Jurnal Pengabdian Nusantara

Vol. 3, No. 4, Desember 2025, pp. 149~159
ISSN: 2987-2413
http://kampungjurnal.org/index.php/JPN/index

Comparative Analysis of XGBoost Method with Gradient Boosting for Vehicle Carbon Emission Prediction

Muhammad Reza Mahendra Laiya [*]**, Bayu Arma Praja, Victoria Beatrice**
Fakultas Informatika, Teknik Informatika, Universitas Mikroskil
Jl. Thamrin No. 140, Medan, (061) 4573767, Indonesia

Article Info

Article history:

Received September 26, 2025
Revised October 22, 2025
Accepted November 3, 2025

Keywords:

CO_2 Emissions, XGBoost, Gradient Boosting, Prediction, vehicles

ABSTRACT

Predicting vehicle carbon dioxide (CO_2) emissions is an important effort to reduce the environmental impact of the transportation sector. This study compares the performance of two boosting models, namely XGBoost and Gradient Boosting, in predicting vehicle CO_2 emissions. The dataset used includes various technical features of the vehicle such as engine size, fuel consumption, and transmission type. The pre-processing stage of data includes normalization and encoding of categorical features to ensure data readiness. The XGBoost and Gradient Boosting models are implemented with 80:20 data sharing for training and testing. Model performance evaluation was conducted using Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2) metrics. The results showed that Gradient Boosting was slightly superior with MAE values of 2,092, MSE 14,414, RMSE 3,796, and R^2 0.9958, compared to XGBoost which achieved MAE 2,098, MSE 14,697, RMSE 3,833, and R^2 0.9957. Both models show excellent performance, with Gradient Boosting more accurate in predicting CO_2 emissions. These findings provide important insights for the development of environmental policies and the design of more environmentally friendly vehicles.

Corresponding Author:

Muhammad Reza Mahendra Laiya
Universitas Mikroskil
Email: mhdreza276@gmail.com

INTRODUCTION

Predicting vehicle carbon dioxide (CO_2) emissions is an important approach in an effort to understand and reduce the environmental impact of the transportation sector. Vehicle emissions in particular CO_2 play a significant role against the problem of global warming and environmental pollution, especially urban areas with transportation being one of the main sources of harmful gases such as carbon monoxide (CO) and nitrogen oxides (NOx) in addition to CO_2 emissions [1]. Private transportation, which uses mostly gasoline fuel, contributes greatly to releasing CO_2 emissions [2] into the atmosphere, where every gallon of fuel burned will produce about 8,887 grams of CO_2 [3]. Despite efforts from the automotive industry to reduce emissions, vehicle emissions levels are still high today and the transportation sector continues to be a major contributor to air pollution and climate change. Therefore, the need to develop predictive [2] [4] machine learning models that are able to accurately estimate vehicle emissions is very important. By considering factors such as engine size, fuel type, and fuel consumption with a prediction model in place, it can help provide insight into the environmental efficiency of different types of vehicles. In addition, the model can be used as a basis for developing stricter emission regulation policies and supporting the development of more environmentally friendly vehicles.

Machine learning-based approaches have proven to deliver promising results in predicting vehicle CO_2 emissions [5]. Traditional regression models often have limitations in capturing complex relationships between variables. In contrast, more advanced [6] machine learning algorithms such as Gradient Boosing and XGBoost (Extreme Gradient Boosting) are able to handle large and complex data more efficiently. Both algorithms use an ensemble learning approach that combines the power of many simple prediction models to improve accuracy and reduce prediction errors. [7]

Gradient Boosting is an algorithm that builds prediction models iteratively, focusing on fixing errors from previous iterations. This algorithm has been known for its good performance in a wide range of prediction applications, including regression analysis and classification. On the other hand, [8] XGBoost is a development of Gradient Boosting designed to improve computing efficiency and overcome overfitting through regularization. [9]

This study aims to fill in the performance comparison between Gradient Boosing and XGBoost by testing and comparing the accuracy, efficiency, and generalization of both algorithms using a dataset that includes the technical features of the vehicle and the CO_2 emissions generated. By understanding the advantages and limitations of each algorithm in this case, the research is expected to provide guidance for practitioners and researchers to choose the most appropriate model in vehicle emission data analysis.

