<u>*Araştırma Makalesi*</u>

www.ejosat.com ISSN:2148-2683

<u>*Research Article*</u>

Vehicle Fuel Emission Efficiency Estimation Using Multi-Linear Regression in Machine Learning

Yunus Egi[1*2]

[1*] Electrical and Electronics Engineering Department, Engineering Faculty, Şırnak University, Şırnak ,Turkey, (ORCID: 0000-0001-5185-8443), yunusegi@sirnak.edu.tr
[2] American University of Middle East, Collage of Engineering, Electrical Engineering Departmetment, Egaila, Kuwait, (ORCID: : 0000-0001-5185-8443), yunusegi@gmail.com

Abstract

Vehicle fuel consumption and emission have been a great deal for global warming and the world economy. The impacts of CO_2 emission can be minimized through optimization of engine design parameters such as Rated horsepower(RHP), Number of Cylinders and Rotors(NCR), Number of Gears(NG), and Equivalent Test Weight(ETW) to the Rounded and Adjusted Fuel Economy (RAFE). This article explorers the weighted impact of the independent variables RHP, NCR, NG, and ETW to RAFE using Multi-Linear Regression(MLR) in machine learning. For the proposed MLR method, the vehicle data is divided into two as training and testing. Then, the data cleanup process was applied to the training data to eliminate outliers that led to incorrect predictions. The proposed method determines the correlation coefficient to compare and seek the variables having less relationships with the dependent variable RAFE. Since there are no insignificant parameters in correlation analysis, MLR training was carried out by taking into account all parameters. Finally, the processed data are trained to create a multi-linear regression model. The obtained model is evaluated through Analysis of Variance(ANOVA). According to the ANOVA, there is a significant relationship between the dependent variable RAFE and the independent variables NG, ETW, and RHP with a p-value of $4.0994e^{-60}$, $1.5887e^{-48}$, and $2.5494e^{-31}$, respectively. Moreover, p-values of NG, ETW, and RHP are supported with F-test results of 227.73, 220.87, and 152.41. On the other hand, the obtained model is also relatively less affected by NCR, with a p-value of 0.031276 and an F-test of 4.94. As a result, the resulting MLR model can be used in new vehicle designs as it reveals which vehicle parameters affect CO_2 emissions.

Keywords: Fuel efficiency, Machine Learning, Multi-linear Regression, CO2 Emission.

Öğreniminde Çok Doğrusal Regresyon Kullanarak Araç Yakıt Emisyon Verimliliği Tahmini

Öz

Araç yakıt tüketimi ve emisyonu küresel ısınma ve dünya ekonomisi için büyük bir olay olmuştur. CO_2 emisyonunun etkileri, Nominal Beygir Gücü (NBG), Silindir ve Rotor Sayısı (SRS), Dişli Sayısı (DS) ve Eşdeğer Test Ağırlığı (ETA) gibi motor tasarım parametrelerinin Yuvarlatılmış ve Ayarlanmış Yakıt Ekonomisine (YAYE) optimizasyonu ile en aza indirilebilir. Bu makalede, makine öğreniminde Çoklu Doğrusal Regresyon (ÇDR) kullanılarak NBG, SRS, DS ve ETA bağımsız değişkenlerinin YAYE'ye ağırlıklı etkisi ortaya çıkarılmıştır. Önerilen ÇDR yöntemi için araç verileri eğitim ve test olarak ikiye ayrılmıştır. Daha sonra yanlış tahminlere yol açan aykırı değerleri ortadan kaldırmak için eğitim verilerine veri temizleme işlemi uygulanmıştır. Önerilen yöntem, bağımlı değişken YAYE ile ilişkisi olan değişkenleri karşılaştırmak ve aramak için korelasyon katsayısını belirlemektedir. Korelasyon analizinde önemsiz parametreler bulunmadığından ÇDR eğitimi tüm parametreler dikkate alınarak gerçekleştirilmiştir. Son olarak, işlenen veriler ÇDR modeli oluşturmak için eğitilmiştir. Elde edilen model Varyans (ANOVA) analizi ile değerlendirilmiştir. ANOVA'ya göre, bağımlı değişken YAYE ile bağımsız değişkenler DS, ETA ve NBG arasında sırasıyla p değeri $4.0994e^{-60}$, $1.5887e^{-48}$ ve $2.5494e^{-31}$ arasında anlamlı bir ilişki vardır. Ayrıca, DS ETA ve NBG'nin p değerleri 227.73, 220.87 ve 152.41 F test sonuçları ile desteklenir. Öte yandan, elde edilen model ayrıca 0.031276 p değeri ve 4.94 F testi ile SRS'den nispeten daha az etkilenir. Sonuç olarak, ortaya çıkan ÇDR modeli, araç parametrelerinin CO_2 emisyonlarını etkilediğini ortaya çıkardığı için yeni araç tasarımlarında kullanılabilir.

