Integra: Journal of Integrated Mathematics and Computer Science 1(2) 60-67

Integra: Journal of Integrated Mathematics and Computer Science

Journal Homepage: https://integrajimcs.com/index.php/integra/index

integra

Journal of Integrated Mathematics and Computer Science

Research Paper

Comparison of Support Vector Regression and Random Forest Regression Performance in Vehicle Fuel Consumption Prediction

Muhaymi Nurdin[1], Wamiliana[2], Akmal Junaidi[1*], Favorisen Rossyking Lumbanraja[1]

[1] Department of Computer Science, Faculty of Mathematics and Natural Science, Universitas Lampung, Lampung, 35145, Indonesia
[2] Department of Mathematics, Faculty of Mathematics and Natural Science, Universitas Lampung, Lampung, 35145, Indonesia
[*] **Corresponding author**: akmal.junaidi@fmipa.unila.ac.id

Keywords

Grid Search, Fuel Consumption, Machine Learning, Random Forest Regression, Support Vector Regression

Abstract

Predicting vehicle fuel consumption is an important aspect in improving energy efficiency and supporting sustainable transportation. This study aims to compare the performance of Support Vector Regression (SVR) and Random Forest Regression (RFR) algorithms in predicting combined vehicle fuel consumption (COMBINED, a combination of 55% urban and 45% highway). The Canadian government's Fuel Consumption Ratings dataset was used, with 2015-2023 data (9,185 entries) for training and testing, and 2024 data (764 entries) for further testing. Pre-processing involved StandardScaler for numerical features and OneHotEncoder for categorical features, followed by hyperparameter optimization using Grid Search, resulting in optimal parameters: SVR (C=100, epsilon=0.5, gamma=1) and RFR (n_estimators=200, max_depth=None, min_samples_split=2). Results show RFR is superior with R^2 0.8845, RMSE 0.9671, and MAE 0.6566, compared to SVR with R^2 0.8648, RMSE 1.0462, and MAE 0.7150. Evaluation on 2024 data and visualization of error distribution corroborate the superiority of RFR. This study concludes RFR is more effective for COMBINED prediction, although SVR is competitive post-optimization, and contributes to the selection of machine learning models for green vehicle technology.

Received: 5 May 2024, Accepted: 14 June 2024
https://doi.org/10.26554/integra.20241221

1. INTRODUCTION

Climate change in recent decades has posed a huge threat to humanity's future. In recent years, consumption-based carbon emissions have dominated studies on environmental challenges and international trade [1]. In the era of globalization, environmental issues and energy efficiency have become major pressing concerns around the world. The overuse of fossil fuels has triggered an increase in air pollution, greenhouse gas emissions, and accelerated the rate of climate change, with the transportation sector being a significant contributor to these issues. Carbon dioxide (CO_2) is a naturally occurring greenhouse gas, meaning it traps heat in the Earth's atmosphere. While essential for life on Earth, increased CO_2 levels due to human activities lead to global warming and climate change, with potential impacts on human health, ecosystems, and the environment. There are many researches regarding the effect of CO_2 emission for human life including [2, 3, 4, 5] and many more.

Carbon emissions have enormous influence on Indonesia's ecology and economy. While Indonesia has committed to lowering emissions and made some progress, there are still issues in policy implementation and enforcement, notably in matching emission reduction policies with economic development and energy security.

This condition emphasizes the need for serious efforts to improve vehicle fuel use efficiency through government policies and technological innovations in the automotive industry. The Indonesian government has taken steps such as encouraging the use of biofuels and energy efficiency technologies [6], while in the logistics sector, some companies have started to develop prediction systems to anticipate fuel theft by drivers, and optimize fuel usage management [7]. Accurate prediction systems are key to helping companies and regulators monitor fuel consumption, reduce operational costs, and mitigate environmental impacts, thus supporting the national sustainability agenda. Not only do carbon emissions have a huge impact on the environment, they also affect the Indonesian economy. The impact of CO_2 emissions and exports on Indonesian economic growth had been investigated by [8].