LITERATURE REVIEW

2.1. Vehicle Carbon Dioxide Emissions

Carbon dioxide (CO_2) gas accounts for around 72% of greenhouse gas production globally and the vehicle sector generates 20% of CO_2 emissions being a major challenge. The European Union is implementing a number of regulations to address this, such as a vehicle emission limit of 95g/km from 2020 and a target of 70g/km by 2030, this effort requires innovation in vehicle design and technology to achieve a reduction in greenhouse gas emissions of up to 60% by 2050. In addition to the impact of global warming, CO_2 can have side effects on humans that need attention. It was recorded that in 2019 the WHO [10] (World Health Organization) stated that air pollution causes around 7 million deaths every year worldwide with 9 out of 10 people breathing polluted air. Air pollution can trigger a variety of serious diseases such as stroke, heart disease, lung cancer, chronic obstructive pulmonary disease and acute respiratory infections. The impact of air pollution is caused by exposure to carbon monoxide (CO) and carbon dioxide (CO_2) gas accumulation from vehicle emissions causing necrosis, chlorosis and growth retardation in plants. The following are the factors that affect vehicle emissions and air pollution: [2] [11]

a. Temperature and humidity: Meteorological conditions such as temperature and humidity have a significant influence on vehicle emissions because they both affect the pollutants released into the atmosphere.
b. Vehicle age: Older vehicles typically produce higher emissions due to component wear and tear that contribute to increased air pollution.
c. Road conditions: Road conditions such as urban or rural roads will affect vehicle emissions as fuel efficiency and driving patterns will differ.
d. Fuel type: The type of fuel used by the vehicle is an important factor because some types of fuel produce higher emissions than others.

And. Speed and mileage: The average speed of a vehicle and its mileage affect the level of emissions, as fuel consumption patterns differ at different speeds and distances traveled.

Several efforts have been made to reduce the use of carbon gas emissions such as India has implemented Bharat Stage (BS) emission standards since 2000 with the switch to BS-VI norms in 2020 which succeeded in reducing CO emissions by 7%, NMVOC 9.7%, OC 20% and BC 5% despite Nox emissions increasing by 4.9% due to the growth in the number of vehicles. In addition, the industry also has an important role in efforts to reduce the use of carbon gas emissions by improving energy efficiency and using green energy during production, adopting electric vehicles powered by electricity during use, and supporting the waste recycling cycle through environmentally friendly technology for efficient recovery and reuse of materials. [12] [13]

ISSN: 2654-8127

2.2. Machine Learning

Machine learning (ML) is a technique for building models that connect predictors with responses and algorithms that are used to reveal input/output relationships based on data. Machine learning is the part of data analytics (DA) that focuses on data analysis to understand hidden patterns in complex datasets, ML is often used in predictive and predictive analytics to aid in decision-making without human intervention and enable model updates based on feedback. [14] Machine learning is divided into several types, namely: [15]

a. Supervised learning: A method that has results already known to the system, by reusing data that has been entered by the user before.
b. Unsupervised Learning: The concept of an unsupervised learning method where a method whose results are not known by the user, and depends on the weight value that was compiled at the beginning of the system development.

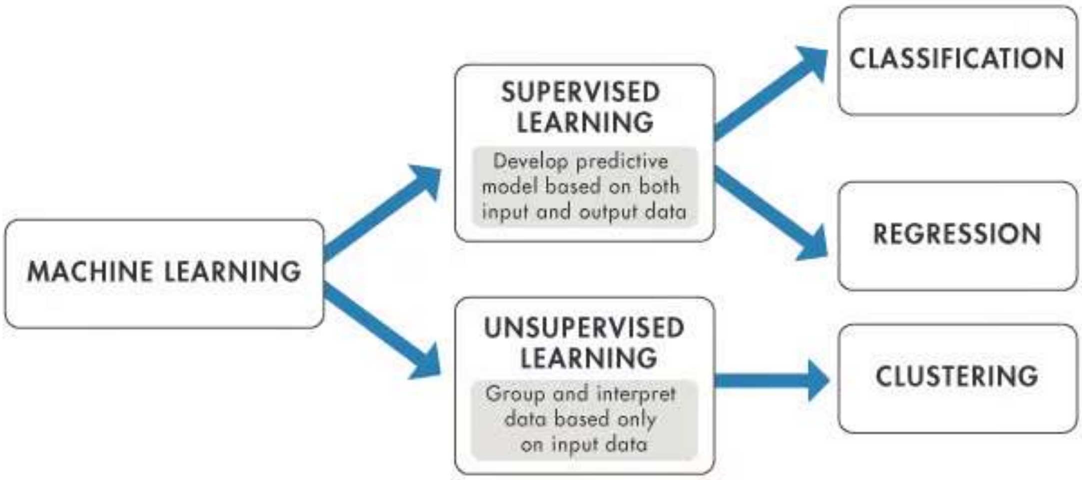

Figure 1. Types of machine learning [16]