Anahtar Kelimeler: Yakıt verimliliği, Makine Öğrenimi, Çoklu lineer Regresyon, CO2 Emisyonu

* Corresponding Author: yunusegi@sirnak.edu.tr

1. Introduction

The fuel efficiency of vehicles has become one of the most fundamental issues for urban sustainability and its adverse effect on human health due to CO2 emission [1]. Intergovernmental Panel on Climate Change (IPCC) states that urban transportation significantly impacts CO2 emission by releasing 20−30% of total greenhouse gases [2]. These emission rates also indicate how large the fuel consumption and cost for the global economy. In this regard, the Environment Protection Agency (EPA) asks car producers to provide corporate average fuel economy (CAFE) test in the National Vehicle and Fuel Emissions Laboratory located in Michigan [3]. Every year, EPA supplies car efficiency data to the Internal Revenue Service (IRS), the Department of Transportation (DOT), and the Department of Energy (DOE) to manage their car efficiency associated with programs [4]. Thus, the car manufacturers tend to minimize the CO2 emission to enlarge the fuel-efficient vehicles by applying quality assurance practices to fulfill the EPA's standards and customers' expectations. Besides, the demand for fuel efficiency increases as the costumers pay more attention to fuel efficiency and CO2 emission. Therefore, it is essential to measure the facility's performance by illustrating sufficient evidence through engine parameters including RHP, NCR, NG, ETW, and RAFE. These parameters have been applied in numerous fields for years in the industry.

Nevertheless, determining the optimum parameters need a meticulous investigation [5]. In the literature, there are many methods to estimate fuel efficiency. However, most studies are based on road conditions, engine speed, and fuel consumption based on GPS location rather than engine parameters. Mahsa et al. [6] introduce a new mechanical model for fuel consumption. Instead of using standard vehicles parameters such as engine type, road conditions, and wind effect, they utilize temperature, driving, road condition, and fuel efficiency. Shamsuddin et al. [7] performed response surface methodology to estimate CNC-Diesel dual-fuel engine emission. According to their study, the engine speed having a fraction in the range of 0% and 40% is used as an independent variable to see the effects of fuel emission on brake torque, brake power, brake specific energy consumption. Nguyen et al. [8] predict the co-collection and the regular garbage packer fuel consumption per km using GPS data and average fuel rates. Based on their estimation, rural areas have 2-4 times higher fuel consumption. Treiber et al. [9] demonstrate an instantaneous fuel consumption model consisting of car types, gear selection, and vehicle properties. According to their model, microscopic traffic simulation software determines fuel consumption trajectories and derived emission (CO2).

In recent years, the implementation of Machine Learning (ML) algorithms such as Linear Regression, Support Vector Machines, and Artificial Neural Networks is skyrocketing due to advancements in computer technology [10]. The main idea behind ML is to create a predictive model that minimizes the residual error through optimization algorithms such as gradient descent, least square, and Adam. The whole process is called Training. In addition, the training process should be tested. In the data preparation process, the data is divided into two parts as Training and Test sets. In this article, Multi-Linear Regression (MLR) in a machine learning and least square optimization algorithm will be used to predict dependent variables through independent explanatory variables. The training process MLR comprises many steps, including preparation of training data set $(x, y) \in R^{n+1}$

having m elements, finding the correlation of data set for every independent variable, computing cost to minimization of prediction error, and finally obtaining a model.

