dalam satuan kilometer per liter (km/L) atau mil per galon (mpg). Efisiensi bahan bakar menjadi faktor penting dalam industri otomotif karena berpengaruh langsung terhadap biaya operasional kendaraan dan dampaknya terhadap lingkungan (Setiawan & Marlinda, 2024).

Secara ekonomis, kendaraan dengan efisiensi bahan bakar yang lebih tinggi dapat mengurangi pengeluaran untuk bahan bakar. Hal ini sangat penting bagi pengguna kendaraan yang sering melakukan perjalanan jauh atau bagi operator armada kendaraan komersial. Pengurangan konsumsi bahan bakar ini dapat menekan biaya operasional, meningkatkan efisiensi ekonomi, dan memberikan manfaat finansial yang signifikan (Syahrir et al., 2024).

Penggunaan bahan bakar yang efisien juga penting untuk mengurangi emisi gas rumah kaca dan polutan berbahaya lainnya, yang pada akhirnya dapat memperbaiki kualitas udara dan mengurangi dampak perubahan iklim. Kesadaran yang semakin meningkat terhadap masalah lingkungan telah mendorong pemerintah untuk mengimplementasikan kebijakan dan regulasi yang mendukung pengembangan kendaraan dengan efisiensi bahan bakar yang lebih baik (Altrinaldo et al., 2024).

Di sisi lain, prediksi konsumsi bahan bakar kendaraan yang akurat menjadi sangat penting, terutama dalam merencanakan biaya operasional dan strategi pengelolaan armada kendaraan. Dengan menggunakan model prediksi yang tepat, perusahaan atau individu dapat lebih mudah memperkirakan kebutuhan bahan bakar, mengoptimalkan penggunaan kendaraan, dan mengurangi pemborosan. Selain itu, prediksi yang akurat juga memungkinkan pengambilan keputusan yang lebih baik terkait pemeliharaan kendaraan, serta dapat berkontribusi pada pengurangan dampak negatif terhadap lingkungan melalui pengelolaan yang lebih efisien.

#### 1. Faktor-Faktor yang Memengaruhi Konsumsi Bahan Bakar

Faktor internal kendaraan memegang peranan penting dalam memengaruhi tingkat konsumsi bahan bakar kendaraan. Salah satu faktor utama yang memengaruhi adalah jenis mesin dan kapasitas silinder yang dimiliki oleh kendaraan. Mesin dengan kapasitas silinder yang lebih besar cenderung mengonsumsi lebih banyak bahan bakar karena membutuhkan energi yang lebih besar untuk menghasilkan tenaga. Hal ini diperkuat oleh studi yang dilakukan oleh (Xu et al., 2023) yang menyatakan bahwa kapasitas mesin memiliki korelasi positif dengan tingkat konsumsi bahan bakar dan emisi karbon dioksida yang dihasilkan, selain itu berat kendaraan juga merupakan faktor yang memengaruhi konsumsi bahan bakar. Semakin berat kendaraan, semakin banyak energi yang diperlukan untuk menggerakkannya, yang pada akhirnya meningkatkan jumlah bahan bakar yang digunakan. Menurut penelitian yang dilakukan oleh (Sharma et al., 2021), pengurangan berat kendaraan hingga 10% dapat meningkatkan efisiensi bahan bakar hingga 6%. Desain aerodinamis kendaraan juga memainkan peran penting dalam mengurangi hambatan udara selama berkendara.

Selain faktor internal, terdapat pula faktor eksternal yang memengaruhi konsumsi bahan bakar kendaraan. Kondisi jalan yang dilalui kendaraan menjadi salah satu faktor eksternal utama. Jalan yang berbukit, berlumpur, atau bergelombang memerlukan tenaga mesin yang lebih besar untuk dilalui, yang berarti lebih banyak bahan bakar akan digunakan. (Xu et al. 2020), menemukan bahwa kendaraan yang digunakan di daerah dengan infrastruktur jalan buruk memiliki tingkat konsumsi bahan bakar yang lebih tinggi dibandingkan dengan kendaraan yang digunakan di jalan datar dan mulus. Cuaca dan iklim juga memainkan peranan penting dalam efisiensi bahan bakar kendaraan. Dalam kondisi suhu dingin, mesin kendaraan memerlukan waktu yang lebih lama untuk mencapai suhu operasi optimal, sehingga meningkatkan konsumsi bahan bakar. Sebaliknya, dalam kondisi cuaca panas, sistem pendingin mesin akan bekerja lebih keras untuk menjaga suhu mesin tetap stabil, yang juga dapat memengaruhi konsumsi bahan bakar.

Gaya berkendara atau perilaku pengemudi merupakan faktor eksternal lainnya yang memengaruhi tingkat konsumsi bahan bakar. Perilaku seperti akselerasi mendadak, pengereman tiba-tiba, dan menjaga kecepatan tinggi secara terus-menerus dapat meningkatkan konsumsi bahan bakar secara signifikan. Penelitian oleh Yao et al (2021) menunjukkan bahwa pengemudi yang menerapkan gaya berkendara hemat energi dapat mengurangi konsumsi bahan bakar hingga 20%. Oleh karena itu, pelatihan pengemudi yang baik dapat menjadi solusi efektif untuk mengurangi penggunaan bahan bakar yang berlebihan. Faktor terakhir yang perlu diperhatikan adalah beban kendaraan. Beban tambahan seperti penumpang atau barang bawaan akan meningkatkan energi yang dibutuhkan untuk menggerakkan kendaraan, sehingga berdampak pada konsumsi bahan bakar yang lebih tinggi.

## 2. Prediksi Konsumsi Bahan Bakar Kendaraan

Pentingnya prediksi konsumsi bahan bakar dalam efisiensi energi tidak dapat diabaikan dalam konteks pengelolaan sumber daya energi yang berkelanjutan. Prediksi konsumsi bahan bakar memungkinkan perencanaan yang lebih akurat terkait kebutuhan bahan bakar di masa mendatang, yang pada akhirnya dapat mengurangi pemborosan dan meningkatkan efisiensi penggunaan energi. Menurut penelitian yang dilakukan oleh (Li et al., 2022), metode prediktif yang menggunakan algoritma pembelajaran mesin dapat meningkatkan akurasi dalam memperkirakan kebutuhan bahan bakar hingga 90%. Hal ini sangat bermanfaat dalam sektor transportasi, di mana efisiensi bahan bakar berperan penting dalam mengurangi emisi karbon dan menghemat biaya operasional.

Prediksi konsumsi bahan bakar juga membantu dalam pengembangan kebijakan energi yang lebih efektif. Pemerintah dan pembuat kebijakan dapat menggunakan hasil prediksi ini untuk merancang strategi yang mendukung pengurangan penggunaan bahan bakar fosil serta mendorong penggunaan energi terbarukan. Selain itu, produsen kendaraan dapat memanfaatkan data prediksi ini untuk mengembangkan teknologi mesin yang lebih hemat energi. Sebagai contoh, penggunaan model prediksi konsumsi bahan bakar dapat membantu produsen dalam mengoptimalkan desain kendaraan agar lebih efisien dalam penggunaan bahan bakar (Li et al., 2022).

Dalam konteks individu, prediksi konsumsi bahan bakar dapat membantu pengemudi dalam mengadopsi gaya berkendara yang lebih hemat energi. Dengan adanya sistem yang mampu memprediksi konsumsi bahan bakar secara *real-time*, pengemudi dapat menyesuaikan perilaku mengemudi untuk mengurangi penggunaan bahan bakar, seperti menghindari akselerasi yang tiba-tiba atau menjaga kecepatan yang stabil. Hal ini tidak hanya mengurangi biaya bahan bakar, tetapi juga berkontribusi pada pengurangan

emisi gas rumah kaca. Secara keseluruhan, kemampuan untuk memprediksi konsumsi bahan bakar memiliki dampak yang luas dalam meningkatkan efisiensi energi di berbagai sektor. Implementasi teknologi prediktif ini dapat membantu mengurangi ketergantungan pada bahan bakar fosil, mengurangi emisi, dan meningkatkan keberlanjutan energi di masa depan.

## 2.2 Machine Learning

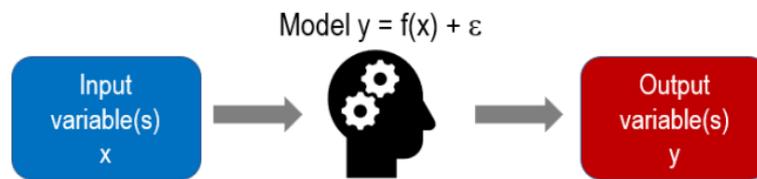
Pembelajaran mesin (*machine learning*) adalah cabang dari kecerdasan buatan yang memungkinkan sistem untuk belajar dan membuat prediksi atau keputusan berdasarkan data tanpa diprogram secara eksplisit. Pembelajaran mesin mengandalkan algoritma yang memanfaatkan data untuk mengidentifikasi pola dan membuat keputusan atau prediksi di masa depan (Alpaydin, 2020). Prinsip dasar pembelajaran mesin adalah pengembangan model matematis yang dapat mengambil input berupa data, memprosesnya, dan menghasilkan output berupa prediksi atau keputusan yang akurat berdasarkan data yang telah dilatih (Goodfellow et al., 2016). Pembelajaran mesin umumnya dibagi menjadi tiga kategori utama, *supervised learning*, *unsupervised learning* dan *reinforcement learning* (Mitchell, 1997). Pada *supervised learning*, model dilatih menggunakan dataset yang terdiri dari input dan output yang telah diketahui atau diberi label. Algoritma akan belajar untuk memetakan input ke output berdasarkan data pelatihan, sehingga ketika diberikan data baru, model dapat memprediksi output dengan akurasi tinggi. Contoh algoritma yang sering digunakan dalam supervised learning adalah *linear regression*, *decision trees*, *support vector regression* dan *random forest regression*.