2.3. Gradient Boosting

Gradient boosting is a machine learning technique for binary classification that builds decision trees sequentially, where each tree will correct the mistakes made by the previous tree. Gradient boosting predicts the terminal nodes of the tree not based on residual averages, but rather is calculated using equations involving the residual and probabilities estimated in previous iterations to optimize the mode in reducing prediction errors. In the boosting gradient there is a terminal node of the tree, the formula for calculating the prediction in the terminal node is as follows:jT_m [17]

$$Y_j = \frac{\sum_{i\epsilon\Omega_{mj}} r_m^{(i)}}{\sum_{i\epsilon\Omega_{mj}} p_{m-1}^{(i)}(1- p_{m-1}^{(i)}} \tag{1}$$

The basic concept of gradient boosting is to combine a number of simple models or weak learners into one stronger model. In general, for the model optimization process, mean squared error (MSE) plays a role in performing regression and log-loss tasks for classification. Each iteration, the model calculates residual errors by subtracting the predicted results of the target value and then adding a new weak learner to handle the error. With this approach, gradient boosting gradually improves the model by incorporating contributions from various weak learners [18].

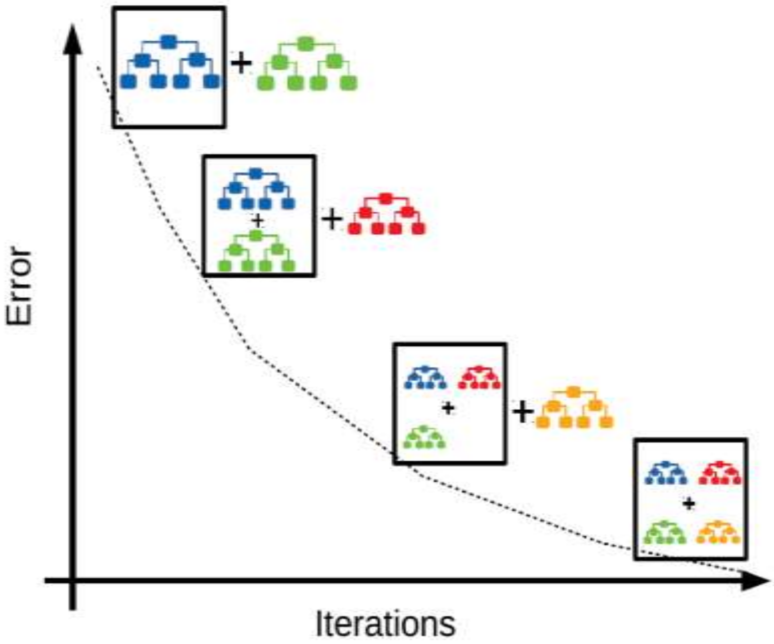

Figure 2. Example of gradient boosting iteration [18]

2.4. XGBoost

XGBoost is an effective machine learning algorithm for structured data. By building the model sequentially and XGBoost will correct the data errors that have been made beforehand. XGBoost applies [19] ensemble learning techniques by combining predictions from various models to improve overall. This approach is based on the boosting method, where simple models such as decision trees (weak learners) are combined into one stronger model (strong learners). As an improved version of the [20] Gradient Boosting Decision Tree (GBDT) algorithm, XGBoost shapes its model through the merging of several decision trees, where the training process involves the continuous addition of trees, with each new tree being formed based on the residue of the previous tree. [21]