The paper is organized as follows: Section II. Material and Methods, Section III. Multiple Linear Regression Method, Analysis, and Results.

2. Material and Method

In this article, Multiple Regression Analysis is performed to predict quality results based on linear regression analysis and determine whether the predictions are suitable to CAFE standards. The correlation and ANOVA analysis will be used to evaluate the proposed model. Our purpose is to estimate which car is more efficient in the market and their quality requirements. To begin the analysis, the independent and dependent variables should be understood to determine their correlation.

2.1. Independent variables

Rated horsepower: The horsepower is a measure of performance that shows how capable a car is when moving between two points[11]. While moving, 550 ft/lbs per second of energy per horsepower is generated. The more horsepower vehicle has, the better acceleration will be achieved. On the other hand, the vehicle Rated Horsepower(RHP) is defined by the engine's maximum produced horsepower under nominal conditions. Thus, it is an essential measure for vehicle performance. The RHP is indicated in Equation (1).

$$P_{hp} = \frac{T[ft.lbf]N[rpm]}{5252} \qquad (1)$$

Where T is torque in pound-foot, and N is the rotational speed.

The number of cylinders and Rotors(NCR): The cylinder is an engine component placed in the engine's center for alternation and movement. The space between the cylinder lets the pistons move up and down. The mathematical calculation which gives us the cylinder volume(V_c) and Engine Displacement(Ed) is given by Equation (2) and (3)[12].

$$V_c = \pi \left(\frac{bore}{2}\right)^2 Stroke \qquad (2)$$

$$E_d = V_c N_c \qquad (3)$$

A rotor is a disc that is very crucial for the braking system. At every end of the axle, there is a brake rotor belonging to the braking system[13]. Thus, the reliability of this part will play a vital role in successfully stopping the car. If these parts lose their integrity by ignoring their maintenance, the braking system will cause unintentional consequences, and sometimes it can even cost some people lives. Besides, the aim is to see the effect of the rotors on emission efficiency, not its safety.

Number of Gears(NG): Gear is part of the machine that has cut teeth meshed with another toothed to transmit torque. In many situations, the teeth on the gears have the same shape and are connected. If there are more than two gears connected to the system, it is called transmission. The main task of the gears is to change the speed and torque by transmitting motion between shafts which may affect the CO_2 emission[14].

Equivalent Test Weight (lbs.)(ETW): Equivalent test weight is such an important factor for our estimation since the car's weight will cause more gas consumption leading to less efficiency[15].

2.2 Dependent variables

Rounded and Adjusted Fuel Economy(RAFE): Fuel economy is a measure of distance that vehicles use a certain amount of fuel to travel[16]. The distance and fuel units can be changed from country to country. For instance, in the USA, the fuel economy is miles per gallon, whereas, in Europe, it is liters per kilometer. The fuel economy is an indication of CO_2 efficiency. In our research, the rounded and adjusted fuel economy (RAFE) is our dependent value influenced by our independent values. By providing the expected value of fuel economy, it is aimed to reduce CO_2 emission.

3. Multiple Linear Regression In Machine Learning

Multiple Linear Regression (MLR) is a machine learning algorithm that explains the relationship between dependent and independent variables [17]. The primary purpose of MLR is to model-independent variables to predict dependent variables. The general formula for multiple regression is demonstrated in Equation(4).

$$Y_i = \beta_0 X_{i0} + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots\ldots + \beta_n X_{in} + \varepsilon \qquad (4)$$

where i is the number of observations, Y_i is the dependent variable, X_{in} is independent variables, β_0 is the intercept, βn is slope coefficients for each independent variable, ε is the model's error term (also known as the residuals).