Fuel usage varies substantially according to vehicle type. In general, smaller and lighter vehicles, such as compact cars and hybrids, have higher fuel efficiency than bigger vehicles like trucks and SUVs. Diesel engines are often more fuel efficient than gasoline engines. Vehicle fuel consumption is an important aspect of the automotive industry that is closely related to energy efficiency and environmental impact. With the increasing global awareness to reduce greenhouse gas emissions, research focusing on fuel consumption prediction is increasingly relevant to support the development of environmentally friendly technologies. Precise predictions allow vehicle manufacturers to design more efficient engines and help users optimally manage fuel consumption [9].

Fuel consumption is influenced by a variety of complex factors, including traffic conditions, environmental factors, vehicle specifications (such as engine and transmission type), and driver behavior [10]. This complexity demands the development of predictive models capable of handling non-linear data and the relationships between variables that influence each other often pose a challenge to simple statistical approaches such as linear regression. In recent years, machine learning has emerged as a promising approach to modeling such phenomena, offering flexibility in capturing patterns that are difficult to analyze manually. Several studies provide prediction and strategies to use fuel more efficiently, including [11, 12, 13], and others.

Machine learning algorithms are collections of instructions that allow computers to learn from data and improve their performance on certain tasks without requiring explicit programming. They are simply mathematical models that analyse data to detect patterns, generate predictions, and classify it. These algorithms are broadly classified as supervised, unsupervised, and reinforcement learning, each having unique properties and uses. Sarker [14] describes the fundamentals of numerous machine learning techniques and their relevance in various real-world application fields, such as healthcare, e-commerce, agriculture, cybersecurity systems, and many more.

Two popular machine learning algorithms for regression tasks are Support Vector Regression (SVR) and Random Forest Regression (RFR). SVR, derived from Support Vector Machine, is designed to find the optimal hyperplane that minimizes prediction error in a transformed feature space, often using a Radial Basis Function (RBF) kernel [15].

The advantage of SVR lies in its ability to handle non-linear data, although its performance is highly dependent on hyperparameter settings such as C (regularization), epsilon (tolerance margin), and gamma (kernel parameters). In contrast, RFR is an ensemble method that combines results from multiple decision trees to improve accuracy and reduce the risk of overfitting [16].

RFR excels at managing data with noise and categorical features, and provides insights through feature importance, but requires optimization of hyperparameters such as the number of trees (n_estimators) and tree depth (max_depth). The selection and optimization of these two algorithms is key to achieving accurate predictions, especially in the context of fuel consumption involving multiple variables.

The Fuel Consumption Ratings dataset from the Canadian government forms the main basis of this research, providing comprehensive data on the fuel consumption of vehicles, both private and commercial. The dataset includes variables such as engine type, engine capacity, fuel type, transmission, vehicle weight, exhaust emissions, as well as technical information such as year, model and manufacturer. This diversity of data enables comparative analysis between vehicles with different specifications, providing a clear picture of the factors that influence variations in fuel consumption [17].

Although originating from Canada, this dataset is relevant for application in Indonesia given similar vehicle characteristics, such as comparable fuel use and similar technical specifications. Adapting this prediction model can support more efficient and environmentally friendly transportation policy planning, including in the context of logistics in Indonesia.

The literature review shows that previous studies have compared SVR and RFR in various contexts. Tualeka [18] found that SVR with RBF kernel excels in credit risk prediction, with a Mean Absolute Percentage Error (MAPE) of 11.63% and Mean Squared Error (MSE) of 0.2486, compared to RFR which tends to overfitting on test data. In contrast, Penalun et al. [19] reported that RFR was more accurate in predicting evaporation rate, with an R^2 of 0.81 and an RMSE of 0.53, compared to SVR with an R^2 of 0.56 and an RMSE of 0.81. However, specific research on predicting vehicle fuel consumption using these two algorithms is still limited, especially with a focus on combined fuel consumption in 55% urban areas and 45% inter-city highways.

The importance of this research lies in two aspects. Theoretically, this research expands the understanding of the performance of SVR and RFR in modeling nonlinear data, with a focus on improving accuracy through hyperparameter optimization using Grid Search. This is in line with the conclusion put forward by Yang et al in [20], which shows that the utilization of multidimensional data-based machine learning models effectively improves the accuracy of vehicle fuel consumption prediction. The results can be used by the automotive industry to design more efficient vehicles and by the logistics sector, to detect fuel theft by drivers and optimize operations. With a focus on the fuel consumption of combined city streets and inter-city highways, this research provides practical insights for everyday vehicle usage conditions.