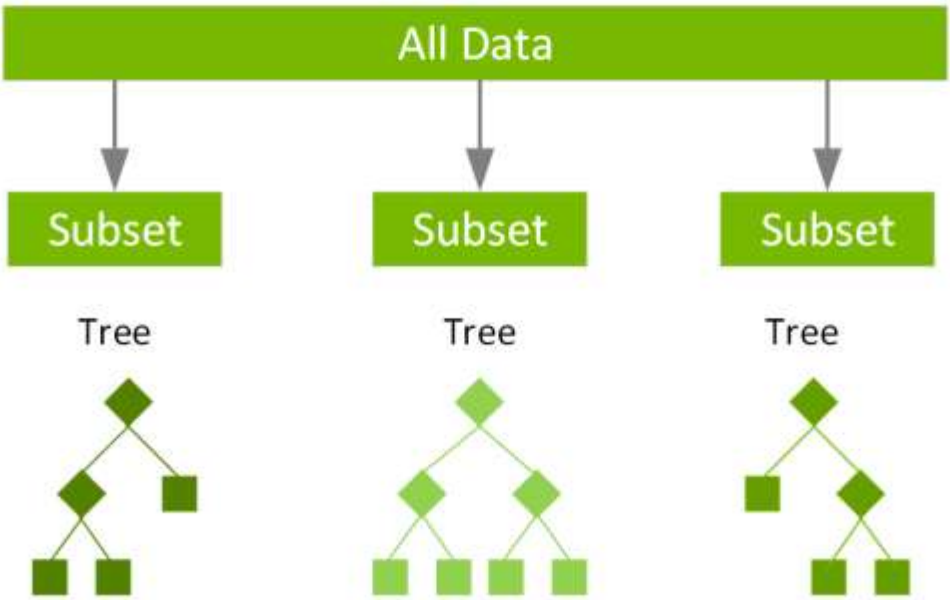

Figure 3. Source: NVIDIA [22]

METHOD

The research stage involves a series of systematic steps taken to achieve the research objectives. This study aims to compare the performance of two [23] boosting methods, namely Gradient Boosting, and XGBoost, in predicting CO_2 emissions in vehicles. Therefore, the sequence of research stages used is visually depicted in Figure 3.

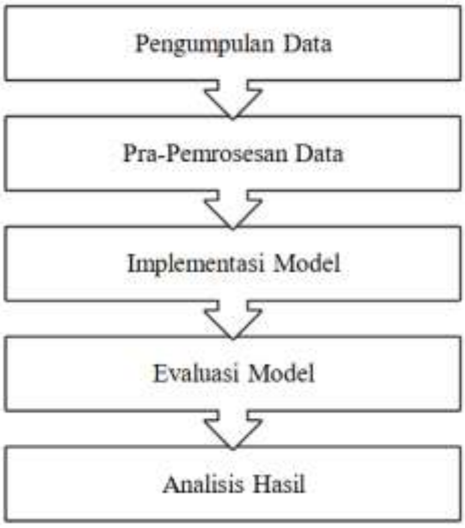

Figure 4. Research Steps

Figure 3 shows the stages of conducting the research. The following is a more detailed explanation of the research steps carried out.

Data Collection

This research stage begins with data collection. The data used is taken from a public dataset available on the Kaggle platform called "CO_2 Emission by Vehicles", containing information related to vehicle CO_2 emissions in Canada, along with parameters such as engine size, fuel consumption under various conditions, transmission type, and fuel type. [24]

Pre-Processing of Data

Once the data is available, the pre-processing step of the data becomes crucial. This step includes an initial exploration of the data structure, including attribute types, value distribution, and identification of anomalies or missing values. In this context, data distribution visualization is performed for several key features such as Engine Size(L) and CO_2 Emissions(g/km). The goal is to understand the characteristics of the dataset more deeply and recognize patterns that may be relevant for predictive analysis. After that, categorical features such as Vehicle Class and Fuel Type are converted into numerical form using one-hot encoding techniques so that they can be used by machine learning algorithms. Columns that are considered less relevant to the target variable, such as Make and Model, are removed based on domain considerations. This process ensures that only truly significant information will be presented to the model.

Model Implementation

In the next step, the implementation of the model is carried out using two approaches that focus on boosting, namely XGBoost and Gradient Boosting. The dataset that has gone through the cleaning and processing stages is divided into two parts, namely 80% training data and 20% test data. This is done to avoid overfitting while allowing for an objective performance evaluation of data that the model has never seen. Numerical data is normalized using StandardScaler so that each feature has a comparable scale.