3.1 Correlation Coefficients

The correlation coefficient represents correlations between two random variables, X and Y. The correlation range is distributed between −1 and 1. 1 represents positive correlation, and −1 represents negative correlation. The correlation coefficients formula is represented in Equation (5)[18]:

$$\rho(X,Y) = \frac{\mathbf{Cov}(X,Y)}{\sqrt{Var(X)Var(Y)}} \qquad (5)$$

3.2 Linear Model Testing

There is just a single variable in simple linear regression that affects the dependent variable. Moreover, the R^2 describes the model by dividing the Sum of Squares (SSR) to the Total Sum of Squares (SST) as indicated in Equation (6)[19].

$$R^2 = \frac{SSR}{SST} = \frac{\sum(\hat{y} - \overline{y})^2}{\sum(y - \overline{y})^2} \qquad (6)$$

If there is a small linear relationship among dependent and independent variables, R^2 will incline to approaches to zero. In contrast, if the association between variables is high, then R^2 will incline close to 1. If linearity between y and x is 100%, R^2 will be equal to 1. MATLAB will be used in this research to see that relationship in a table. The statistical summary of data will estimate the model to make sure that the model is trustworthy.

3.3 Confidence Interval

Confidence interval is used to determine how reliable our estimation is [20]. By itself, the confidence interval is calculated by observations, and it changes from one data point to another data point. If $X_i \sim N(\mu, \sigma^2)$ with known σ^2 and unknown μ. $100(1 - \alpha)\%$ confidence interval for μ can be calculated as Equation (7).

$$\widehat{\mu_L} = \bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n}, \; \widehat{\mu_U} = \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n} \qquad (7)$$

Where quantile $z_{\alpha/2}$ is given by $P(Z > z_{\alpha/2}) = \alpha/2$. For instance, a 95% *confidence interval* for μ is represented in Equation (8).

$$\widehat{\mu_L} = \bar{x} - 1.96 \cdot \sigma/\sqrt{n}, \; \widehat{\mu_U} = \bar{x} + 1.96 \cdot \sigma/\sqrt{n}. \qquad (8)$$

3.4 ANOVA

Analysis of Variance(ANOVA) determines significant differences in means between groups [21]. Several factors such as p-value and F statistics determine the significance of the model. If the p-value is less than the desired confidence(α), there will be enough evidence to claim that the null hypothesis is true. F statistic can also be used for the evaluation of the model. The higher f value will have a better linear relationship. The ANOVA table can be constructed via treatments and Error-values, which will help us calculate SSE, DoF, MS, F, and p-value. A representation of ANOVA is presented in Table 1.

Table 1. ANOVA

Source	SSE	DoF	Ms	F	p-value
Treatment	SSE_W	k-1	$\dfrac{SSE_W}{k-1}$	$\dfrac{MS_W}{MS_B}$	P<0.05
Error	SSE_B	n-k	$\dfrac{SSE_B}{n-k}$		
Total	SSE_T	n-1			

3.5 Training Process

Training in MLR is required for the optimization of the predictions. Therefore, the raw data should be cleaned from outliers leading to faulty results. After the data cleaning process, the data is divided into Training and test set. The data proportion of each part depends on the training options. This article distributes the data as 70% training and 30% test sets. Once the training data set is prepared, it will be multiplied with the proposed model's randomly selected weights(β0, β1, β2, β3, and β4) to make predictions. Then these predictions will be iteratively compared with actual values to minimize the cost(Error). The random weights are adjusted based on the cost values, and predictions are remade until the optimum result is found. The training pipeline of the proposed method is represented in Figure 1.
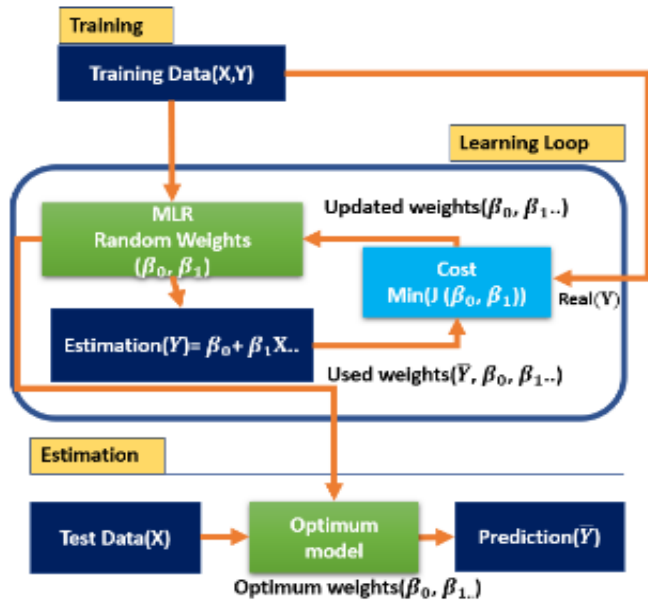
Figure 2. MLR Learning Structure