Misalnya, dalam konteks prediksi efisiensi bahan bakar, dataset berisi informasi tentang fitur-fitur kendaraan (seperti kelas kendaraan, jenis transmisi, dan ukuran mesin) sebagai input, dan konsumsi bahan bakar sebagai output. Model yang dilatih dengan data ini dapat digunakan untuk memprediksi konsumsi bahan bakar kendaraan baru. Berbeda dengan *supervised learning*, *unsupervised learning* bekerja dengan data yang tidak diberi label, artinya algoritma tidak memiliki output

yang diketahui sebelumnya. Tujuannya adalah untuk menemukan struktur atau pola yang tersembunyi dalam data. Contoh penerapannya adalah *clustering*, di mana data dikelompokkan berdasarkan kesamaan antar data. Algoritma seperti *k-means clustering* atau *hierarchical clustering* sering digunakan dalam *unsupervised learning*. Ini dapat digunakan, misalnya, untuk mengidentifikasi kelompok kendaraan dengan pola konsumsi bahan bakar yang serupa.

Selain *supervised learning* dan *unsupervised learning*, terdapat kategori lain dalam pembelajaran mesin yang disebut *reinforcement learning*. *Reinforcement learning* adalah tipe *machine learning* yang menggunakan konsep *reward/punishment*, di mana model akan berusaha memaksimalkan *reward* (atau mengurangi *penalty/punishment*) jika berhasil mencapai tujuan yang telah ditentukan. Dalam *reinforcement learning*, model belajar melalui interaksi dengan lingkungan dan mendapatkan umpan balik berupa *reward* atau *punishment* berdasarkan tindakannya. Model ini tidak dilatih dengan data yang memiliki label atau output yang sudah diketahui, melainkan dengan mencoba berbagai tindakan dan memperbaiki perilakunya seiring waktu berdasarkan hasil yang diperoleh. Penerapan *reinforcement learning* sering ditemui dalam bidang robotik, seperti misalnya mengajarkan bagaimana robot dapat berjalan, berlari, atau melakukan aktivitas lainnya dengan cara yang efisien. Dalam hal ini, robot akan mendapatkan *reward* ketika melakukan tugas dengan benar, dan *punishment* ketika melakukan kesalahan, seperti jatuh atau gagal mencapai tujuan yang diinginkan. Tujuan dari *reinforcement learning* adalah untuk menemukan kebijakan (*policy*) yang dapat memaksimalkan total *reward* dalam jangka panjang, yang mengarah pada pengambilan keputusan yang optimal (Maesaroh, 2024).

Implementasi *machine learning* yang digunakan di dunia nyata berupa model *machine learning* (Gambar 1). Model ini merupakan algoritma statistik yang bisa merupakan hasil dari implementasi *supervised*, *unsupervised*, atau *reinforcement learning*.

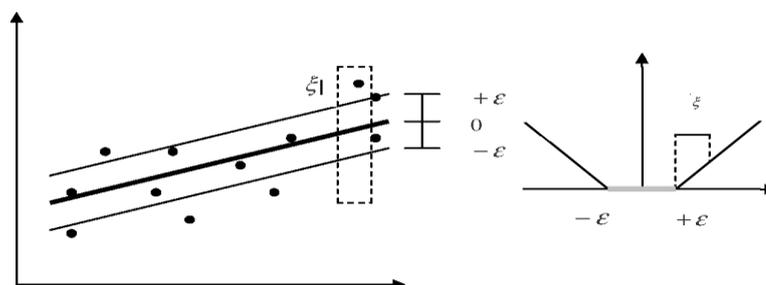


Gambar 1. Ilustrasi model machine learning.

*Machine Learning (ML)* adalah cabang kecerdasan buatan yang memungkinkan sistem belajar dari data dan membuat keputusan tanpa diprogram secara eksplisit. Dengan kemampuan mengenali pola dalam data, *ML* telah menjadi inti dari berbagai aplikasi teknologi modern seperti sistem rekomendasi, pengenalan wajah, dan kendaraan otonom. Teknik seperti deep learning memungkinkan pencapaian besar dalam pengenalan gambar dan pemrosesan bahasa alami. *ML* digunakan di berbagai sektor seperti kesehatan, keuangan, dan manufaktur, untuk analisis data, deteksi penipuan, dan prediksi perawatan mesin. Kemampuannya untuk beradaptasi dengan data baru menjadikannya semakin penting di masa depan.

### 2.3 Support Vector Regression (SVR)

*Support Vector Regression (SVR)* adalah metode regresi yang berbasis pada *Support Vector Machine (SVM)*. Tujuan utama *SVR* adalah memprediksi nilai output *Y* berdasarkan input *X* dengan toleransi kesalahan tertentu ( $\epsilon$  *epsilon*). Tidak seperti regresi linier biasa, *SVR* tidak hanya mencari garis terbaik untuk mendekati data, tetapi juga mempertimbangkan margin toleransi (Vapnik, 1995).



Gambar 2. Ilustrasi konsep SVR (Elsa, 2023).

Pada Gambar 2 garis hitam tebal merupakan *hyperplane* sedangkan untuk dua garis yang mengapitnya adalah *soft margin*. Jarak antara *hyperplane* dan *soft margin* adalah sebesar  $\epsilon$  dan titik-titik yang berada pada  $+\epsilon$  sampai  $-\epsilon$  merupakan *support vector*, namun untuk titik yang melewati *soft margin* dibutuhkan adanya variabel slack  $\xi$  (Elsa, 2023).

*Support Vector Regression (SVR)* merupakan pengembangan dari *Support Vector Machine (SVM)*, yang awalnya dirancang untuk masalah klasifikasi. Sementara *SVM* digunakan untuk memisahkan data ke dalam kategori yang berbeda dengan margin yang maksimal, *SVR* digunakan untuk masalah regresi, yaitu memprediksi nilai kontinu. Tujuan utama dari *SVR* adalah mencari fungsi regresi yang dapat meminimalkan kesalahan prediksi, namun dengan memperkenalkan margin toleransi. Margin toleransi ini diwakili oleh parameter  $\epsilon$  (*epsilon*), yang menentukan seberapa besar kesalahan yang masih diterima tanpa memberikan penalti. Dalam hal ini, model *SVR* berusaha untuk memprediksi nilai yang berada dalam rentang margin yang dapat diterima. Jika prediksi keluar dari margin toleransi tersebut, maka kesalahan tersebut akan dikenakan penalti yang sebanding. Dengan kata lain, *SVR* tidak hanya berfokus pada meminimalkan kesalahan secara keseluruhan, tetapi juga mengontrol seberapa besar kesalahan yang diperbolehkan, dengan mengabaikan kesalahan kecil yang berada dalam rentang margin  $\epsilon$ . Berikut adalah rumus yang digunakan dalam *SVR* (Elsa, 2023):

1. Fungsi Regresi *SVR* (Umum):

$$f(x) = w \cdot x + b \quad (1)$$

Dimana:

- $f(x)$  adalah fungsi regresi yang dihasilkan dari model.
- $w$  adalah vektor pembobot.
- $x$  adalah vektor input.
- $b$  adalah bias.

Fungsi ini digunakan untuk menghitung nilai output berdasarkan input  $x$ .

## 2. Masalah Optimasi SVR:

Untuk meminimalkan kesalahan dan mengoptimalkan fungsi regresi, kita meminimalkan norma dari vektor pembobot  $w$  dan penalti kesalahan melalui variabel slack  $\xi_i$ , dengan rumus:

$$\min \|w\|^2 + C \sum \xi_i \quad (2)$$

dengan syarat:

$$|y_i - (w \cdot x_i + b)| \leq \epsilon + \xi_i \quad (3)$$

- $y_i$  adalah nilai target (nilai yang sebenarnya) untuk data ke- $i$ .
- $x_i$  adalah vektor input untuk data ke- $i$ .
- $\xi_i$  adalah variabel slack yang menunjukkan margin kesalahan pada data ke- $i$ .
- $\epsilon$  adalah toleransi margin kesalahan, yaitu seberapa besar kesalahan yang dapat diterima.
- $C$  adalah parameter cost yang mengontrol penalti kesalahan pada data pelatihan.

Tujuan optimasi ini adalah untuk meminimalkan bobot  $w$  (agar model lebih sederhana) dan penalti  $\xi_i$  (agar kesalahan prediksi tetap kecil).

