Coding: *Journal of Computing and Software Engineering*
Vol. 1, No. 2, Desember 2025, Hal: 90-96
E-ISSN: 3109-869X
Doi: https://doi.org/10.25181/coding.v1i2.4379
Journal homepage: https://jurnal.polinela.ac.id/coding

Perbandingan Kinerja Model *Machine Learning* dalam Memprediksi Konsumsi Bahan Bakar Kendaraan

Affan Prasetya[1] , Oki Arifin[2]

[1]Program Studi Magister Terapan Ketahanan Pangan, Politeknik Negeri Lampung, Indonesia.
[2]Program Studi Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Lampung, Indonesia.

Artikel Info

Kata Kunci:
Decision Tree;
Konsumsi Bahan Bakar;
Machine learning;
Random Forest;
Support Vector Machine.

Keywords:
Decision Tree;
Fuel Consumption;
Machine Learning;
Random Forest;
Support Vector Machine.

Riwayat Artikel:

Submitted: 10 Agustus 2025
Accepted: 30 November 2025
Published: 31 Desember 2025

Abstrak: Penelitian ini membahas upaya prediksi konsumsi bahan bakar kendaraan berbasis algoritma *machine learning* untuk mendukung efisiensi energi dan pengurangan emisi karbon. Sektor transportasi merupakan kontributor signifikan terhadap konsumsi energi nasional, sehingga diperlukan model prediksi akurat yang dapat digunakan dalam perencanaan kebijakan. Penelitian ini bertujuan membandingkan kinerja tiga algoritma regresi *Decision Tree* (DT), *Random Forest* (RF), dan *Support Vector Machine* (SVM) dalam memprediksi konsumsi bahan bakar berdasarkan fitur teknis kendaraan. Dataset yang digunakan bersumber dari *Natural Resources Canada*, terdiri dari variabel seperti kapasitas mesin, silinder, transmisi, emisi CO_2, dan konsumsi bahan bakar. Proses pra-pemrosesan mencakup *one-hot encoding*, normalisasi, dan pembagian data (*hold-out* 80:20). Evaluasi dilakukan menggunakan metrik RMSE, MAE, dan koefisien determinasi (R^2). Hasil menunjukkan bahwa model *Decision Tree* memiliki performa terbaik dengan nilai RMSE terendah (0.326), MAE terendah (0.22), dan R^2 tertinggi (0.960), mengungguli *Random Forest* dan SVR. Temuan ini menunjukkan bahwa model sederhana namun tepat guna dapat lebih unggul dari model kompleks dalam kondisi dataset tertentu. Penelitian ini memberikan kontribusi terhadap pengembangan sistem prediktif efisiensi energi kendaraan dan dapat digunakan sebagai dasar untuk kebijakan transportasi berkelanjutan.

Abstract: *This study investigates the application of machine learning algorithms to predict vehicle fuel consumption as part of efforts to improve energy efficiency and reduce carbon emissions. The transportation sector is a major contributor to national energy usage, necessitating accurate predictive models for policy planning and energy management. This research compares the performance of three regression algorithms Decision Tree (DT), Random Forest (RF), and Support Vector Machine (SVM) in predicting fuel consumption based on vehicle technical features. The dataset used is sourced from Natural Resources Canada, consisting of variables such as engine size, number of cylinders, transmission type, CO_2 emissions, and fuel consumption. Preprocessing steps include one-hot encoding, normalization, and data splitting (80% training, 20% testing using hold-out validation). Model performance is evaluated using RMSE, MAE, and the coefficient of determination (R^2). Results show that the Decision Tree model outperforms the others, achieving the lowest RMSE (0.326), lowest MAE (0.22), and highest R^2 (0.960). These findings indicate that a well-suited, simple model can outperform more complex algorithms when aligned with the dataset's characteristics. This research contributes to the*

development of predictive systems for vehicle energy efficiency and offers insights for sustainable transportation policy planning.

Corresponding Author:
Affan Prasetya
Email: affanparsetya6@gmmail.com

PENDAHULUAN

Konsumsi bahan bakar kendaraan merupakan salah satu indikator utama dalam analisis efisiensi energi dan emisi karbon suatu negara. Sektor transportasi, khususnya kendaraan bermotor pribadi, menjadi salah satu kontributor terbesar terhadap konsumsi energi final dan emisi gas rumah kaca (GHG), termasuk karbon dioksida (CO_2). Menurut Kementerian Energi dan Sumber Daya Mineral, (2023), konsumsi energi di sektor transportasi di Indonesia mencapai lebih dari 40% dari total konsumsi energi nasional, dengan sebagian besar berasal dari bahan bakar fosil.