This article will optimize the cost through the Normal Equation since this learning algorithm has a closed form of solution as indicated in Equation (9).

$$\hat{\beta} = \left(\hat{X}^T \hat{X}\right)^{-1} \hat{X}^T \hat{Y} \qquad (9)$$

4. Analysis and Results

This article uses the MATLAB 2019a student version and HP OMEN Laptop to train and test the proposed linear model. The car efficiency data is obtained from the US Department of Energy website, which is publicly available. The raw data(3722 data points) containing NCR, NG, and ETW as dependent parameters are selected and cleaned from outliers. The process reduced the data set to 3583 data points. To understand which variable contributes to the dependent variable(RAFE), how each independent variable is related to one another should be evaluated. This evaluation can be done with scatter plots, using simple linear regression. Each independent variable is compared to the RAFE and summarized in the linear model. This summary comprises R^2, which determines the weight of importance from 0-1, as seen in Table II and Figure 2.

Table 2. Correlation Analysis

	RAFE	RHP	NCR	NG	ETW
RAFE	1.0000	-0.5063	-0.5360	0.0535	-0.4969
RHP	-0.5063	1.0000	0.8668	0.3123	0.5953
NCR	-0.5360	0.8668	1.0000	0.1930	0.7266
NG	0.0535	0.3123	0.1930	1.0000	0.2169
ETW	-0.4969	0.5953	0.7266	0.2169	1.0000

According to Table II, It is seen that RAFE has a negative correlation with RHP, NCR, and ETW, while the NG has a slight positive correlation. While seeing this relationship from correlation analysis, every single parameter can also be evaluated with single linear regression. In Figure 2, the actual and predicted values are represented as blue and yellow points, respectively. From Figures 2a-2d, it is seen that predicted values of RAFE are within the average of actual values and are in the same direction.
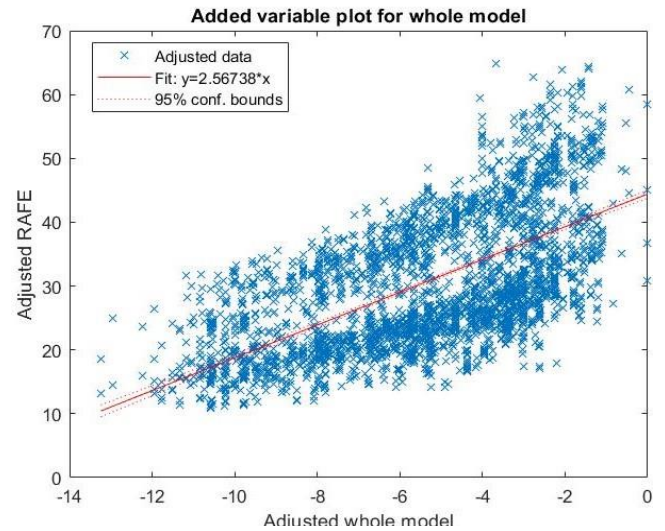
Figure 1. 95% confidence interval for RAFE versus adjusted whole model

These similar behaviors mean that all the parameters will contribute to the final model estimation of MLR.

4.1 Development and Model Estimation with MLR

The training process is performed through MLR and the Normal Equation. The MLR model is obtained as indicated in Table III, which summarizes our regression analysis. From the summary, β_0, β_1, β_2, and β_3 are given in the first column of the table under the name of estimation. The corresponding SSE, t-state, and p-values are also represented in the 3-5th columns.