## 3. Fungsi *Loss Insensitive*:

Fungsi *loss insensitive* digunakan untuk menentukan toleransi kesalahan dalam prediksi model. Toleransi ini diwakili oleh parameter  $\epsilon$  (*epsilon*) yang berfungsi sebagai ambang batas atau margin kesalahan yang masih dapat diterima tanpa dikenakan penalti.

Apabila nilai prediksi model berada dalam rentang margin yang ditentukan oleh  $\epsilon$ , maka kesalahan tersebut dianggap tidak signifikan dan tidak memengaruhi fungsi loss. Sebaliknya, jika kesalahan prediksi melebihi nilai  $\epsilon$ , maka fungsi loss akan memberikan penalti sesuai dengan selisih antara nilai prediksi dan nilai aktual. Dengan demikian, fungsi ini membantu

model memfokuskan perhatian pada kesalahan yang substansial, sambil mengabaikan kesalahan kecil yang masih berada dalam batas toleransi.

rumus:

$$L(w, b, \alpha) = \frac{1}{2} \|w\|^2 - \sum_{i=1}^n \alpha_i (y_i - w \cdot x_i - b) \quad (4)$$

Dimana:

- $\alpha_i$  adalah *Lagrange Multiplier* yang menunjukkan kontribusi setiap data terhadap fungsi objektif.

Dengan menggunakan metode *Lagrange*, kita bisa mendapatkan solusi optimal untuk  $w$  dan  $b$ .

#### 4. Transformasi Non-Linear ke Feature Space

Dalam permasalahan *non-linear*, data tidak dapat dipisahkan dengan menggunakan garis lurus atau *hyperplane* di ruang input. Oleh karena itu, digunakan fungsi kernel yang memetakan data ke dalam ruang fitur berdimensi lebih tinggi. Transformasi ini memungkinkan data yang tidak dapat dipisahkan secara linear di ruang input menjadi dapat dipisahkan secara linear setelah dipetakan ke ruang fitur. Fungsi kernel  $K(x, x')$  digunakan untuk menghitung *dot product* antara dua vektor di ruang fitur tanpa perlu menghitung transformasi eksplisitnya, yaitu:

$$K(x, x') = \langle \phi(x), \phi(x') \rangle \quad (5)$$

Dimana  $\phi(x)$  adalah transformasi data ke dalam ruang fitur yang lebih tinggi. Dengan menggunakan kernel, rumus untuk prediksi pada SVR menjadi:

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) \quad (6)$$

Dengan demikian, model SVR dapat menangani data *non-linear* dengan menggunakan kernel untuk menghitung *dot product* di ruang fitur yang lebih tinggi.

## 5. Fungsi Kernel

Fungsi kernel yang umum digunakan dalam *SVM* dan *SVR* adalah fungsi yang memetakan data ke ruang dimensi tinggi secara efisien, contohnya adalah:

$$\text{Linear} : K(x, x') = x \cdot x' \quad (7)$$

Kernel linear adalah fungsi kernel yang paling sederhana, di mana pemisahan data dilakukan secara langsung dalam ruang fitur asli tanpa transformasi tambahan. Kernel linear cocok digunakan ketika data sudah dapat dipisahkan secara linier dalam ruang fitur asli.  $x$  dan  $x'$  adalah vektor fitur dari dua sampel data.

$$\text{Radial Basis Function (RBF)} : K(x, x') = \exp(-\gamma \|x - x'\|^2) \quad (8)$$

Kernel RBF, juga dikenal sebagai kernel Gaussian, digunakan untuk klasifikasi data non-linier. Kernel ini memiliki kemampuan untuk menangani data yang tidak dapat dipisahkan secara linier dengan baik. Parameter  $\gamma$  memengaruhi sebaran kernel, nilai  $\gamma$  yang lebih besar membuat kernel lebih sensitif terhadap jarak antar titik data, sementara nilai  $\gamma$  yang lebih kecil membuat kernel lebih luas. Penyetelan parameter  $\gamma$  harus dilakukan dengan hati-hati karena berpengaruh besar terhadap kinerja kernel, sedangkan  $\|x - x'\|^2$  adalah kuadrat jarak Euclidean antara dua vektor fitur.

$$\text{Polynomial} : K(x_i, x_j) = (x_i \cdot x_j + c)^d \quad (9)$$

Kernel polinomial digunakan untuk menangani data yang memiliki hubungan *non-linier* yang dapat dipisahkan dengan fungsi polinomial. Parameter  $c$  dan  $d$  memengaruhi fleksibilitas dan kompleksitas model *SVM*. Nilai  $c$  yang lebih besar meningkatkan pengaruh suku bias, sementara nilai  $d$  yang lebih tinggi memungkinkan model menangkap interaksi fitur yang lebih kompleks. Namun, penggunaan nilai  $d$  yang terlalu tinggi dapat menyebabkan *overfitting*, di mana model terlalu menyesuaikan diri dengan data pelatihan dan kehilangan kemampuan generalisasi pada data baru, d

adalah derajat polinomial yang menentukan kompleksitas pemisahan dan  $c$  adalah konstanta bebas yang menambahkan bias.

Fungsi kernel memungkinkan model *SVM* dan *SVR* untuk mengatasi masalah *non-linear* tanpa perlu menghitung transformasi data secara eksplisit ke dalam ruang fitur tinggi. Dengan menggunakan kernel, persamaan untuk prediksi pada *SVR* menjadi:

$$f(x) = \sum_{i=1}^n \alpha_i K(x_i, x) \quad (10)$$

dimana  $K(x_i, x)$  adalah hasil kernel antara data pelatihan  $x_i$  dan data uji  $x$ .

*SVR* merupakan metode regresi yang kuat dalam menangani masalah prediksi nilai kontinu dengan margin kesalahan yang terkendali. Dengan menggunakan konsep margin dan support vector, *SVR* dapat menghasilkan model yang akurat meskipun pada masalah *non-linear*, berkat penggunaan fungsi kernel. Melalui optimasi dan teknik *Lagrange Multiplier*, *SVR* mampu mengatasi kendala dan mencapai solusi yang optimal untuk masalah regresi.

*Support Vector Regression (SVR)* memiliki beberapa kelebihan, di antaranya kemampuannya untuk menangani masalah *non-linear* dengan menggunakan fungsi kernel, serta menghindari *overfitting* dengan mengontrol margin toleransi kesalahan. *SVR* juga efektif pada data berdimensi tinggi dan memberikan hasil yang lebih stabil karena dapat mengabaikan kesalahan kecil. Selain itu, fleksibilitas dalam memilih kernel sesuai karakteristik data membuatnya cocok untuk berbagai jenis masalah regresi.

Namun, *SVR* juga memiliki kekurangan, seperti waktu komputasi yang lebih lama pada dataset besar, serta ketergantungan pada pemilihan parameter yang tepat untuk menghasilkan model yang optimal. Model ini juga cenderung sulit diinterpretasikan, terutama ketika menggunakan kernel *non-linear*, dan bisa menjadi kurang efektif jika data mengandung banyak *noise* atau *outlier*. Pemilihan kernel yang tepat juga sering menjadi tantangan dalam penggunaannya.

## 2.4 *Random Forest Regression (RFR)*

*Random Forest* adalah metode *ensemble* yang terdiri dari banyak pohon keputusan (*decision trees*). Setiap pohon keputusan dalam *Random Forest* dibuat menggunakan subset acak dari data pelatihan, dan prediksi yang dihasilkan oleh *Random Forest* adalah hasil rata-rata dari prediksi seluruh pohon keputusan. Pada *Random Forest Regression*, masing-masing pohon keputusan menghasilkan prediksi untuk nilai kontinu, dan hasil akhirnya dihitung dengan mengambil rata-rata dari seluruh prediksi pohon.

Pembuatan pohon keputusan dalam *Random Forest* dilakukan dengan dua tahap penting (Zahra, 2023):

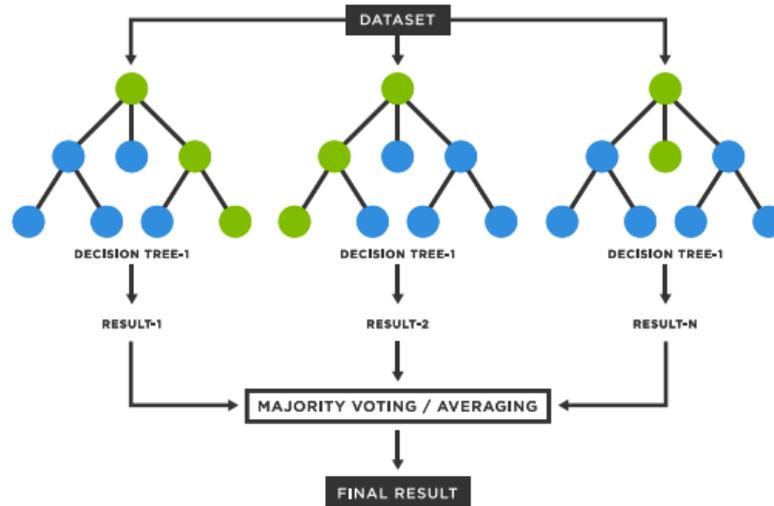
1. *Bootstrapping*:

Setiap pohon keputusan dibangun dengan subset acak dari data pelatihan yang diperoleh melalui teknik *bootstrapping*, yaitu pengambilan sampel dengan penggantian.