Perencanaan kebijakan energi nasional dan upaya mitigasi perubahan iklim sangat membutuhkan prediksi yang akurat terhadap pola konsumsi bahan bakar. Dalam konteks ini, pendekatan berbasis data seperti *machine learning* (ML) menjadi alat yang menjanjikan. *Machine learning* dapat mengidentifikasi pola kompleks dalam data kendaraan dan perilaku konsumsi bahan bakar yang tidak mudah ditangkap oleh model statistik konvensional (Yoo *et al.*, 2025).

Berbagai algoritma *machine learning* telah digunakan untuk memprediksi konsumsi bahan bakar kendaraan, di antaranya *Decision Tree (DT), Random Forest (RF),* dan *Support Vector Machine (SVM). Random Forest* sering digunakan karena kemampuannya dalam menangani data *non-linear* dan *overfitting* yang rendah (Hassan *et al.,* 2023). Sementara itu, SVR dikenal memiliki akurasi yang baik dalam regresi dengan data berdimensi tinggi dan jumlah sampel terbatas (Yuli Artini et al., 2024). *Decision Tree*, meskipun sederhana, tetap populer karena interpretabilitasnya yang tinggi dan kemudahan dalam visualisasi model (Jia et al., 2023).

Namun, performa dari setiap algoritma sangat bergantung pada karakteristik dataset yang digunakan, seperti jumlah fitur, tipe data (numerik atau kategorikal), dan korelasi antarfitur. Oleh karena itu, penting dilakukan evaluasi komparatif untuk menentukan algoritma mana yang paling optimal untuk digunakan dalam konteks data tertentu (Hamed & Khafagy, 2021; Amenhar, 2025).

Penelitian ini bertujuan untuk membandingkan performa tiga algoritma *machine learning* yaitu *Decision Tree, Random Forest, dan Support Vector Regression* dalam memprediksi konsumsi bahan bakar kendaraan berbasis data teknis kendaraan seperti kapasitas mesin, berat kendaraan, tipe bahan bakar, dan lain-lain. Dengan melakukan evaluasi menggunakan metrik seperti *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), dan koefisien determinasi (R^2), hasil penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan model prediktif efisien untuk mendukung kebijakan energi dan transportasi berkelanjutan.

METODE

Metode yang digunakan dalam penelitian ini dilaksanakan dengan cara dimulai dari pengumpulan dataset terlebih dahulu, kemudian data tersebut diproses untuk mendapatkan hasil, setelah hasil ditemukan, tahap selanjutnya dilakukan evaluasi terhadap hasil yang didapat. Semua proses tersebut dapat terlihat pada penjelasan berikut:

2.1. Data

Penelitian ini menggunakan dataset sekunder yang bersifat terbuka dan dapat diakses publik, yaitu data kendaraan dari *Natural Resources Canada (NRCan) (Dataset* ini tersedia dalam format CSV dan dapat diakses melalui laman resmi pemerintah Kanada *https://open.canada.ca/data/en/dataset/0867f3cf-7c10-49b2-a4b3-8584c701b3a5)*. *Dataset* ini mencakup data teknis kendaraan ringan yang diproduksi pada rentang tahun 2000 dan telah digunakan secara luas dalam studi terkait konsumsi bahan bakar dan emisi gas rumah kaca oleh kendaraan (Yoo *et al.,* 2025). Data ini dipilih karena kelengkapannya, akurasi pengukuran dari lembaga resmi, serta karena variabel-variabelnya mencerminkan faktor-faktor utama yang memengaruhi konsumsi bahan bakar kendaraan.

Adapun variabel-variabel yang digunakan dalam penelitian ini terdiri dari lima fitur, yaitu *Engine Size (L)* yang merepresentasikan kapasitas mesin dalam liter, *Cylinders* yang menunjukkan jumlah silinder dalam mesin, *Transmission* yang merupakan tipe transmisi kendaraan (otomatis/manual dan variannya), *Fuel Consumption (L/100 km)* sebagai target variabel yang menunjukkan tingkat konsumsi bahan bakar dalam liter per 100 kilometer, serta CO_2 *Emissions (g/km)* sebagai indikator emisi karbon yang dikeluarkan kendaraan. Meskipun CO_2 bukan variabel target, ia tetap disertakan sebagai fitur tambahan karena memiliki korelasi tinggi dengan konsumsi bahan bakar, sebagaimana ditemukan dalam studi-studi sebelumnya (Hien & Ah-Lian, 2022).