Table 3. Estimated Coefficients for MLR

	B_i	SSE	t-stat	p-value
Intercept	44.436	1.1615	38.256	9.5276e-269
RHP	-0.036953	0.0029932	-12.346	2.5494e-34
NCR	-0.48951	0.22721	-2.1544	0.03276
NG	2.52	0.15121	16.665	4.0994e-60
ETW	-0.004022	0.00027063	-14.862	1.5887e-48

Number of observations: 3583, Error degrees of freedom: 3578
Root Mean Squared Error: 8.73
R-squared: 0.371, Adjusted R-Squared: 0.37
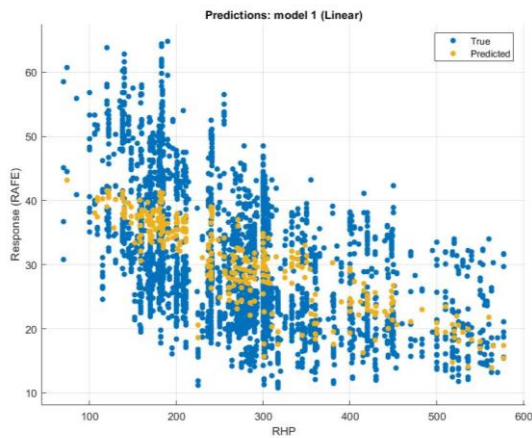F-statistic vs. constant model: 527, p-value = 0

The model also explains that the number of observations is 3583 with an Error DoF 3578. RMSE value of the model is obtained as 8.73. In addition, the R^2=0.37 means that the model explains 37% of the observations. Lastly, the obtained p-values are all less than 0.05, which means that all the independent variables significantly affect RAFE, directly affecting the CO_2 emission. Since all the values are calculated, the values can be replaced to construct the model as indicated in Equation (10)

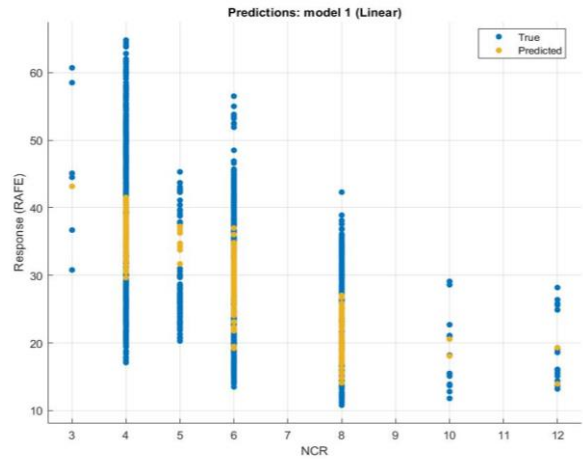$$RAFE_i = 44.4 - 0.03 RHP_i - 0.4 NCR_i - 2.528 NG_i - 0.004 ETW_i \qquad (9)$$

The constructed RAFE estimation model is evaluated using the ANOVA table as seen in Table IV. According to the ANOVA table, the contribution of every independent variable can be observed

through the F test and p-value. It is seen that NG has the highest contribution with an F value of 277.73, while the
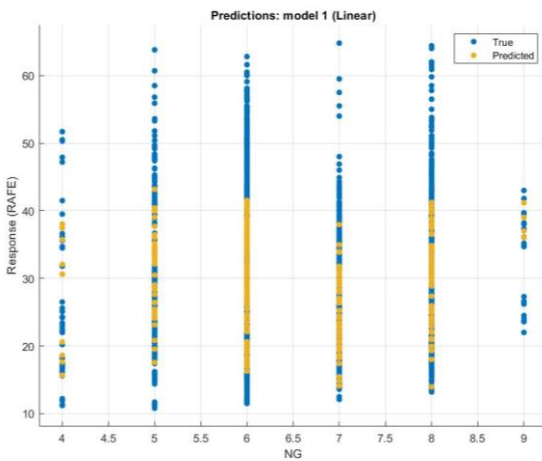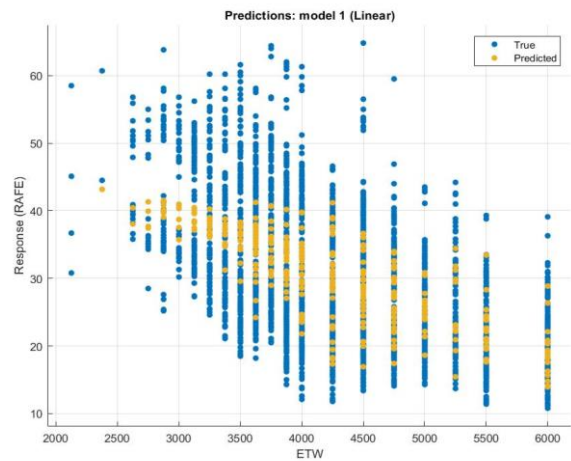