The implementation of the XGBoost model involves setting parameters such as the maximum depth of the decision tree, the learning speed, and the number of estimates. The working principle of the XGBoost algorithm consists of several main components, namely the objective function, regulation, loss function, and model update. In the objective function section, the objective of XGBoost is formulated using two elements, namely the loss function and regulation. The function of this purpose is expressed by the equation:

$$Obj(t) = \sum_{\{i=1\}}^{n} L\left(y_i, \hat{y}_i^t\right) + \sum_{\{k=1\}}^{t} \Omega(f_k)$$

Information:
L = Loss function between actual and predicted values
$\Omega(f_k)$ = Regulation

The term regulation or regularization in XGBoost adds a penalty to the model based on its complexity. There are two types of regulations applied, namely (L_1 *Lasso*) and (L_2 *Ridge*) regulations. This regularization helps to make the model more general to data that has never been seen before. Mathematically, regulation is expressed by:

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j}^{T} w_j^2$$

Information:
T = Number of leaves of the tree.
w_j= Leaf weight
γ , λ = Regulation parameters

Next, the loss function is used to calculate the error in each iteration. This function is updated through equations:

$$Obj(t) \approx \sum_{i=1}^{n} \left[g_i f_{t(x_i)} + \left(\frac{1}{2}\right) h_i f_t^2(x_i)\right] + \Omega(f_t)$$

Information:
g_i = Gradien kesalahan
h_i = Hessian

At the model update stage, the predictions are updated for each iteration through the equation:t

$$\hat{y}_i^{(t)} = \hat{y}_i^{(t-1)} + \eta f_{t(x_i)}$$

Information:
η = Learning rate

The Gradient Boosting model follows a similar pattern, but uses a different algorithm in its boosting approach. The working principle of the Gradient Boosting algorithm can be explained through three main stages, namely weight initialization, learning iteration, and model combination.

In the early stages, the model is initialized with a simple weight. This initial model, denoted by , is usually only a target average for regression cases or a baseline probability for classification. This stage aims to provide a simple starting point so that subsequent iterations can gradually correct errors or residues. Next, the algorithm enters the iteration stage. In each iteration, the first step is to calculate the current prediction residue or error using the equation: $F_0(x)t$

$$r_{\{i,t\}} = -\left[\frac{\partial L\left(y_i, F(x_i)\right)}{\partial F(x_i)}\right]_{F=F_{t-1}}$$

Information:
L = Loss function
y_i = Actual target
$F(x_i)$ = Prediction from the model

ISSN: 2654-8127

This residue represents the magnitude of the errors that must be minimized in that iteration. Then, this residue is used as an input to train the Weak Learner. Weak Learners are tasked with studying patterns in previous prediction errors and generating new models aimed at improving predictions. Next, the model's predictions are updated by adding contributions from the newly trained *h_t(x)Weak Learner*, as per the equation:

$$F_{t(x)} = F_{\{t-1\}(x)} + vh_{t(x)}$$

Information:
v = Learning rate

The final stage is to combine all the weak models that have been produced. The final model of Gradient Boosting is formulated as:

$$F(x) = F_0(x) + \sum_{t-1}^{T} vh_{t(x)}$$

The formula above is a combination of the initial model and all Weak Learners with their respective weights adjusted by the learning rate. This final model represents the optimal prediction based on all the information that has been learned from previous iterations.

Model Evaluation

The evaluation was carried out to assess the performance of the model by utilizing four main evaluation metrics, namely Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). These four metrics provide different perspectives on the quality of the predictions, from the mean size of the error to the proportion of variance that the model successfully explains. For example, the R^2 value indicates the extent to which available features can explain changes in CO_2 emissions, while RMSE measures prediction errors on the same scale as the target.

Results Analysis

The final step in the study is the analysis of the results. Feature analysis is performed using the built-in attributes of the boosting model to understand the significance of each feature in influencing the target variable. As a result, features such as combined fuel consumption and engine size were found to have a dominant role in predicting CO_2 emissions. The performance of XGBoost and Gradient Boosting is compared visually, which shows that both models are capable of delivering high performance but have slightly different metric results. Bar graphs containing MAE, MSE, RMSE, and R^2 values for each model show substantial differences between the two approaches.