2.2. Pra-pemrosesan Data

Sebelum digunakan untuk pelatihan model pembelajaran mesin, data terlebih dahulu melalui proses pra-pemrosesan. Proses ini bertujuan untuk mengubah data mentah menjadi bentuk yang sesuai dan optimal bagi algoritma yang akan digunakan. Beberapa langkah pra-pemrosesan yang dilakukan dalam penelitian ini adalah sebagai berikut:

Pertama, dilakukan *one-hot encoding* terhadap variabel kategorikal *Transmission*. Teknik ini digunakan untuk mengubah kategori menjadi representasi numerik biner tanpa memberikan bobot ordinal, sehingga menghindari asumsi hubungan matematis antar kategori. Langkah ini sangat penting dalam pembelajaran mesin karena sebagian besar algoritma hanya dapat memproses data numerik (Ostermann *et al.,* 2024).

Kedua, semua fitur numerik seperti *Engine Size, Cylinders*, dan CO_2 *Emissions* dinormalisasi ke dalam rentang nilai 0 hingga 1 menggunakan teknik *Min-Max Scaling*. Normalisasi ini penting untuk mencegah dominasi fitur tertentu akibat skala yang berbeda, serta mempercepat konvergensi dalam algoritma yang peka terhadap skala seperti *Support Vector Machine*. Selanjutnya, *dataset* dibagi menjadi dua bagian utama, yaitu data pelatihan (training data) sebesar 80% dan data pengujian (testing data) sebesar 20%. Pembagian ini dilakukan secara acak, namun stratifikasi dipertahankan agar distribusi target tetap representatif. Teknik ini mengacu pada metode *hold-out validation* yang umum digunakan dalam pembelajaran mesin untuk mengukur generalisasi model terhadap data yang belum pernah dilihat (Haoxing & System, n.d.).

2.3. Arsitektur Eksperimen

Penelitian ini menggunakan pendekatan eksperimental dengan membandingkan tiga algoritma regresi yang populer dalam prediksi berbasis data teknis kendaraan, yaitu *Decision Tree (DT), Random Forest (RF),* dan *Support Vector Machine (SVM)*. Ketiga model ini dibangun dan diimplementasikan menggunakan perangkat lunak *RapidMiner Studio (Altair AI Studio Educational)*, sebuah platform visual berbasis *workflow* yang mendukung pemodelan data sains tanpa harus menulis kode secara manual.

2.4. Evaluasi Model

Evaluasi performa model dilakukan dengan membandingkan nilai aktual dari data uji terhadap prediksi model. Karena eksperimen menggunakan metode *hold-out validation*, hasil evaluasi berasal dari satu kali pembagian data (tanpa validasi silang).

Doi: http://doi.org/10.26594/coding.v1i2.4379

1. *Mean Absolute Error* (MAE)

MAE mengukur rata-rata dari selisih absolut antara nilai aktual dan prediksi. MAE bersifat linier terhadap kesalahan dan mudah diinterpretasikan.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|$$

2. *Root Mean Squared Error* (RMSE)

RMSE mengukur rata-rata kuadrat selisih antara nilai aktual dan prediksi, lalu diakarkan. RMSE memberikan penalti lebih tinggi terhadap kesalahan besar.

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

3. *Coefficient of Determination* (R^2 *Score*)

R^2 digunakan untuk mengukur proporsi variansi dari variabel target yang dapat dijelaskan oleh model. Nilai R^2 berkisar antara 0 hingga 1, di mana nilai mendekati 1 menunjukkan bahwa model menjelaskan sebagian besar variasi target.