This study aims to analyse and compare the performance of SVR and RFR in predicting vehicle fuel consumption using Fuel Consumption Ratings dataset, with Grid Search optimization, validation on recent data, and error distribution analysis. The results are expected to support the development of environmen-

Table 1. Description of Dataset Variables

Data	Data Type	Description
Transmission	Mixed	Covers transmission type and gear count: A = Automatic, AM = Automated Manual, AS = Automatic with Selective Shift, AV = Continuously Variable, M = Manual; includes number of gears.
Fuel type	Categorical	Type of fuel used: X = Regular gasoline; Z = Premium gasoline; D = Diesel; E = E85; N = Natural Gas. For Flexible Fuel Vehicles (FFV), fuel consumption values are provided for both gasoline and E85.
Make	Categorical	Vehicle brand.
Model	Categorical	Vehicle model.
Vehicle class	Categorical	Vehicle class.
Year	Numeric	Year the vehicle was manufactured.
Engine size (L)	Numeric	Engine size in litres.
Cylinders	Numeric	Number of engine cylinders.
CO_2 rating	Numeric	CO_2 emission rating, rated on a scale from 1 (worst) to 10 (best).
Smog rating	Numeric	Smog rating, rated on a scale from 1 (worst) to 10 (best).
CO_2 emissions (g/km)	Numeric	CO_2 emissions in grams per km.
Highway (L/100 km)	Numeric	Fuel consumption on highway roads in litres per 100 km.
City (L/100 km)	Numeric	Fuel consumption on city roads in litres per 100 km, includes stop-and-go traffic.
Combined (mpg)	Numeric	Combined fuel consumption in miles per gallon.
Combined (L/100 km)	Numeric	Combined fuel consumption in litres per 100 km (55% city and 45% highway driving).

tally friendly vehicle technology and sustainable transportation policies.

2. METHODS

This study compares the performance of Support Vector Regression (SVR) and Random Forest Regression (RFR) to predict the combined (COMBINED) fuel consumption, calculated as 55% CITY + 45% HIGHWAY. The approach uses machine learning with Grid Search optimisation. The Fuel Consumption Ratings dataset from the Canadian government was used, with 9,185 entries (2015-2023) for training and initial testing, and 764 entries (2024) for further testing. Figure 1 shows the flowchart of the research.

Table 2. Tuned Parameters and Optimal Value of SVR Model

Parameter	Tested Value	Optimal Value
C	0.1, 1, 10, 100	100
Epsilon	0.01, 0.1, 0.2, 0.5	0.5
Gamma	'scale', 'auto', 0.1, 1	1

The process includes data collection, pre-processing, model building, optimisation, evaluation, model saving, retesting, and error analysis. The research was conducted using a laptop with specifications (Intel Core i5, 16 GB RAM, 512 GB SSD, Windows 11) and the model computing process was carried out using Google Collaboratory, Python 3.9, and libraries scikit-learn, pandas, matplotlib, seaborn, numpy.

2.1 Dataset and Data Source

The Fuel Consumption Ratings dataset is taken from the official website of the Canadian government, containing 9,949 entries (2015-2024). The data is divided into 9,185 entries (2015-2023) for initial training/testing and 764 entries (2024) for advanced testing. Table 1 displays the dataset description of fuel consumption ratings.

2.2 Data Pre-Processing

The 2015-2016 data was removed because the SMOG RATING and CO_2 RATING attributes were empty, leaving 6,951 entries (2017-2023) and 764 entries (2024). Categorical features were converted with OneHotEncoder (drop='first') to prevent duplication of information.

Numerical features were normalised using StandardScaler. The 2017-2023 data was split 80% training (5,561 entries) and 20% testing (1,390 entries) using train_test_split with random_state=42.