RESULTS AND DISCUSSION

This study predicts CO_2 emissions produced by vehicles, especially cars by analyzing the performance of 2 boosting techniques, namely Gradient Boosting and *XGBoost*. The dataset used to create the CO_2 emission prediction model includes many features, including Engine Size(L) and CO_2 Emissions(g/km). The first stage will be pre-processing of the data, where this step includes an initial exploration of the data structure, including attribute types, value distribution, identification of anomalies or missing values, and applying one-hot encoding to categorical features. To provide an overview of the distribution of the dataset used, data visualizations were made for several features. The following visualization of the data distribution for the Engine Size(L) feature is shown in Figure 4.

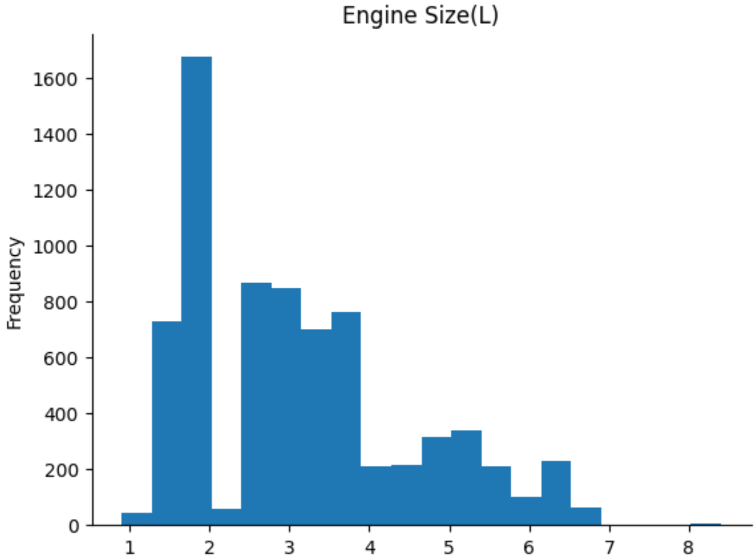

Figure 5. Data Engine Size(L) Distribution

The bar graph in Figure 5 shows the distribution of engine size data, where the data shows a right-skewed pattern, with a data range between 1 and 8 liters. The highest frequency is centered on the engine size of about 2 liters with more than 1600 vehicles, and the frequency decreases gradually for larger engine sizes. The majority of vehicles have engine sizes between 1.5 to 4 liters, while vehicles with engine sizes above 4 liters exhibit much lower frequencies. This distribution pattern reflects the trend of vehicles with lower carbon emissions, where smaller engine sizes tend to result in fewer CO_2 emissions. Furthermore, for data distribution, the Fuel Consumption City (L/100km) feature will be shown in Figure 5.

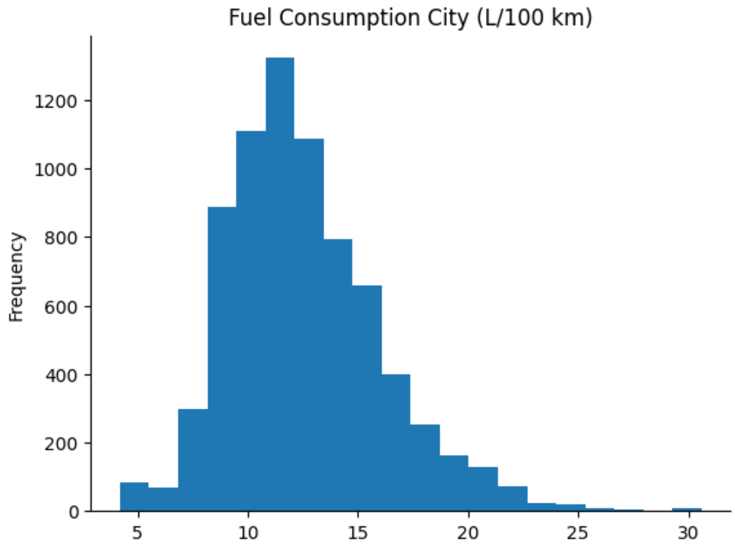

Figure 6. Fuel Consumption City Data Distribution (L/100km)

Figure 6 shows a distribution that is close to normal with a slight slope to the right, where the most consumption is in the range of 10-12 L/100 km with a frequency of up to 1300 vehicles. This pattern indicates a direct correlation with CO_2 emission levels, where vehicles with higher fuel consumption (>15 L/100 km) have a decreasing frequency, indicating the market's preference for more efficient and environmentally friendly vehicles. Next, it is for the distribution of CO_2 Emissions (g/km) data which will be visualized in Figure 6.