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

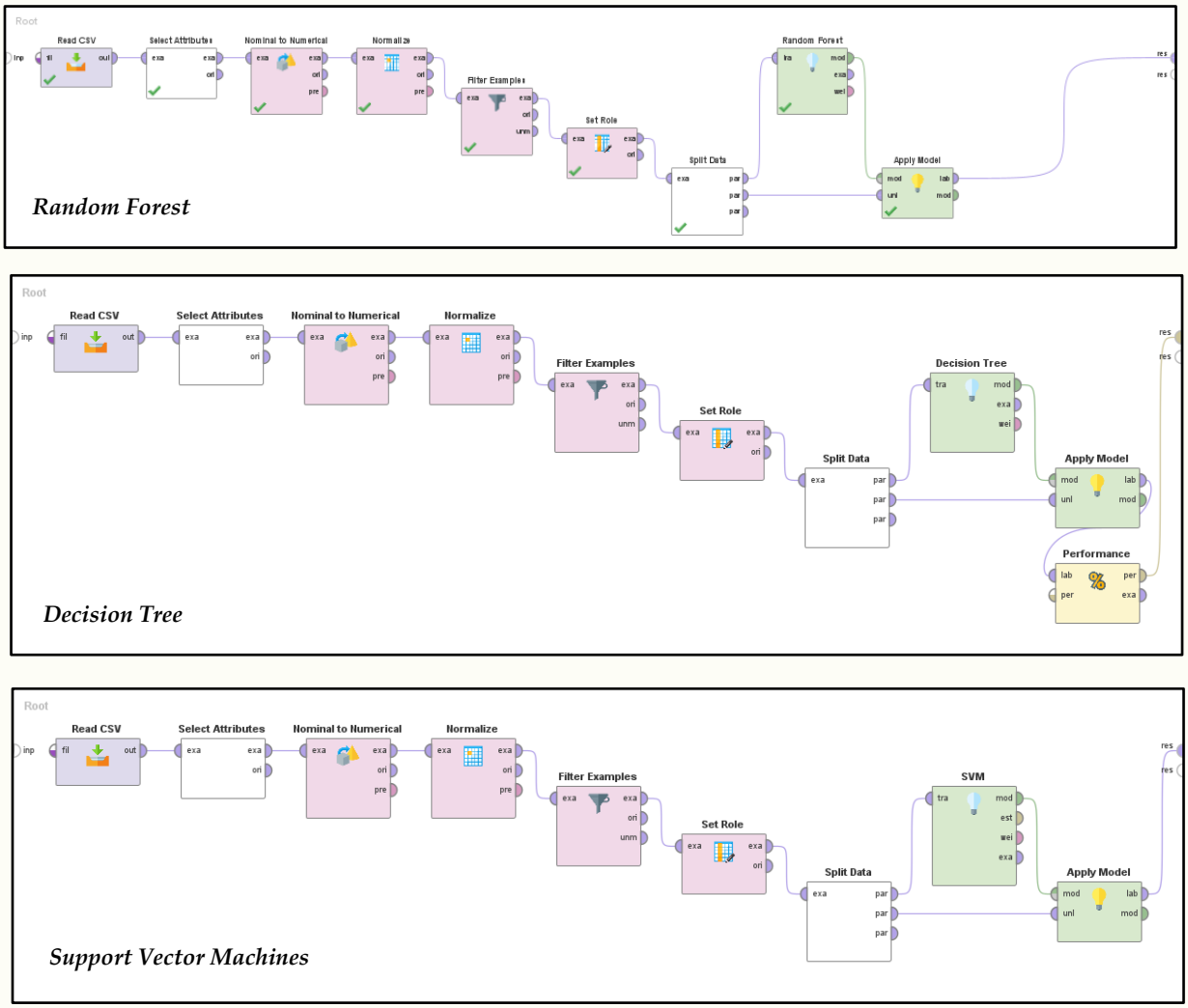
Keterangan:
y_i adalah nilai aktual
$\hat{y}_i$ adalah nilai prediksi
$\bar{y}_i$ adalah nilai rata-rata dari target

Evaluasi dilakukan dengan menambahkan operator *Performance (Regression)* setelah *Apply Model*. Operator ini secara otomatis menghitung nilai MAE, RMSE, dan R^2 berdasarkan hasil prediksi dan nilai aktual dari data uji. Karena model menggunakan pendekatan *hold-out validation*, evaluasi hanya dilakukan satu kali berdasarkan hasil dari subset pengujian. Meskipun validasi silang (seperti *k-fold CV)* dapat memberikan estimasi yang lebih stabil, pendekatan *hold-out* tetap valid dan sering digunakan sebagai langkah awal untuk menilai kelayakan model (Haoxing & System, 2023).

HASIL DAN PEMBAHASAN

3.1 Tampilan Proses di *RapidMiner*

Eksperimen dilakukan menggunakan *RapidMiner Studio*, sebuah perangkat lunak pemodelan analitik berbasis visual yang memungkinkan pengguna membangun alur *machine learning* tanpa menulis kode pemrograman. Pada Gambar 1 berikut ditampilkan alur proses pemodelan menggunakan algoritma *Random Forest*, *Decision Tree*, dan *Support Vector Machines*.