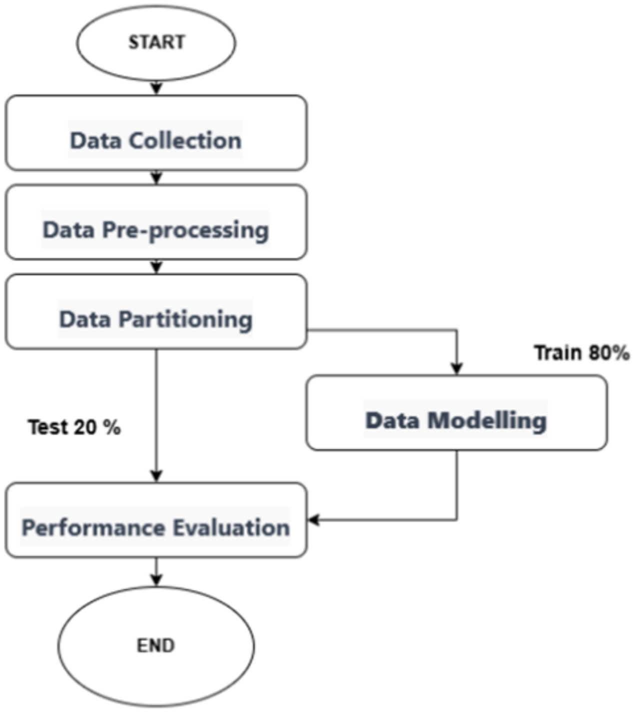

Figure 1. The Flowchart of the Research

2.3 Model Building

The prediction model was built using the scikit-learn library, with a Support Vector Regression (SVR) algorithm implementing a Radial Basis Function (RBF) kernel and parameters C = 1, epsilon = 0.1, and gamma = 'scale' to accommodate non-linear data characteristics. Meanwhile, the Random Forest Regression (RFR) model is configured with n_estimators = 100, max_depth = None, and min_samples_split = 2 to produce a flexible model that can handle data complexity.

The entire data transformation process is performed consistently through a scikit-learn pipeline that integrates StandardScaler for numerical features and OneHotEncoder for categorical features, ensuring uniform data during both training and testing.

2.4 Hyperparameter Optimisation

Hyperparameter optimisation is performed to improve model performance in the training and prediction process. The two models used in this study, namely Support Vector Regression (SVR) and Random Forest Regression (RFR), each require parameter adjustments in order to produce an optimal prediction model. The optimisation process is performed using the Grid Search technique, which evaluates a combination of parameter values thoroughly based on the highest accuracy score obtained from cross-validation.

In SVR, tests were conducted on three main parameters, namely the value of C (regulation), epsilon (error threshold allowed in the margin), and gamma (radial basis function kernel coefficient). Table 2 shows the tuned parameters and optimal value of SVR Model, while Table 3 displays the range of values tested and the best configuration.

Table 3. Tuned Parameters and Optimal Value of RFR Model

Parameter	Tested Value	Optimal Value
n_estimators	50, 100, 200	200
max_depth	None, 10, 20	None
min_samples_split	2, 5, 10	2

Meanwhile, in the RFR model, three main parameters were explored, namely n_estimators (number of trees in the ensemble), max_depth (maximum depth of the tree), and min_samples_split (minimum number of samples to split internal nodes). Table 3 shows the results of parameter exploration for RFR.

From the test results, the optimal parameter combinations for SVR and RFR showed improved predictive performance on vehicle fuel consumption.

2.5 Model Evaluation

Model performance evaluation is conducted using three main metrics, namely the coefficient of determination (R^2), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These three metrics are commonly used measures to assess the accuracy of regression models.

The coefficient of determination R^2 measures the proportion of variance of the dependent variable that can be explained by the model. R^2 values range from 0 to 1, with higher values indicating a more accurate model. According to Chicco et al., the coefficient of determination is more informative than other error metrics in regression evaluation [21].

RMSE calculates the square root of the mean square of the difference between the actual and predicted values. This metric is sensitive to large errors and imposes a larger penalty on predictions that are far off the actual value, whereas MAE measures the average of the absolute value of errors, and is more stable against outliers because all errors are weighted equally [22].

In this study, the model was trained and tested using 5,561 data entries from the period 2017 to 2023, while 1,390 entries were used as test data. The selection of these three metrics was done to provide a thorough evaluation of the regression model performance.

The model with the best parameters was saved using joblib, and then tested again using 764 data entries for 2024. The prediction error was calculated as the difference between the predicted value of the saved model and the actual value of the COMBINED variable in the 2024 data, and then further analysed through bar plot visualisation to illustrate the error distribution in detail.