ISSN: 2654-8127

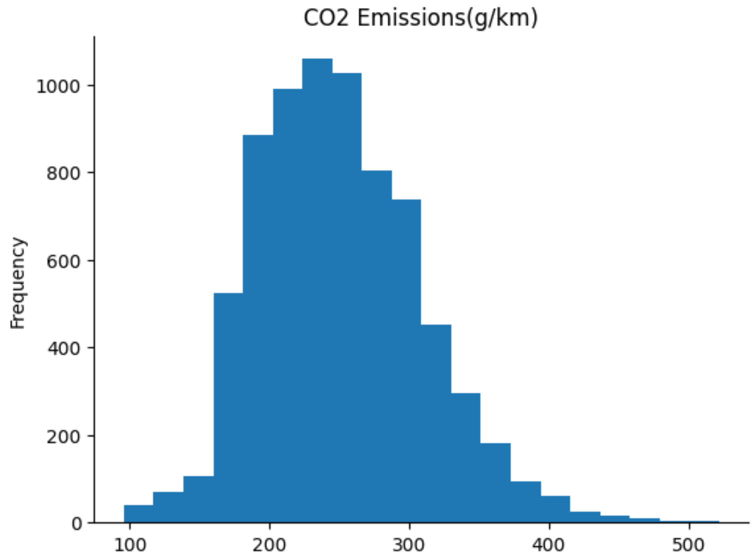

Figure 7. Distribution of CO_2 Emissions Data (g/km)

The distribution of CO_2 Emissions (g/km) data shown in Figure 7 shows the frequency distribution of carbon dioxide emissions from various cars, where emission frequencies range from 100 to 500 g/km, with a significant frequency peak of around 200 g/km. This shows that most of the cars in this dataset have moderate levels of CO_2 emissions. However, there are also a number of cars with higher emissions, reaching up to 500 g/km, which may reflect vehicles with lower fuel efficiency or larger engines.

The next process is to build a CO_2 emission prediction model using two boosting techniques, namely *XGBoost* and *Gradient Boosting*. The process of building the model is carried out using the Python programming languages and Google Colab. The first step is to separate the features and targets, where features such as engine size, fuel consumption, and more are separated from the target, namely CO_2 emissions. After that, the features are standardized using 'StandardScaler' to ensure that all features are on the same scale. The dataset was then divided into training data and test data with an 80:20 ratio. The XGBoost model is trained using 'XGBRegressor' from the 'xgboost' library, while the Gradient Boosting model is trained with the 'GradientBoostingRegressor' from the 'sklearn' library. Both models were trained using training data (*'X_train' and 'y_train'*). After training, the model was evaluated on test data by calculating metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R^2). The evaluation results for each of the models tested are shown in Figure 7.

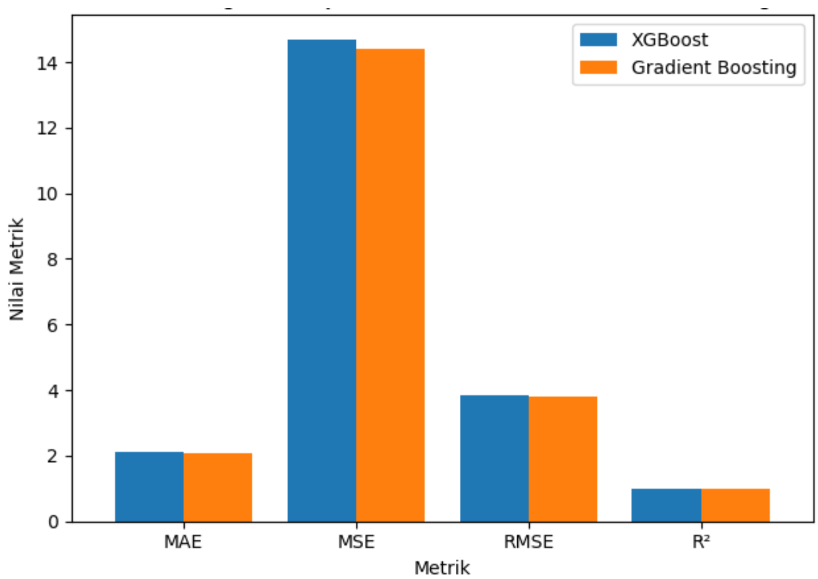

Figure 8. Model Performance Comparison

Figure 8 shows and compares the performance of two machine learning models, XGBoost and Gradient Boosting, using evaluation metrics such as MAE, MSE, RMSE, and R^2. XGBoost shows slightly higher MAE, MSE, and RMSE values than Gradient Boosting, which indicates a slightly larger prediction error compared to Gradient Boosting. The results of the evaluation of the two models in more detail are shown in Table 1.