2. Pemilihan Fitur Secara Acak:

Dalam setiap *split* di setiap pohon keputusan, hanya subset acak dari fitur yang dipertimbangkan, yang meningkatkan keberagaman pohon keputusan dalam *Random Forest*.

Proses pembuatan *Random Forest Regression* dapat dijelaskan dengan langkah-langkah berikut (Zahra, 2023):



Gambar 3 Kontruksi *Random Forest Regression* (Zahra, 2023).

1. Pembuatan Pohon Keputusan:

Untuk setiap pohon keputusan  $T$ , algoritma akan memilih subset acak dari data pelatihan  $D_t$  dan subset acak dari fitur. Pohon keputusan akan membagi data pada setiap node berdasarkan kriteria pembagian yang optimal (misalnya, *Mean Squared Error* pada regresi).

2. Prediksi untuk Setiap Pohon:

Setiap pohon keputusan akan menghasilkan prediksi  $y_t$  untuk input data  $x$ .  
 Prediksi Akhir *Random Forest*: Prediksi akhir untuk data  $x$  dihitung sebagai rata-rata prediksi dari semua pohon keputusan:

$$\hat{y} = \frac{1}{T} \sum_{t=1}^T \hat{y}_t \quad (11)$$

dimana:

- $T$  adalah jumlah pohon keputusan dalam *Random Forest*.
- $\hat{y}_t$  adalah prediksi yang dihasilkan oleh pohon keputusan ke- $t$ .

*Random Forest* memiliki beberapa keuntungan yang membuatnya sangat populer dalam masalah regresi:

1. Ketahanan terhadap *Overfitting*:

Karena setiap pohon keputusan dibangun menggunakan subset acak dari data dan fitur, Random Forest mampu mengurangi *overfitting*, meskipun pohon keputusan individu mungkin sangat dalam dan *overfitting* pada data pelatihan.

2. Kemampuan untuk Menangani Data yang Tidak Terstruktur:

*Random Forest* dapat menangani data dengan banyak fitur atau fitur yang berskala sangat berbeda tanpa memerlukan normalisasi data.

3. Fitur Penting:

*Random Forest* dapat digunakan untuk mengukur pentingnya fitur dengan menghitung seberapa sering suatu fitur digunakan dalam pembagian pohon, yang membantu dalam pemilihan fitur.

Meskipun *Random Forest* memiliki banyak keuntungan, ada beberapa hal yang perlu diperhatikan:

1. Kompleksitas dan Waktu Komputasi:

Membangun banyak pohon keputusan dan menghitung prediksi dari masing-masing pohon membutuhkan waktu komputasi yang lebih lama, terutama untuk dataset besar.

2. Interpretasi Model:

Meskipun Random Forest dapat memberikan akurasi yang baik, model ini cenderung menjadi “*black box*” yang berarti bahwa interpretasi model untuk mengetahui bagaimana keputusan diambil bisa sulit dilakukan.

Untuk mengatasi masalah ini, teknik feature selection dapat digunakan yaitu proses memilih subset fitur yang paling relevan dan signifikan dari sekumpulan data untuk digunakan dalam proses pelatihan model. Dalam konteks *machine learning*, khususnya pada algoritma seperti *Random Forest*, *feature selection* sangat penting karena berfungsi untuk meningkatkan akurasi model dan mengurangi kompleksitas

komputasi. Proses ini bertujuan untuk menghilangkan fitur yang tidak relevan atau redundan yang dapat menyebabkan *overfitting* dan memperlambat waktu komputasi.

## 2.5 Metrik Evaluasi

Dalam proses evaluasi model regresi, penggunaan metrik evaluasi yang tepat sangat penting untuk memahami sejauh mana model dapat memprediksi nilai target secara akurat. Setiap metrik memiliki karakteristik dan kegunaannya masing-masing yang memberikan perspektif berbeda terhadap performa model. Tiga metrik yang umum digunakan adalah *Root Mean Squared Error (RMSE)*, *Mean Absolute Percentage Error (MAPE)*, dan koefisien determinasi ( $R^2$ ).

### 1. *Root Mean Squared Error (RMSE)*:

*RMSE* mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai aktual. Semakin kecil nilai *RMSE*, semakin baik kinerja model dalam memprediksi nilai target. *RMSE* dihitung dengan rumus:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (12)$$

Di mana:

- $n$  adalah jumlah data,
- $y_i$  adalah nilai aktual,
- $\hat{y}_i$  adalah nilai prediksi.

*RMSE* memberikan gambaran tentang seberapa besar kesalahan prediksi model dalam satuan yang sama dengan variabel target. *RMSE* mengukur rata-rata kesalahan kuadrat antara nilai prediksi dan nilai aktual. Semakin kecil nilai *RMSE*, semakin baik kinerja model dalam memprediksi nilai target. Nilai *RMSE* yang rendah menunjukkan bahwa variasi nilai yang dihasilkan oleh suatu model prakiraan mendekati variasi nilai observasinya.

## 2. Mean Absolute Percentage Error (MAPE):

MAPE mengukur rata-rata persentase kesalahan absolut antara nilai prediksi dan nilai aktual. MAPE dihitung dengan rumus:

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \times 100\% \quad (13)$$

Di mana:

- $n$  adalah jumlah data,
- $y_i$  adalah nilai aktual,
- $\hat{y}_i$  adalah nilai prediksi.

Nilai MAPE yang lebih kecil menunjukkan akurasi model yang lebih baik. Namun, MAPE dapat menjadi tidak terdefinisi jika terdapat nilai aktual yang nol. MAPE mengukur rata-rata persentase kesalahan absolut antara nilai prediksi dan nilai aktual. Nilai MAPE yang lebih kecil menunjukkan akurasi model yang lebih baik. Secara umum, interpretasi nilai MAPE adalah 0 – 100 %, nilai MAPE lebih dari 50% bisa dikatakan merupakan hasil prediksi yang buruk sedangkan kurang dari 50% mendekati hasil yang layak dan kurang dari 20% merupakan hasil prediksi yang sangat baik.

## 3. Koefisien Determinasi ( $R^2$ ):

$R^2$  mengukur proporsi variasi dalam data target yang dapat dijelaskan oleh model. Nilai  $R^2$  berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan model yang lebih baik dalam menjelaskan variasi data.  $R^2$  dihitung dengan rumus:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (14)$$

Di mana:

- $n$  adalah jumlah data,
- $y_i$  adalah nilai aktual,
- $\hat{y}_i$  adalah nilai prediksi.
- $\bar{y}$  adalah rata-rata nilai aktual.

$R^2$  mengukur proporsi variasi dalam data target yang dapat dijelaskan oleh model. Nilai  $R^2$  berkisar antara 0 hingga 1, di mana nilai yang lebih tinggi menunjukkan model yang lebih baik dalam menjelaskan variasi data. Nilai  $R^2$  yang lebih tinggi menunjukkan bahwa model dapat menjelaskan variasi data dengan lebih baik.

Pemilihan metrik evaluasi yang tepat sangat penting dalam menilai kinerja model regresi. Setiap metrik memiliki kelebihan dan kekurangan, sehingga sering kali digunakan secara bersamaan untuk mendapatkan gambaran yang lebih komprehensif tentang akurasi model.

## 2.6 Penelitian Terdahulu

Dalam beberapa tahun terakhir, perkembangan teknologi *machine learning* telah membuka peluang baru di berbagai bidang, salah satunya dalam prediksi konsumsi bahan bakar kendaraan. Berbagai metode *machine learning* telah diterapkan untuk tujuan ini, di antaranya *Random Forest Regression* (RFR) dan *Support Vector Regression* (SVR). Namun, studi yang secara khusus membandingkan kinerja RFR dan SVR dalam konteks prediksi konsumsi bahan bakar kendaraan masih terbatas. Sebagian besar penelitian cenderung berfokus pada penggunaan RFR, yang dikenal memiliki kemampuan baik dalam menangani data kompleks dan tidak linier. Oleh karena itu, penelitian yang membandingkan kedua metode ini menjadi sangat relevan untuk memberikan wawasan lebih mendalam mengenai kelebihan dan kekurangan masing-masing dalam memprediksi konsumsi bahan bakar kendaraan.

### 1. Studi Perbandingan RFR dan SVR dalam Prediksi Laju Penguapan

Salah satu penelitian yang relevan adalah studi berjudul "Perbandingan *Random Forest Regression* dan *Support Vector Regression* pada Prediksi Laju Penguapan" (Penalun et al., 2022). Penelitian ini menggunakan data pengamatan meteorologi harian untuk membandingkan kinerja RFR dan SVR dalam memprediksi laju penguapan. Hasil optimasi *hyperparameter* pada data pelatihan menunjukkan bahwa model RFR menghasilkan skor *Root Mean Square Error* (RMSE) sebesar -0,67, sedangkan model SVR dengan kernel *Radial Basis Function* (RBF) mencatat skor RMSE sebesar -0,57. Pada tahap evaluasi lebih lanjut dengan data pengujian, model RFR menghasilkan nilai *Coefficient of Determination* ( $R^2$ ) sebesar 0,79 dan RMSE sebesar 0,56, sementara SVR mencapai  $R^2$  sebesar 0,81 dan RMSE sebesar 0,53. Berdasarkan hasil ini, SVR menunjukkan kinerja yang lebih baik dalam memprediksi laju penguapan harian. Meskipun fokus penelitian

ini berbeda dengan prediksi konsumsi bahan bakar kendaraan, hasilnya menegaskan bahwa SVR memiliki potensi besar dan relevansi tinggi dalam prediksi berbasis data kompleks.