a.) RAFE vs RHP

b.) RAFE vs NCR

c.) RAFE vs NG

d.) RAFE vs ETW

Figure 2. Single Linear Regression Results

NCR has the lowest contribution with an F value of 4.64. Finally, ETW and RHP can be considered second and third highest contributors for the model, respectively.

Table 4 ANOVA Results

Source	SSE	DoF	Ms	F	p-value
RHP	11608	1	11608	152.41	$2.5494e^{-34}$
NCR	353.51	1	353.51	4.6414	0.03276
NG	21153	1	21153	277.73	$4.0994e^{-60}$
ETW	16823	1	16823	220.87	$1.5887e^{-48}$
Total	27252	4	76164		
All Model	y=2.56738x				

After obtaining the model and evaluating the ANOVA, the 95% confidence interval for the adjusted RAFE and

whole adjusted model y=2.56738x is plotted in Figure 3.

5. Discussion

It should be noted that many important characteristics can make the model stronger in terms of car emission efficiency. However,

this research chooses the essential characteristics to achieve optimum performance. If NG is evaluated individually, it will be seen that it is the essential characteristic of this project since the correlation between NG and RAFE have a higher F statistic value. The other three characteristics also play a critical role in fuel emission efficiency, but their properties are not considered crucial as NG. ETW is the second important characteristic that affects the quality of car emission efficiency. According to the results, it is seen that as ETW is increased, the car efficiency decreases. Finally, to achieve the best fuel efficiency and reduce CO_2 emission, the obtained MLR model needs to be used for predictions.

6. Conclusion

MLR in machine learning is used in this research to estimate vehicle fuel efficiency. The prediction process went through several processes such as data acquisition, removing outliers, training, testing, and performance evaluation. After the processes, an MLR model is obtained and evaluated by the ANOVA table. The ANOVA results indicate a strong relationship between RAFE and NG, ETW, and RHP with the p- values of $4.0994e^{-60}$, $1.5887e^{-48}$, and $2.5494e^{-34}$, respectively. On the other hand, there is a slight relationship between RAFE and NCR since the p-value is between

0.01 and 0.05. From the ANOVA table, it is understood which population group has more contribution by observing the F value. The F value 277.73 means that NG is more weighted than the others since it has the highest group value. The other parameters weights can be ordered as ETW, RHP, and NCR. This estimation can be implemented in software and help the clients and producers achieve minimum CO_2 emission in future work.

References

[1] Kan, Z., Tang, L., Kwan, M. P., Zhang, X. (2018). Estimating vehicle fuel consumption and emissions using GPS big data. International journal of environmental research and public health, 15(4), 566.

[2] He, H., Liang, X. Z., Wuebbles, D. J. (2018). Effects of emissions change, climate change and long-range transport on regional modeling of future US particulate matter pollution and speciation. Atmospheric Environment, 179, 166-176.

[3] Hula, A., Maguire, A., Bunker, A., Rojeck, T., Harrison, S. (2021). The 2021 EPA Automotive Trends Report: Greenhouse Gas Emissions, Fuel Economy, and Technology since 1975 (No. EPA-420-R-21-023).