2.6 Data Analysis

Analysis includes correlation matrix for numerical features (ENGINE SIZE, CYLINDERS, CITY, HIGHWAY, COMBINED, EMISSIONS) using pandas corr() with Pearson method, visualised as heatmap with seaborn (v0.12.2), scale -1 to 1, annotation of correlation values. Error distribution visualised with matplotlib (v4.0.0) using a 30 bin histograms, displaying the average error for SVR and RFR on the 2024 data.

3. RESULTS AND DISCUSSION

Figure 2 displays the relationship between numerical features was analysed in this study using a correlation matrix with the Pearson method, visualised as a heatmap for ENGINE SIZE, CYLINDERS, CITY, HIGHWAY, COMBINED, EMISSIONS, CO_2 RATING, and SMOG RATING features.

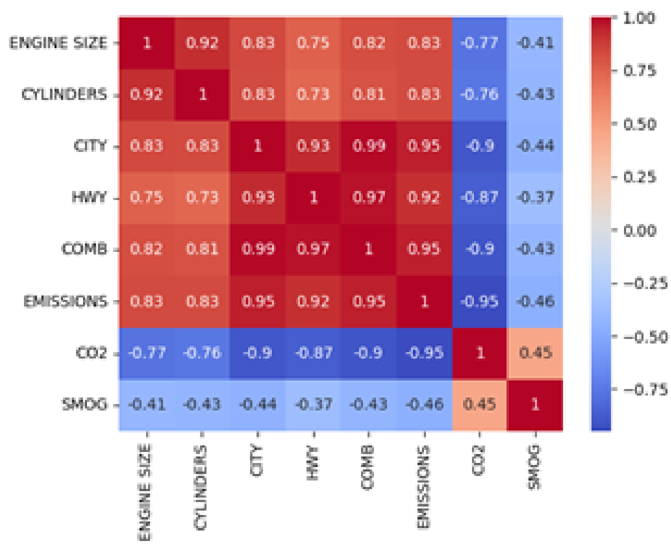

Figure 2. Correlation Matrix

Figure 2 shows that the highest correlation is between COMB and CITY (0.99), followed by COMB and HWY (0.97), and COMB and EMISSIONS (0.95).This confirms that combined fuel consumption is strongly influenced by consumption in urban areas and highways, in line with its definition as a composite of CITY (55%) and HWY (45%). The strong correlation with EMISSIONS suggests that increased fuel consumption has a direct impact on exhaust emissions, which is relevant in the context of emissions regulation and vehicle efficiency.

A high correlation was also observed between ENGINE SIZE and CYLINDERS (0.92), indicating multicollinearity between engine dimensions and cylinder configuration. Both contribute to increased fuel consumption, as vehicles with larger engines and more cylinders are generally high-performing but less efficient.

In contrast, negative correlations are shown by CO_2 to COMB (-0.95) and SMOG to COMB (-0.43). These negative values indicate that vehicles with higher fuel consumption tend to have worse CO_2 and SMOG ratings, meaning greater environmental impact. This consistent negative correlation is also reflected against other features such as CITY and HWY, indicating that the environmental rating is inversely proportional to fuel consumption.

These correlations suggest that features such as EMISSIONS, ENGINE SIZE, and CYLINDERS are highly influential in the prediction of fuel consumption, and are important indicators in the evaluation of vehicle performance and environmental impact. In addition, the detected correlation pattern supports the importance of careful feature selection to avoid redundancies and maintain the stability of the predictive model.

Table 4. Model Performance Before and After Tuning (Target: COMBINED)

Model	Status	R^2	RMSE	MAE	Main Parameter
SVR	before	0.8951	0.8953	0.6116	Default
SVR	after	0.9171	0.7958	0.5675	C=100, ε=0.5, γ=1
RFR	before	0.9341	0.7093	0.4823	Default
RFR	after	0.9340	0.7101	0.4829	n_estimators=200, max_depth=None

3.1 Comparison of Model Performance

The performance of Support Vector Regression (SVR) and Random Forest Regression (RFR) was evaluated against the COMBINED target, using the 2017-2023 period data (before and after tuning), as well as the 2024 data. Table 4 shows the evaluation results based on the R^2, RMSE, and MAE metrics for model performance before and after tuning, while Table 5 shows for model performance using year 2024 data.