## 2. Aplikasi SVR dalam Prediksi Konsumsi Bahan Bakar PLTGU

Penelitian lain yang relevan berjudul "Prediksi Penggunaan Bahan Bakar pada PLTGU menggunakan Metode *Support Vector Regression (SVR)*" (Novianti et al., 2022). Studi ini memanfaatkan data penggunaan bahan bakar gas dari PT. Indonesia Power untuk memprediksi konsumsi bahan bakar pada Pembangkit Listrik Tenaga Gas dan Uap (PLTGU). Hasilnya menunjukkan bahwa model SVR memiliki kinerja baik dengan *Mean Absolute Percentage Error (MAPE)* sebesar 7,7513%. Nilai MAPE di bawah 10% ini mengindikasikan akurasi prediksi yang sangat baik. Penelitian ini menegaskan bahwa SVR dapat menjadi solusi efektif untuk memprediksi konsumsi bahan bakar dalam konteks spesifik seperti pembangkit listrik.

## 3. Kinerja RFR dalam Prediksi Konsumsi Bahan Bakar Kendaraan Berbensin

Penelitian berjudul "Predicting Gasoline Vehicle Fuel Consumption in Energy and Environmental Impact Based on *Machine Learning* and Multidimensional Big Data" (Yang et al., 2022) juga memberikan kontribusi penting. Studi ini membandingkan kinerja berbagai metode *machine learning*, termasuk RFR, Regresi Linier, Regresi *Naïve Bayes*, Regresi Jaringan Syaraf Tiruan, dan *LightGBM*, dalam memprediksi konsumsi bahan bakar kendaraan bermotor berbahan bakar bensin. Data yang digunakan mencakup faktor kendaraan, lingkungan, serta perilaku mengemudi, yang dikumpulkan melalui aplikasi *BearOil*, survei perilaku mengemudi, serta data iklim dan kondisi jalan di berbagai kota di Tiongkok. Hasil evaluasi menunjukkan bahwa RFR memiliki performa terbaik dengan MAPE sebesar 7,5%, *Mean Squared Error (MSE)* sebesar 0,805, dan  $R^2$  sebesar 0,776. Nilai-nilai ini menunjukkan bahwa RFR sangat efektif dalam konteks prediksi konsumsi bahan bakar kendaraan berbensin.

RFR dan SVR efektif untuk memprediksi konsumsi bahan bakar, dengan keunggulan masing-masing. RFR unggul pada data kompleks, sedangkan SVR cocok untuk prediksi spesifik. Penelitian ini akan membandingkan kinerja keduanya dalam konteks konsumsi bahan bakar kendaraan untuk memberikan wawasan lebih mendalam.

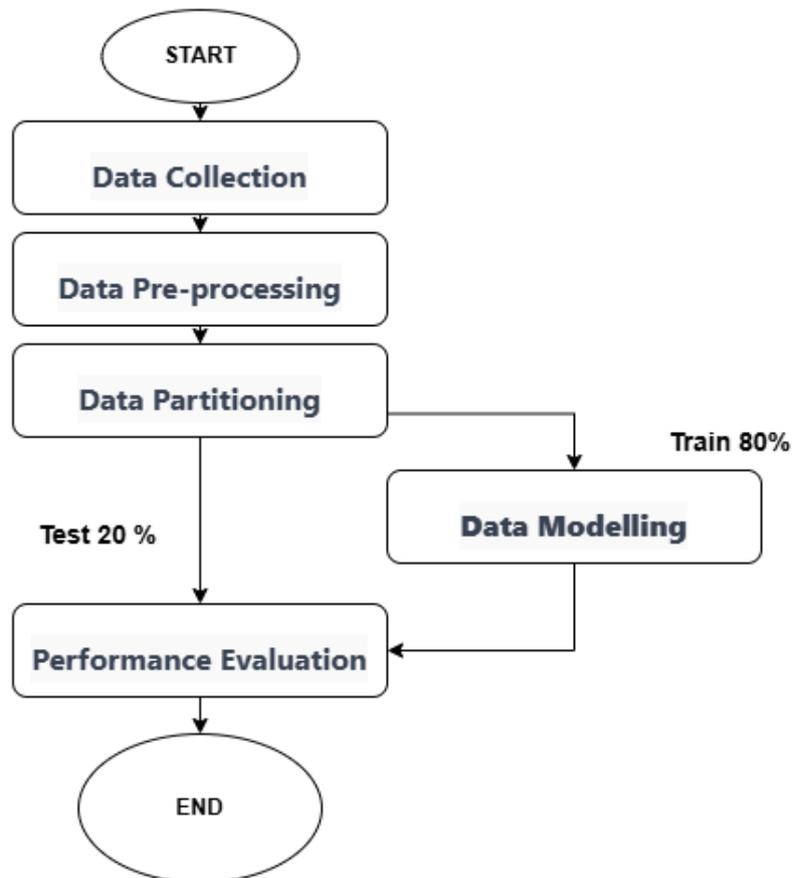
## BAB III METODOLOGI PENELITIAN

### **3.1 Waktu dan Tempat Penelitian**

Penelitian ini dilaksanakan pada semester genap tahun ajaran 2023/2024 bertempat di Jurusan Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lampung.

### **3.2 Alur Kerja Penelitian**

Pada bagian ini, disajikan alur kerja penelitian secara keseluruhan, yang mencakup tahapan dari pengumpulan data hingga analisis hasil. Penelitian ini bertujuan untuk mengevaluasi kinerja *Support Vector Regression (SVR)* dan *Random Forest Regression (RFR)* dalam memprediksi konsumsi bahan bakar kendaraan. Setiap tahap dalam penelitian ini dirancang secara sistematis untuk memastikan validitas data, akurasi model, serta interpretasi hasil yang optimal. Adapun alur kerja penelitian ini dapat dilihat pada Gambar 4, yang memberikan ilustrasi mengenai proses yang dilakukan pada setiap tahap.



Gambar 4. Alur Kerja Penelitian.

## 1. Data Penelitian

Sumber data yang digunakan dalam penelitian ini adalah dataset yang diunduh dari situs [open.data.ca](https://open.canada.ca/data/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64/resource/ec44c747-a640-4481-bcf0-a2ea7ca8f249), khususnya dari "*Fuel Consumption Ratings*" dataset, <https://open.canada.ca/data/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64/resource/ec44c747-a640-4481-bcf0-a2ea7ca8f249>.

Dataset ini disediakan oleh pemerintah Kanada dan merupakan sumber yang kaya informasi mengenai konsumsi bahan bakar kendaraan bermotor di Kanada. Dataset ini mencakup berbagai data terkait efisiensi bahan bakar,

emisi, dan karakteristik kendaraan yang diperlukan untuk analisis dan model prediksi.

Deskripsi Dataset		
Data	Tipe data	Deskripsi
Transmission	Gabungan	mencakup tipe transmisi dan jumlah gigi transmisi kendaraan : A = Automatic (Otomatis), AM = Automated Manual (Manual Otomatis), AS = Automatic with Selective Shift (Otomatis dengan Perpindahan Pilih), AV = Continuously Variable (Variabel Kontinu), M = Manual (Manual) , disertai Gear (Gigi Transmisi).
Fuel type	Kategorikal	nama jenis bahan bakar yang digunakan : X = Bensin biasa; Z = Bensin premium; D = Diesel; E = E85; N = Gas Alam Untuk FFV, nilai konsumsi disediakan untuk bensin dan E85.
Make	Kategorikal	merek kendaraan.
Model	Kategorikal	model kendaraan.
Vehicle class	Kategorikal	kelas kendaraan.
Year	Numerik	tahun pembuatan kendaraan.
Engine size (L)	Numerik	ukuran mesin dalam liter.
Cylinders	Numerik	jumlah silinder mesin.
CO2 rating	Numerik	rating emisi CO2, dinilai pada skala dari 1 (terburuk) hingga 10 (terbaik).
Smog rating	Numerik	rating polusi udara dari asap knalpot , dinilai pada skala dari 1 (terburuk) hingga 10 (terbaik).
CO2 emissions (g/km)	Numerik	emisi CO2 kendaraan dalam gram per kilometer
Highway (L/100 km)	Numerik	konsumsi bahan bakar kendaraan di jalan raya dalam liter per 100 kilometer .
City (L/100 km)	Numerik	konsumsi bahan bakar kendaraan di kota dalam liter per 100 kilometer, mewakili mengemudi perkotaan dalam lalu lintas stop-and-go.
Combined (mpg)	Numerik	konsumsi bahan bakar gabungan dalam mil per galon
Combined (L/100 km)	Numerik	konsumsi bahan bakar gabungan dalam liter per 100 kilometer ( 55% mengemudi di kota dan 45% mengemudi di jalan raya).