Gambar 1. Alur Proses Permodelan di *RapidMiner*

3.2 Hasil Evaluasi Model

Dalam penelitian ini, dilakukan eksperimen terhadap tiga algoritma *machine learning* yaitu *Random Forest*, *Decision Tree*, dan *Support Vector Machine* (SVM) untuk memprediksi konsumsi bahan bakar kendaraan berdasarkan fitur teknis seperti ukuran mesin, jumlah silinder, jenis transmisi, dan emisi CO_2. Evaluasi model dilakukan menggunakan metrik regresi yaitu *Root Mean Squared Error (RMSE), Mean Absolute Error (MAE),* dan *Koefisien Determinasi* (R^2). Hasil lengkap dari eksperimen ditampilkan dalam Tabel 1 berikut:

Tabel 1. Hasil Evaluasi Model Regresi

Model	**RMSE**	**MAE**	**R^2**
Random Forest	0.533	0.32	0.908
Decision Tree	0.326	0.22	0.960
Support Vector Machines	0.512	0.33	0.901

Root Mean Square Error (RMSE) memberikan penalti yang lebih besar terhadap kesalahan prediksi yang besar, karena perhitungan didasarkan pada kuadrat dari selisih prediksi dan nilai aktual. Dari hasil di atas, *Decision Tree* memperoleh nilai RMSE terendah (0.326), menunjukkan bahwa secara statistik model ini paling stabil dalam memprediksi konsumsi bahan bakar dengan kesalahan rata-rata

Doi: http://doi.org/10.26594/coding.v1i2.4379

paling kecil. Temuan ini sejalan dengan studi oleh (Yuli Artini et al., 2024) menunjukkan bahwa model *Decision Tree* yang dioptimasi mampu mengungguli *Random Forest* dalam kasus *dataset* berskala kecil hingga menengah yang memiliki variabel input dengan korelasi tinggi, seperti pada data teknis kendaraan.

Mean Absolute Error (MAE) menunjukkan rata-rata kesalahan absolut antara prediksi dan nilai aktual, tanpa memperhitungkan arah kesalahan. Kembali, *Decision Tree* unggul dengan MAE sebesar 0.22, lebih kecil dibandingkan dengan *Random Forest* (0.32) dan SVM (0.33). Nilai MAE yang rendah penting karena memberikan interpretasi langsung dalam konteks satuan liter/100 km. Artinya, kesalahan rata-rata prediksi *Decision Tree* hanya ± 0.22 liter per 100 km, angka yang sangat baik untuk aplikasi nyata seperti perencanaan efisiensi energi atau uji emisi kendaraan. Menurut Mahi-al-rashid *et al.,* (2022), MAE adalah metrik yang paling informatif untuk sistem berbasis regulasi seperti konsumsi energi, karena nilai absolut kesalahan dapat langsung digunakan dalam evaluasi kebijakan atau penyesuaian target teknis kendaraan.

Nilai R^2 mengindikasikan seberapa baik model menjelaskan variansi pada data target. Semakin mendekati 1, semakin baik model dalam menjelaskan hubungan antara *input* dan *output*. *Decision Tree* memperoleh R^2 sebesar 0.960, yang berarti 96% variansi dalam konsumsi bahan bakar dapat dijelaskan oleh model. Sebaliknya, *Random Forest* dan SVM mencatat nilai R^2 masing-masing sebesar 0.908 dan 0.901. Meskipun tetap dalam kategori baik (di atas 0.9), nilai tersebut menunjukkan bahwa kedua model sedikit lebih rendah dalam kemampuan generalisasi dibandingkan dengan *Decision Tree* dalam konteks dataset ini.

Menurut Mao & Cao (2024) sebuah *Decision Tree* yang dioptimasi dengan teknik gradient-based (oblique regression tree) berhasil mengungguli *Random Forest* sebanyak 2.03 % pada rata-rata akurasi di 16 dataset. Selain itu, penelitian oleh Wang & Yao (2025), memperkenalkan *FoLDTree*, sebuah pohon tunggal dengan struktur oblique dan pemilihan fitur efisien, yang menunjukkan performa yang setara bahkan melebihi *Random Forest* pada dataset nyata. Temuan ini mendukung prinsip bahwa dalam *dataset* berskala kecil hingga menengah dengan fitur berkorelasi tinggi, pemilihan arsitektur bukan kompleksitas semata menentukan performa model.