Table 5. Model Performance using Year 2024 Data (Target: COMBINED)

Model	R^2	RMSE	MAE
SVR	0.8649	1.0462	0.7151
RFR	0.8845	0.9671	0.6566

From both tables, it can be concluded that RFR shows a more consistent and superior performance in predicting combined fuel consumption, both before and after tuning, as well as when tested on actual 2024 data. Meanwhile, although SVR improved after tuning, it still performed below RFR on the actual data.

The decrease in model performance on the 2024 data is due to the shift in data patterns, such as the increase in vehicle technology efficiency which makes the distribution and relationship between features change compared to the 2017-2023 training data.

SVR experienced a greater decline due to its reliance on the RBF kernel function, which tends to be less flexible in adjusting to new data variations. In contrast, RFR is more resilient due to its ensemble-based approach, so it is able to maintain performance even when the data at hand has different characteristics.

3.2 The Error Distribution

Figure 3 and Figure 4 show the distribution of SVR error and RFR of prediction data vs actual data for the COMBINED target. Figure 3 shows the bar plot of the SVR error with an average

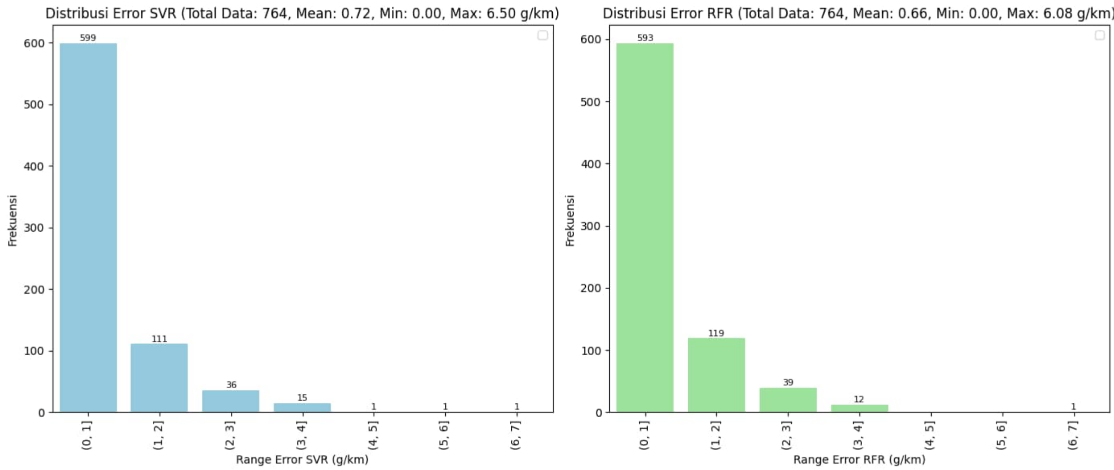

Figure 3. Error Distribution of Predicted vs Actual SVR and RFR (COMBINED)