Gambar 5. Deskripsi Dataset *Fuel Consumption Ratings*.

Dataset yang akan digunakan dalam analisis ini merupakan data *Fuel Consumption Ratings* yang mencakup periode tahun 2015 hingga 2023, dengan total sekitar 9.185 data poin yang tersebar di 15 kolom. Data ini mengandung berbagai atribut yang berkaitan dengan karakteristik atau fitur kendaraan. Tentu sebelum digunakan, dataset harus melewati prapemrosesan terlebih dahulu untuk memastikan data yang digunakan relevan dan dapat menghasilkan model prediksi yang baik.

Deskripsi Dataset		
Data	Variasi Data	Contoh Dataset
Fuel type	5	X = Bensin biasa; Z = Bensin premium; D = Diesel; E = E85, N = Gas Alam Untuk FFV
Cylinders	8	[ 4 6 8 12 10 16 3 5]
Year	9	[2023 2022 2021 2020 2015 2016 2017 2018 2019]
Vehicle class	15	['Full-size' 'Sport utility vehicle: Small' 'Sport utility vehicle: Standard' 'Compact' 'Mid-size' 'Minicompact' 'Two-seater' 'Subcompact' 'Station wagon: Small' 'Station wagon: Mid-size' 'Pickup truck: Small' 'Pickup truck: Standard' 'Minivan' 'Special purpose vehicle' 'Van: Passenger']
Engine size (L)	49	[1.5 3.5 3. 2. 2.9 4. 5.2 6. 4.4 8. 3.6 1.2 1.3 2.7 6.2 5.5 5.3 5.7 6.4 2.3 2.5 3.3 5. 1.6 3.8 5.6 2.4 4.6 3.4 6.7 1.8 6.6 1.4 2.8 3.2 6.5 3.7 4.3 1. 6.8 4.2 4.7 5.9 6.3 4.8 8.4 2.1 0.9 2.2]
Transmission	28	['AV7' 'M6' 'AS10' 'A8' 'A9' 'AM7' 'AS8' 'AM8' 'AV' 'AS9' 'A10' 'A6' 'M7' 'AV1' 'AM6' 'AS7' 'AV8' 'AV6' 'AS6' 'AV10' 'M5' 'AS5' 'A7' 'AM9' 'A5' 'A4' 'AS4' 'AM5']
Model	1978	['Integra' 'Integra A-SPEC' 'MDX SH-AWD' ... '1500 Classic 4X4 EcoDiesel' '1500 Classic 4X4 FFV' 'Corolla LE Eco']
Make	41	['Acura' 'Alfa Romeo' 'Aston Martin' 'Audi' 'Bentley' 'BMW' 'Bugatti' 'Buick' 'Cadillac' 'Chevrolet' 'Chrysler' 'Dodge' 'FIAT' 'Ford' ..... 'Volvo' 'Scion' 'smart']

Gambar 6. Deskripsi Variasi Dataset *Fuel Consumption Ratings*.

## 2. Pra-pemrosesan Data

Pra-pemrosesan data adalah tahap penting dalam analisis data yang bertujuan untuk mempersiapkan dataset agar siap digunakan dalam pembuatan model prediksi atau analisis lainnya. Langkah pertama dalam pra-pemrosesan adalah pembersihan data, yang mencakup identifikasi dan penghapusan data yang tidak relevan, duplikat, atau memiliki nilai yang hilang. Hal ini penting untuk memastikan bahwa dataset yang digunakan akurat, konsisten, dan dapat memberikan hasil yang valid. Selanjutnya, dilakukan transformasi data, seperti normalisasi atau standarisasi nilai, serta konversi format data, untuk memastikan bahwa data berada dalam bentuk yang sesuai dengan kebutuhan analisis. Selain itu, seleksi fitur (*feature selection*) juga dilakukan untuk memilih fitur yang relevan dan menghapus fitur yang tidak memberikan kontribusi berarti terhadap model, guna menghindari *overfitting*.

Terakhir, penanganan data tidak lengkap dilakukan dengan cara menghapus atau mengimputasi nilai yang hilang, agar tidak mengganggu analisis lebih lanjut. Tujuan utama dari pra-pemrosesan data adalah untuk meningkatkan kualitas dataset dan memaksimalkan efektivitas model yang akan dibangun, dengan memastikan bahwa data yang digunakan bebas dari gangguan yang dapat menurunkan kualitas hasil analisis.

a. Pembersihan Data

Proses pembersihan data dilakukan untuk memastikan bahwa dataset yang digunakan dalam penelitian bebas dari duplikasi, kesalahan, dan data yang tidak relevan. Sebagai contoh, data *MAKE* dan *MODEL* pada Gambar 6, tidak akan digunakan karena keduanya berpotensi membatasi penelitian. Data tersebut terlalu spesifik pada jenis kendaraan tertentu, yang dapat menyebabkan model menjadi kurang generalisasi dalam memprediksi konsumsi bahan bakar untuk berbagai jenis kendaraan.

b. *Encoding* Fitur Kategorikal

Fitur kategorikal perlu diubah menjadi format numerik agar dapat digunakan dalam model *machine learning*. Metode encoding yang digunakan adalah *One-Hot Encoding* yang mengubah fitur kategorikal menjadi format biner dengan membuat kolom baru untuk setiap kategori. Setiap kategori diwakili oleh nilai 1 atau 0 pada kolom yang sesuai. Teknik ini digunakan untuk fitur seperti '*VEHICLE CLASS*', '*TRANSMISSION*', dan '*FUEL\_NAME*'.

c. Penanganan Data Tidak Lengkap

Nilai-nilai yang hilang dalam dataset dapat memberikan dampak negatif terhadap kualitas dan keakuratan model yang dibangun. Ketika terdapat data yang hilang, model dapat menghadapi kesulitan dalam mempelajari pola yang ada, dan dapat menyebabkan keputusan yang bias atau bahkan kesalahan dalam prediksi. Oleh karena itu, penting untuk melakukan penanganan yang tepat terhadap nilai yang hilang sebelum data digunakan dalam analisis

lebih lanjut, dikarenakan pada tahun 2015 data *SMOG RATING* dan data *CO2 RATING* tidak diukur dan pada tahun 2016 juga data *SMOG RATING* tidak diukur seperti yang tampak pada Gambar 7. Dengan kondisi ini, dalam penelitian ini data 2015 dan 2016 tidak digunakan untuk menjaga konsistensi fitur dalam dataset sehingga semua data poin memiliki atribut yang lengkap dan relevan dengan variabel yang digunakan dalam penelitian. Dengan demikian, dataset yang digunakan adalah dataset dari tahun 2017-2022, yaitu sebanyak 6951 data poin. Penyesuaian ini memastikan bahwa model yang dibangun tidak dipengaruhi oleh ketidakseimbangan data akibat nilai-nilai yang hilang pada periode tertentu.

Dataset total data poin keseluruhan (pertahun) :				
Model year	Smog rating	CO2 rating	Make, Model, Vehicle class, Engine size (L), Cylinders, Transmission, Fuel type, City (L/100 km), Highway (L/100 km), Combined (L/100 km), Combined (mpg), CO2 emissions (g/km)	Total
2015	0	0	1128	2234
2016	0	1106	1106	
2017	1058	1058	1058	6951
2018	1083	1083	1083	
2019	1056	1056	1056	
2020	975	975	975	
2021	970	970	970	
2022	976	976	976	
2023	833	833	833	
Total				9185

Gambar 7. Total Dataset *Fuel Consumption Ratings*.

Untuk memastikan bahwa tahapan *preprocessing* data dilakukan secara lebih rapi dan terstruktur, sebaiknya seluruh proses dijalankan melalui pipeline. Pipeline ini memungkinkan urutan proses *preprocessing* untuk dieksekusi secara otomatis dan berurutan, serta memastikan bahwa setiap langkah dilakukan dengan konsisten di seluruh data.

### 3. Pembuatan Pipeline

Pembuatan pipeline bertujuan untuk mengotomasi langkah-langkah dalam proses analisis dan pembuatan model, sehingga dapat menghindari kesalahan manual dan meningkatkan efisiensi. Pipeline ini mencakup tahap-tahap berikut:

- a. *Preprocessing Data*: Melakukan normalisasi dan pembersihan data.
- b. *Encoding Data* : merubah data kategorikal menjadi numerik
- c. Pelatihan Model: Menerapkan algoritma *Support Vector Regression (SVR)* dan *Random Forest Regression (RFR)*.
- d. Evaluasi Model: Menggunakan data pengujian untuk mengevaluasi kinerja model.

Pembuatan pipeline ini memungkinkan untuk mengubah data mentah menjadi input model secara otomatis dan efisien, serta melakukan evaluasi secara berulang dengan parameter yang berbeda. Dalam penelitian ini, pipeline dibangun menggunakan pustaka *Scikit-learn*, yang memungkinkan integrasi langkah-langkah *preprocessing*, pelatihan, dan evaluasi dalam satu objek yang mudah digunakan.