KESIMPULAN

Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma *machine learning* dapat digunakan secara efektif untuk memprediksi konsumsi bahan bakar kendaraan berbasis data teknis. Di antara tiga algoritma yang diuji, *yaitu Decision Tree, Random Forest*, dan *Support Vector Machine*, model *Decision Tree* menunjukkan performa terbaik dengan nilai RMSE dan MAE yang paling rendah serta nilai koefisien determinasi tertinggi. Hasil ini menunjukkan bahwa meskipun *Decision Tree* merupakan model yang relatif sederhana, ia mampu memberikan akurasi tinggi dalam konteks *dataset* yang digunakan, yang memiliki struktur teratur dan fitur berkorelasi kuat.

Implikasi dari temuan ini menunjukkan bahwa pemilihan model prediktif tidak selalu harus bergantung pada kompleksitas algoritma, tetapi lebih pada kesesuaian karakteristik data dengan pendekatan yang digunakan. Selain itu, hasil penelitian ini memberikan kontribusi praktis dalam pengembangan sistem pemantauan efisiensi bahan bakar kendaraan serta dalam penyusunan kebijakan berbasis data untuk sektor transportasi. Penelitian selanjutnya disarankan untuk mengintegrasikan teknik tuning parameter dan validasi silang, serta mengeksplorasi penggunaan ensemble model adaptif atau arsitektur neural network ringan yang kompatibel dengan implementasi real-time pada perangkat otomotif.

DAFTAR PUSTAKA

Amenhar, H. (2025). Comparative Analysis of Regression Models for Vehicle Fuel Consumption Prediction. *Https://Www.Researchgate.Net/*.

HAMED, M., & Khafagy, M. (2021). Fuel Consumption Prediction Model using Machine Learning. *International Journal of Advanced Computer Science and Applications*, *12*. https://doi.org/10.14569/IJACSA.2021.0121146

Haoxing, Z., & System, C. (2023). Model selection with bootstrap validation. *Willey*. https://doi.org/10.1002/sam.11606

Hassan, M. A., Salem, H., Bailek, N., & Kisi, O. (2023). Random Forest Ensemble-Based Predictions of On-Road Vehicular Emissions and Fuel Consumption in Developing Urban Areas. *Sustainability*, *15*(2). https://doi.org/10.3390/su15021503

Hien, N. L. H., & Ah-Lian, K. (2022). Analysis and prediction model of fuel consumption and carbon dioxide emissions of light-duty vehicles. *Applied Sciences*, *12*(2), 803.

Jia, Q., Li, X., Yu, L., Bian, J., Zhao, P., Li, S., Xiong, H., & Dou, D. (2023). Learning from Training Dynamics: Identifying Mislabeled Data beyond Manually Designed Features. *Proceedings of the 37th AAAI Conference on Artificial Intelligence, AAAI 2023*, *37*, 8041–8049. https://doi.org/10.1609/aaai.v37i7.25972

Kementerian Energi dan Sumber Daya Mineral. (2023). Handbook of Energy & Economy Statistics of Indonesia. In *Kementerian Energi dan Sumber Daya Mineral*.

Mahi-al-rashid, A., Hossain, F., Anwar, A., & Azam, S. (2022). False Data Injection Attack Detection in Smart Grid Using Energy Consumption Forecasting. *Energies*, *15*(13). https://doi.org/10.3390/en15134877

Mao, Q., & Cao, Y. (2024). Can A Single Tree Outperform An Entire For- Est? *ArXiv Preprint ArXiv:2410.23147*, 1–20.

Ostermann, A., Bajrami, A., & Bogensperger, A. (2024). Short - term forecasting of German generation - based - CO 2 emission factors using parametric and non - parametric time series models. *Energy Informatics*. https://doi.org/10.1186/s42162-024-00303-9

Wang, S., & Yao, K. (2025). FoLDTree : A ULDA-Based Decision Tree Framework for Efficient Oblique Splits and Feature Selection. *ArXiv Preprint ArXiv:2410.23147*, *1*, 1–19.

Yoo, S.-R., Shin, J.-W., & Choi, S.-H. (2025). Machine learning vehicle fuel efficiency prediction. *Scientific Reports*, *15*(1), 14815. https://doi.org/10.1038/s41598-025-96999-0

Yuli Artini, N. P. S., Sumarjaya, I. W., & Nilakusmawati, D. P. E. (2024). Penerapan Metode Support Vector Regression (Svr) Dengan Algoritma Grid Search Dalam Peramalan Harga Saham. *E-Jurnal Matematika*, *13*(2), 94. https://doi.org/10.24843/mtk.2024.v13.i02.p447

Doi: http://doi.org/10.26594/coding.v1i2.4379