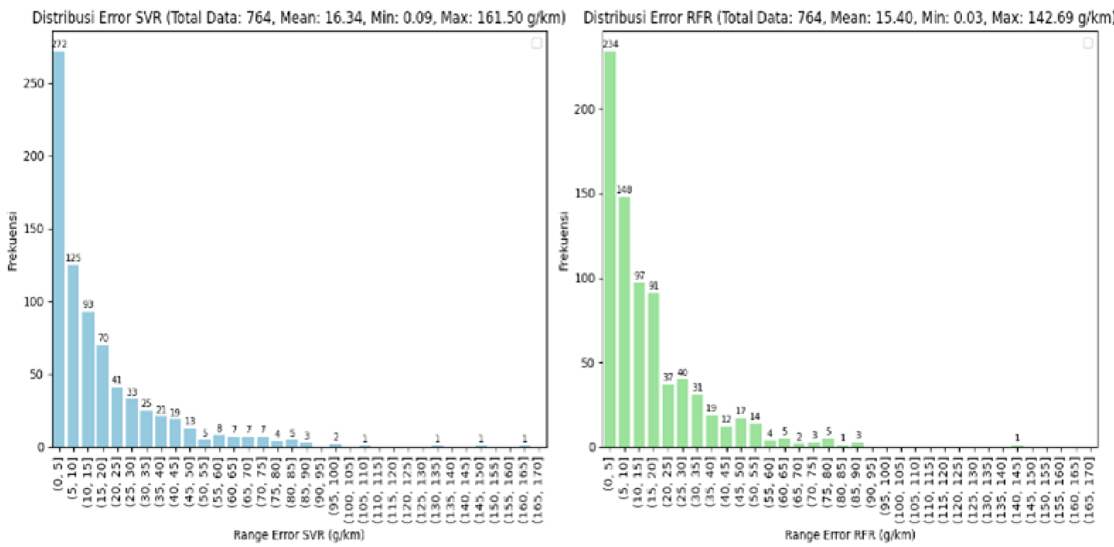

Figure 4. Error Distribution of Predicted vs Actual SVR and RFR (EMISSIONS)

error of 0.72, the graph shows there are 599 data that have a COMBINED target prediction error in the range of 0-1.

The RFR error bar plot with an average error of 0.66, shows there are 593 data that have a COMBINED target prediction error with a range of 0-1. RFR error with more data concentrated around zero explains the lower RMSE and MAE compared to SVR, indicating more accurate predictions.

In addition to creating a prediction model for combined city and inter-city highway fuel consumption, a prediction model for EMISSIONS was also created, the error distribution of SVR and RFR of predicted data vs actual data for target emissions in figure 4, was visualised to understand the error pattern.

Figure 4 shows the SVR error bar plot with an average error of 16.34, where there are 272 data that have target emissions prediction errors in the range of 0-5. Meanwhile, the RFR error bar plot has an average error of 15.40 with 234 data in the same error range.

Although most of the data is concentrated in the low range, the error distribution of the target emissions is wider than that of COMBINED, with higher maximum values. This suggests that predicting emissions is more difficult than COMBINED due to more complex variables and a wider range of values, resulting in generally higher errors.

3.3 Discussion of Key Findings

RFR consistently outperformed SVR in predicting COMBINED and EMISSIONS at all stages of evaluation, including before tuning, after tuning, and in testing the prediction model using 2024 data. RFR's advantage lies in its ensemble approach, which effectively handles non-linear relationships as well as relationships between features, such as ENGINE SIZE and CYLINDERS that are highly correlated (0.87).

Hyperparameter tuning with Grid Search had more impact on SVR, increasing the COMBINED R^2 from 0.8951 to 0.9171, as the optimal parameters (C=100, ε=0.5, γ=1) allowed SVR to capture non-linear patterns better. However, RFR remains more reliable with a stable R^2 at 0.9340-0.9341, demonstrating its stability against tuning. The drop in performance in the prediction model using the 2024 data (RFR: R^2 0.8845, SVR: R^2 0.8649) is due to the change in data patterns, which is not fully reflected in the 2017-2023 training data.

This decrease is more significant in SVR due to the dependency of the RBF kernel on the training data, whereas RFR is more adaptive. To improve performance, the training data needs to be updated with more recent and representative data, and additional hyperparameter exploration can be done to adapt the model to the dynamics of vehicle technology.

3.4 Implications and Application

More accurate RFR could be applied in the automotive industry to design fuel-efficient vehicles, in Indonesia's logistics sector where it could be useful to detect fuel theft, and for environmental policy to set fuel efficiency standards, supporting the 2050 zero emissions target.

4. CONCLUSIONS

This study concludes that Random Forest Regression (RFR) performs better than Support Vector Regression (SVR) in predicting vehicle fuel consumption, in line with the objective to compare the two models in handling fuel consumption data from the Fuel Consumption Ratings dataset. The findings demonstrate the effectiveness of the models in handling non-linear relationships between features, with RFR proving to be more reliable thanks to the ensemble approach that improves the stability of the predictions. Although both models face challenges on recent data, RFR remains consistently superior.

The contribution of this research lies in providing empirical evidence of RFR's superiority in predicting fuel consumption and emissions, while opening up opportunities to develop models that are more adaptive to data dynamics, and support innovation in the automotive, logistics and environmental policy sectors for global sustainability.

5. ACKNOWLEDGEMENT

The author would like to thank Software Engineering Lab, and Mathematics and Statistics Research Lab, Universitas Lampung for the support given.