### 4. Pemilihan Fitur dan Target

Pemilihan fitur adalah langkah penting dalam proses pembangunan model prediksi. Tujuannya adalah untuk mengidentifikasi fitur-fitur yang paling relevan yang memiliki pengaruh signifikan terhadap variabel target, serta mengeliminasi fitur yang kurang penting atau *redundant*. Dengan demikian, model yang dihasilkan akan lebih sederhana, lebih mudah diinterpretasikan, dan memiliki performa yang lebih baik, selain itu dikarenakan penelitian ini disesuaikan dengan PT. Telkomsat yang sedang mengembangkan sistem serupa fitur yang digunakan disesuaikan dengan kebutuhan, fitur yang akan

digunakan antara lain *YEAR*, *VEHICLE CLASS*, *ENGINE SIZE*, *CYLINDERS*, *FUEL\_NAME*, *TRANSMISSION*.

Fitur yang digunakan juga mempertimbangkan penilaian dengan analisis korelasi dengan menggunakan *Corelation Matrix*.

Matriks korelasi dapat dinyatakan dalam bentuk nilai antara -1 hingga 1, di mana:

- a. Nilai positif menunjukkan hubungan searah antara fitur dan target, artinya saat nilai fitur meningkat, nilai target juga cenderung meningkat.
- b. Nilai negatif menunjukkan hubungan berlawanan arah, artinya saat nilai fitur meningkat, nilai target cenderung menurun.
- c. Nilai mendekati 0 menunjukkan bahwa tidak ada hubungan linier yang signifikan antara fitur dan target.

Dalam konteks penelitian ini, matriks korelasi akan digunakan untuk mengidentifikasi fitur-fitur yang memiliki korelasi tinggi dengan variabel target yaitu *CITY (L/100 KM)*, *HWY (L/100 KM)*, *COMB (L/100 KM)*, *COMB (MPG)*, *EMISSIONS*, *CO2 RATING*, dan *SMOG RATING*. Hasil dari analisis ini membantu dalam pemilihan fitur yang relevan dan menghindari multikolinearitas dalam model. Visualisasi korelasi dalam bentuk heatmap juga akan digunakan untuk memperjelas hubungan antar fitur. Kemudian 80% data akan dijadikan data pelatihan dan 20% akan dijadikan data pengujian.

## 5. Optimasi *hyperparameter*

Optimasi *Hyperparameter* dilakukan untuk mengoptimalkan kinerja *Hyperparameter*, pada kedua algoritma (*SVR dan RFR*) parameter yang tidak dipelajari oleh model selama pelatihan, melainkan harus ditentukan sebelumnya. Pemilihan *hyperparameter* yang tepat sangat mempengaruhi performa model. Proses *tuning* ini dilakukan dengan menggunakan *Grid*

*Search*, yang memilih kombinasi parameter secara acak dari berbagai parameter yang ditentukan.

*Tuning* parameter *SVR* bertujuan untuk menemukan kombinasi *C*, kernel, *gamma*, dan *epsilon* yang memberikan hasil terbaik dalam memprediksi konsumsi bahan bakar kendaraan. *Support Vector Regression* memiliki beberapa parameter penting yang dapat memengaruhi kinerja model. Proses *tuning* dilakukan untuk menemukan kombinasi parameter terbaik yang menghasilkan model dengan performa optimal. Berikut adalah parameter utama yang di-*tuning* pada *SVR*:

a. *C (Regularization Parameter)*

Parameter *C* mengontrol keseimbangan antara kesalahan pada data training dan margin keputusan yang dibuat oleh model. Nilai *C* yang lebih tinggi akan memprioritaskan akurasi pada data training tetapi berisiko menyebabkan *overfitting*. Sebaliknya, nilai *C* yang terlalu rendah dapat menyebabkan *underfitting*.

b. *Kernel*

*SVR* menggunakan fungsi kernel untuk memetakan data ke dimensi yang lebih tinggi agar data mendekati linier pada dimensi yang baru sehingga dapat lebih mudah dipisahkan dengan *separating boundary* yang linier. *Kernel* yang digunakan dalam *tuning* meliputi Linear Kernel yang cocok untuk data yang dapat dipisahkan secara linier. *Radial Basis Function (RBF) Kernel* yang digunakan untuk data yang tidak dapat dipisahkan secara linier. *Polynomial Kernel* yang cocok untuk data yang memiliki pola non-linear dengan derajat tertentu.

c. *Gamma ()*

Parameter *gamma* digunakan dalam kernel *RBF* dan *Polynomial* untuk menentukan seberapa jauh pengaruh satu titik data terhadap model. Nilai *gamma* yang lebih besar membuat model lebih kompleks dan sensitif terhadap perubahan kecil pada data,

sementara nilai gamma yang terlalu kecil membuat model kurang fleksibel.

d. *Epsilon* ( $\epsilon$ )

Parameter epsilon menentukan margin di sekitar fungsi prediksi di mana tidak ada penalti yang diberikan untuk kesalahan. Nilai epsilon yang lebih besar mengurangi sensitivitas model terhadap data dan dapat membantu mengurangi *overfitting*.

Sedangkan *tuning* parameter *RFR* bertujuan untuk menyesuaikan jumlah pohon, kedalaman pohon, dan jumlah sampel yang digunakan pada setiap node agar menghasilkan model yang akurat dan tidak *overfitting*. *Random Forest Regression* adalah algoritma berbasis pohon keputusan yang menggunakan banyak pohon untuk membuat prediksi. Proses *tuning* pada *RFR* bertujuan untuk mengoptimalkan kinerja model dengan menyesuaikan parameter-parameter berikut:

a. *n\_estimators*

Parameter *n\_estimators* menentukan jumlah *Tree* dalam *Forest*. Semakin banyak pohon yang digunakan, semakin baik kinerja model, tetapi hal ini juga meningkatkan waktu komputasi. Proses *tuning* bertujuan untuk menemukan jumlah pohon yang optimal agar model tidak terlalu kompleks namun tetap akurat.

b. *max\_depth*

Parameter *max\_depth* membatasi kedalaman maksimum setiap *Tree* dalam *Forest*. Kedalaman yang terlalu besar dapat menyebabkan model *overfitting*, sementara kedalaman yang terlalu kecil dapat menyebabkan *underfitting*.

c. *min\_samples\_split*

Parameter *min\_samples\_split* menentukan jumlah minimum sampel yang diperlukan untuk membagi node dalam pohon. Nilai yang lebih besar mengurangi kompleksitas model dan membantu mencegah *overfitting*.

d. *min\_samples\_leaf*

Parameter *min\_samples\_leaf* menentukan jumlah minimum sampel yang diperlukan pada setiap daun pohon. Pengaturan ini meningkatkan generalisasi model dengan mencegah pembentukan daun yang hanya terdiri dari sedikit sampel.

e. *max\_features*

Parameter *max\_features* menentukan jumlah maksimum fitur yang dipertimbangkan untuk membagi node dalam pohon. Nilai yang lebih kecil meningkatkan kecepatan model tetapi dapat mengurangi akurasi. Proses tuning bertujuan untuk menemukan keseimbangan antara kecepatan komputasi dan akurasi model.

f. *bootstrap*

Parameter *bootstrap* menentukan apakah sampel diambil dengan penggantian saat membangun pohon. Mengaktifkan *bootstrap* umumnya meningkatkan keakuratan model.

## 6. Evaluasi Kinerja Model

Untuk mengevaluasi kinerja model yang dibangun, digunakan tiga metrik evaluasi utama:

a. *Mean Absolute Percentage Error (MAPE)*:

*MAPE* mengukur rata-rata persentase kesalahan absolut antara nilai yang diprediksi dan nilai aktual dari data pengujian. Metrik ini digunakan untuk melihat seberapa besar kesalahan prediksi dalam bentuk persentase terhadap nilai aktual. Semakin rendah nilai *MAPE*, semakin baik kinerja model.

b. *Root Mean Square Error (RMSE)*:

*RMSE* mengukur akar dari rata-rata kuadrat kesalahan prediksi. Metrik ini memberikan gambaran tentang besarnya kesalahan prediksi, di mana semakin rendah nilai *RMSE*, semakin baik kinerja model.

c. R-squared ( $R^2$ ):

$R^2$  mengukur seberapa baik model dapat menjelaskan variasi dalam data. Nilai  $R^2$  yang tinggi menunjukkan bahwa model dapat menjelaskan sebagian besar variasi dalam data yang ada.

## 7. Pengujian dengan Data Baru

Setelah proses pelatihan dan *tuning* model selesai, model akan diuji menggunakan data baru yang tidak terlibat dalam proses pelatihan sebelumnya. Tujuan pengujian dengan data baru adalah untuk menilai kemampuan model dalam memprediksi konsumsi bahan bakar kendaraan pada data yang belum pernah dilihat sebelumnya, sehingga dapat mengukur sejauh mana model dapat menggeneralisasi informasi.

Data baru ini diproses dengan cara yang sama seperti data pelatihan, termasuk langkah preprocessing seperti normalisasi dan pemilihan fitur yang relevan. Model yang sudah terlatih diuji terhadap data baru ini, dan hasilnya dibandingkan dengan nilai aktual untuk mengukur kinerja model. Data baru yang digunakan adalah dataset yang baru dikeluarkan tahun 2023-2025, <https://open.canada.ca/data/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64/resource/ec44c747-a640-4481-bcf0-a2ea7ca8f249>.