[4] Tan, Z., Wu, Y., Gu, Y., Liu, T., Wang, W., Liu, X. (2022). An overview on implementation of environmental tax and related economic instruments in typical countries. Journal of Cleaner Production, 330, 129688.

[5] Choi, Y. Y., Liu, Y., Huang, L. (2015). Safer or Cheaper? Household Safety Concerns, Vehicle Choices, and the Costs of Fuel Economy Stan- dards (No. 330-2016-14033).

[6] Qaemi, M., Heravi, G. (2012). Sustainable Energy Performance Indicators of Green Building in Developing Countries. In Construction Research Congress 2012: Construction Challenges in a Flat World (pp. 1961-1970).

[7] Shamsuddin, M. S., Zulkifli, A. F. H. (2021). Prediction of Performance and Emission of CNG-Diesel Dual Fuel Engine using Response Surface Methodology. Progress in Engineering Application and Technology, 2(2), 790–809. Retrievedfromhttps://publisher.uthm.edu.my/periodicals/index.php/peat/article/view/704

[8] Nguyen TTT, Wilson BG. Fuel consumption estimation for kerbside municipal solid waste (MSW) collection activities. Waste Management Research. 2010;28(4):289 297. doi:10.1177/0734242X09337656

[9] Treiber, M., Kesting, A., Thiemann, C. (2008). How much does traffic congestion increase fuel consumption and emissions? Applying a fuel consumption model to the NGSIM trajectory data. In 87th Annual Meeting of the Transportation Research Board, Washington, DC (Vol. 71, pp. 1-18).

[10] Eyceyurt, E., Zec, J. (2020). Uplink Throughput Prediction in Cellular Mobile Networks. International Journal of Electronics and Communication Engineering, 14(6), 149-153.

[11] McCartt, A. T., Hu, W. (2017). Effects of vehicle power on passenger vehicle speeds. Traffic injury prevention, 18(5), 500-507.

[12] Dolatabadi, N., Forder, M., Morris, N., Rahmani, R., Rahnejat, H., Howell-Smith, S. (2020). Influence of advanced cylinder coatings on vehicular fuel economy and emissions in piston compression ring conjunction. Applied Energy, 259, 114129.

[13] Nascimento, T. P., Saska, M. (2019). Position and attitude control of multi-rotor aerial vehicles: A survey. Annual Reviews in Control, 48, 129-146.

[14] Triantafyllopoulos, G., Kontses, A., Tsokolis, D., Ntziachristos, L., Samaras, Z. (2017). Potential of energy efficiency technologies in reducing vehicle consumption under type approval and real-world conditions. Energy, 140, 365-373.

[15] Thomas, J. (2016). Vehicle efficiency and tractive work: rate of change for the past decade and accelerated progress required for US fuel economy and CO2 regulations. SAE International Journal of Fuels and Lubricants, 9(1), 290-305.

[16] Dreyer, S. J., Teisl, M. F., McCoy, S. K. (2015). Are acceptance, support, and the factors that affect them, different? Examining perceptions of US fuel economy standards. Transportation Research Part D: Transport and Environment, 39, 65-75

[17] Zhao, Q., Caiafa, C. F., Mandic, D. P., Chao, Z. C., Nagasaka, Y., Fujii, N., ... Cichocki, A. (2012). Higher order partial least squares (HOPLS): a generalized multilinear regression method. IEEE transactions on pattern analysis and machine intelligence, 35(7), 1660-1673.

[18] Adler, J., Parmryd, I. (2010). Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. Cytometry Part A, 77(8), 733-742.

[19] Plonsky, L., Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. The Modern Language Journal, 102(4), 713-731.

[20] B. J., Dehghani, H., Shamsi, R. (2020). Predicting silver price by applying a coupled multiple linear regression (MLR) and imperialist competitive algorithm (ICA). Metaheuristic Comput Appl, 1(1), 101-114.

[21] Rouder, J. N., Engelhardt, C. R., McCabe, S., Morey, R. D. (2016). Model comparison in ANOVA. Psychonomic bulletin review, 23(6), 1779-1786.