REFERENCES

[1] Tonmoy Choudhury, Umar Nawaz Kayani, Azeem Gul, Syed Arslan Haider, and Sareer Ahmad. Carbon emissions, environmental distortions, and impact on growth. *Energy Economics*, 126:107040, 2023.

[2] Jiannan Wang and Waseem Azam. Natural resource scarcity, fossil fuel energy consumption, and total greenhouse gas emissions in top emitting countries. *Geoscience Frontiers*, 15(2):101757, 2024.

[3] Karen McKee. Potential effects of elevated atmospheric CO_2 on coastal wetlands. Technical report, Reston, VA, 2006.

[4] Leonel J. R. Nunes. The rising threat of atmospheric CO_2: A review on the causes, impacts, and mitigation strategies. *Environments*, 10(4):66, 2023.

[5] Frederica Perera. Pollution from fossil-fuel combustion is the leading environmental threat to global pediatric health and equity: Solutions exist. *International Journal of Environmental Research and Public Health*, 15(1):16, 2017.

[6] W. Saputra, A. Permatasari, M. Ichsan, and T. Syakira. Strategi pengembangan bahan bakar nabati melalui pemberdayaan perkebunan sawit rakyat di indonesia. *SPOS Indonesia*, pages 16–17, January 2022.

[7] G. M. Nguegnang. Predicting fuel consumption in power generation plants using machine learning and neural networks. *arXiv preprint*, page arXiv:2202.05591, 2022.

[8] Z. Yusuf, Wardhiah, G. Syamni, M. J. A. Siregar, and Y. A. Sitepu. The effect of CO_2 emissions in indonesia. *IOP Conference Series: Earth and Environmental Science*, 922:012026, 2021.

[9] N. F. Alayida, T. Aisyah, R. Deliana, and K. Diva. Pengaruh digitalisasi di era 4.0 terhadap para tenaga kerja di bidang logistik. *Jurnal Economina*, pages 254–268, 2023.

[10] D. S. Pangesty, A. Budiharjo, and P. Rusmandani. Pengaruh kecepatan kendaraan terhadap konsumsi bahan bakar minyak di jalan tol. *Siklus: Jurnal Teknik Sipil*, pages 1–8, 2021.

[11] So rin Yoo, Jae woo Shin, and Seoung Ho Choi. Machine learning vehicle fuel efficiency prediction. *Scientific Reports*, 15:14815, 2025.

[12] Sohaib Zahid and Umar Jamil. Data-driven machine learning techniques for fuel economy prediction in sustainable transportation systems. *Green Energy and Intelligent Transportation*, page 100303, 2025.

[13] R. F. K. Canal, G. Riffel, and G. Gracioli. Machine learning for real-time fuel consumption prediction and driving profile classification based on ecu data. *IEEE Access*, 12:68586–68600, 2024.

[14] Iqbal H. Sarker. Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2:160, 2021.

[15] A. J. Smola. A tutorial on support vector regression. *Statistics and Computing*, 14(3):199–222, 2024.

[16] L. Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001.

[17] Government of Canada. Fuel consumption ratings, 2015–2024. Natural Resources Canada.

[18] Nur Aisha Al Zahra Tualeka, Mohamad Atok, and Aimatul Ummah Alfajriyah. Perbandingan metode random forest regression (rfr) dan support vector regression (svr) dalam memprediksi risiko kredit pada bank xyz. *Jurnal Teknik ITS*, 13(6):540–547, 2024.

[19] F. E. Penalun, A. Hermawan, and D. Avianto. Perbandingan random forest regression dan support vector regression pada prediksi laju penguapan. *Jurnal Fasilkom*, 13(2):104–111, 2023.

[20] Y. Yang, N. Gong, K. Xie, and Q. Liu. Predicting gasoline vehicle fuel consumption in energy and environmental impact based on machine learning and multidimensional big data. *Energies*, 15(5):2–17, 2022.

[21] D. Chicco, M. J. Warrens, and G. Jurman. The coefficient of determination r-squared is more informative than smape, mae, mape, mse and rmse in regression analysis evaluation. *PeerJ Computer Science*, 7(3):e623, 2021.

[22] T. O. Hodson. Root-mean-square error (rmse) or mean absolute error (mae): When to use them or not. *Geoscientific Model Development*, 15(14):5481–5487, 2022.