### 3.3 Perangkat Penelitian

Pada penelitian ini, perangkat keras yang digunakan adalah Laptop *Razer Blade* dengan spesifikasi sebagai berikut: *Processor Intel Core i5* (generasi terbaru), *RAM* 16 GB, *Storage* 512 GB SSD, dan Sistem Operasi *Windows* 11. Perangkat lunak yang digunakan meliputi *Google Colaboratory* untuk pengembangan dan pelatihan model *Random Forest Regression* serta *Support Vector Regression*, *Python* 3.9 sebagai bahasa pemrograman utama, dan beberapa library seperti *Scikit-learn*, *Pandas*, *Matplotlib* dan *Seaborn*, *NumPy*, dll, untuk menunjang pembuatan model prediksi konsumsi bahan bakar kendaraan.

## **BAB V**

### **KESIMPULAN DAN SARAN**

#### **5.1 Kesimpulan**

Berdasarkan hasil penelitian dan analisis yang telah dilakukan terhadap performa model *Support Vector Regression (SVR)* dan *Random Forest Regression (RFR)* dalam memprediksi berbagai target variabel, dapat disimpulkan beberapa poin berikut:

1. Performa Model Secara Umum

*Random Forest Regression (RFR)* secara konsisten menunjukkan performa yang lebih baik dibandingkan *Support Vector Regression (SVR)* pada target dengan pola data yang kompleks atau fluktuasi yang tinggi. Hal ini tercermin dari nilai *Mean Absolute Error (MAE)*, *Root Mean Square Error (RMSE)*, dan *R-squared ( $R^2$ )* yang lebih baik pada sebagian besar target variabel.

2. Target dengan Performa Terendah

Target *EMISSIONS* memiliki tingkat kesalahan prediksi tertinggi dibandingkan variabel lainnya, dengan nilai  $R^2$  yang lebih rendah akibat pola fluktuasi data yang kompleks. Hal ini menegaskan perlunya eksplorasi model atau pendekatan alternatif untuk meningkatkan akurasi prediksi. Dalam konteks rumusan masalah kedua, rendahnya  $R^2$  menunjukkan bahwa prediksi kedua model masih perlu ditingkatkan melalui analisis lebih lanjut terhadap pola data. Selain itu, hasil ini mengungkap keterbatasan kedua algoritma dalam menangani data dengan kompleksitas tinggi, sebagaimana yang dikaji dalam rumusan masalah ketiga terkait evaluasi kelebihan dan kekurangan model.

### 3. Keunggulan dan Kekurangan Model

*RFR* unggul dalam menangani target variabel seperti *COMB\_MPG*, *EMISSIONS*, dan *SMOG*, yang memiliki tingkat kompleksitas tinggi. Model ini mampu memberikan prediksi yang lebih stabil dan akurat meskipun pada data dengan pola yang tidak teratur. *SVR* menunjukkan performa yang baik pada target dengan hubungan yang lebih sederhana, seperti *CITY* dan *HWY*. Namun, model ini cenderung kurang stabil saat menangani target dengan variabilitas tinggi.

## 5.2 Saran

Berdasarkan hasil penelitian ini, beberapa saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut:

### 1. Eksplorasi Model Alternatif

Untuk meningkatkan akurasi prediksi pada target dengan performa rendah, seperti *EMISSIONS*, disarankan untuk mengeksplorasi model alternatif, seperti Deep Learning atau metode ensemble, yang lebih adaptif dalam menangani pola data yang kompleks.

### 2. Implementasi Model Prediksi

Model prediksi yang telah menunjukkan performa baik, seperti prediksi konsumsi bahan bakar kendaraan di perkotaan (*CITY*), jalan raya (*HWY*), dan gabungannya (*COMB*), dapat diimplementasikan dalam aplikasi nyata untuk mendukung pengambilan keputusan yang lebih akurat.

### 3. Penggunaan Dataset yang Lebih Besar

Untuk meningkatkan generalisasi model, penelitian selanjutnya dapat menggunakan dataset yang lebih besar dan lebih beragam sehingga pola data yang dipelajari menjadi lebih representatif dan akurat.

Dengan mengikuti saran-saran ini, diharapkan penelitian mendatang dapat menghasilkan model prediksi yang lebih akurat dan efektif dalam berbagai konteks aplikasi.

## DAFTAR PUSTAKA

- Adi, A. C. (2024, 22 Mei). Indonesia tekankan urgensi dekarbonisasi sektor transportasi. Kementerian Energi dan Sumber Daya Mineral Republik Indonesia. Diakses pada 7 Januari 2025 dari <https://www.esdm.go.id>
- Alayida, N. F., Aisyah, T., Deliana, R., & Diva, K. (2023). Pengaruh digitalisasi di era 4.0 terhadap para tenaga kerja di bidang logistik. *Jurnal Economina*, 2(1), 254-268.
- Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
- Altrinaldo, A., & Abu, R. (2024). Analisis tingkat emisi gas buang, konsumsi bahan bakar dan kinerja mesin sepeda motor 4 langkah menggunakan bahan bakar pertamax dan campuran pertamax etanol. *Ekasakti Engineering Journal*, 4(1), 1-12.
- Bintang, H. (2023, 20 Februari). Dekarbonisasi sektor transportasi: Elektrifikasi dan kendaraan listrik di Indonesia. Institute for Essential Services Reform (IESR). Diakses pada 7 Januari 2025 dari <https://iesr.or.id>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Elsa, R. (2023). Penerapan metode support vector regression (SVR) menggunakan kernel linear, polinomial, dan radial dengan *grid search optimization*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

- Li, Y., Zeng, I. Y., Niu, Z., Shi, J., Wang, Z., & Guan, Z. (2022). *Predicting vehicle fuel consumption based on multi-view deep neural network. Neurocomputing, 502, 140-147.*
- Maesaroh, S. (2024). Pembelajaran mesin (Machine Learning). Dalam Pembelajaran mesin dan kecerdasan buatan: Teori dan aplikasi praktis (hlm. 5-10). Banten: PT Sada Kurnia Pustaka.
- Mitchell, T. M. (1997). *Machine learning*. McGraw-Hill.
- Novianti, F., Ulinuha, N., Hafiyusholeh, M., & Arianto, A. (2022). Prediksi penggunaan bahan bakar pada PLTGU menggunakan metode support vector regression (SVR). *J. Techno. COM, 21(2), 249-255.*
- Pangesty, S., Budiharjo, A., & Rusmandani, P. (2021). Pengaruh kecepatan kendaraan terhadap konsumsi bahan bakar minyak di jalan tol. *Siklus: Jurnal Teknik Sipil, 7(1), 1-8.*
- Penalun, F. E., Hermawan, A., & Avianto, D. (2023). Perbandingan random forest regression dan support vector regression pada prediksi laju penguapan. *Jurnal Fasilkom, 13(2), 104-111.*
- Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning, 1(1), 81-106.*
- Setiawan, S., & Marlinda, L. (2024). Perbandingan efisiensi bahan bakar pada kendaraan low MPV dengan metode COPRAS. *Infotek: Jurnal Informatika dan Teknologi, 7(2), 421-431.*
- Sharma, P., & Sharma, A. K. (2021). Application of response surface methodology for optimization of fuel injection parameters of a dual fuel engine fuelled with producer gas-biodiesel blends. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 1-18.*

- Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, 14(3), 199-222.
- Syahrir, I., Priyono, M. D., & Batutah, M. A. (2024). Analisis perbandingan performa bahan bakar Shell Super dan Shell V-Power pada motor Honda PCX 150 cc tahun 2021. *Journal of Manufacturing in Industrial Engineering & Technology*, 3(1), 24-32.
- Tualeka, N. A. A. Z. (2024). Perbandingan metode *random forest regression (RFR)* dan *support vector regression (SVR)* dalam memprediksi risiko kredit pada Bank XYZ (Disertasi Doktoral). Surabaya: Institut Teknologi Sepuluh Nopember.
- Vapnik, V., & Cortes, C. (1995). Support-vector networks. *Machine learning*, 20, 273-297.
- Xu, X., Yu, H., Sun, Q., & Tam, V. W. (2023). A critical review of occupant energy consumption behavior in buildings: How we got here, where we are, and where we are headed. *Renewable and Sustainable Energy Reviews*, 182, 113396.
- Yang, Y., Gong, N., Xie, K., & Liu, Q. (2022). *Predicting gasoline vehicle fuel consumption in energy and environmental impact based on machine learning and multidimensional big data*. *Energies*, 15(5), 1602.
- Yao, Z., Wang, Y., Liu, B., Zhao, B., & Jiang, Y. (2021). Fuel consumption and transportation emissions evaluation of mixed traffic flow with connected automated vehicles and human-driven vehicles on expressway. *Energy*, 230, 120766.
- Zahra, R. A. (2023). Penerapan algoritma random forest dengan *hyperparameter tuning* untuk memprediksi harga sewa kost di Kota Bandar Lampung.