Table 1. Comparison of Evaluation Results

Model Boosting	MAE	MSE	RMSE	R^2
XGBoost	2.098	14.697	3.833	0.9957
Gradient Boosting	2.092	14.414	3.796	0.9958

Based on the results of the model evaluation in Table 1, the two boosting models, namely XGBoost and Gradient Boosting, showed excellent performance in predicting CO_2 emissions. XGBoost has a MAE value of 2,098, MSE of 14,697, RMSE of 3,833, and R^2 of 0.9957. Meanwhile, Gradient Boosting has a MAE value of 2,092, MSE of 14,414, RMSE of 3,796, and R^2 of 0.9958. Although the difference between the two models is very small, Gradient Boosting is slightly superior in terms of MAE, MSE, and RMSE, which shows a slightly lower prediction error. However, both models have very high R^2 values, close to 1, which indicates that they are both able to explain the variations in target data very well. Overall, both models are effective for predicting CO_2 emissions, with Gradient Boosting showing slightly better performance when it comes to prediction errors.

CONCLUSION

This study has compared two boosting models, namely XGBoost and Gradient Boosting, in predicting vehicle CO_2 emissions. The highest results in the evaluation of CO_2 emission prediction were obtained from the Gradient Boosting model with a data sharing ratio of 80:20, where this model achieved a MAE value of 2,092, MSE of 14,414, RMSE of 3,796, and R^2 of 0.9958. The excellent performance of Gradient Boosting is due to its ability to handle data iteratively and focus on correcting errors from previous iterations, resulting in more accurate predictions. In addition, Gradient Boosting also shows high consistency in various evaluation metrics, making it a reliable model for CO_2 emission predictions.

On the other hand, XGBoost also showed excellent performance with a MAE value of 2.098, MSE of 14.697, RMSE of 3.833, and R^2 of 0.9957. Although slightly inferior to Gradient Boosting, XGBoost still excels in terms of computing efficiency and the ability to handle big data through parallelization and regularization techniques that prevent overfitting. The performance of both models shows that they are highly effective at predicting CO_2 emissions, with Gradient Boosting slightly superior in terms of prediction errors. These findings show that Gradient Boosting is the best model for predicting vehicle CO_2 emissions, followed by XGBoost.

Suggestion

Here are some suggestions for further research:

1. Exploring more diverse datasets from different geographic regions with different environmental conditions and emission regulations. In addition, the application of hyperparameter tuning techniques such as Grid Search or Random Search can be considered to optimize the performance of other models.
2. Using a more diverse dataset covering different regions with different environmental conditions, emission regulations and vehicle types.
3. Exploring other machine learning algorithms to compare performance with models studied by previous researchers.
4. The research can include several external variables such as weather conditions, fuel quality and traffic conditions to see the impact of CO_2 emissions.
5. Conducting model testing on electric and hybrid vehicles to understand how the model works on different types of vehicles with different emissions.
6. Apply model testing to datasets in real-time to validate model capabilities in a variety of scenarios.

ISSN: 2654-8127

REFERENCES

[1] D. D. Sari, A. Irfan, B. Muslim, and B. A. Seno, “Pengaruh Jumlah Kendaraan Bermotor Dengan Kadar Karbon Monoksida (CO) Di Udara Pada Jalan Perintis Kemerdekaan Dan Jalan Lubuk Begalung,” Jurnal Kesehatan Lingkungan Mandiri, vol. 2, no. 1, pp. 19–27, Apr. 2023, doi: 10.33761/jklm.v2i1.676.

[2] N. F. F. Yaacob, M. R. Mat Yazid, K. N. Abdul Maulud, and N. E. Abdul Basri, “Spatial temporal pattern of carbon dioxide emission from vehicle,” IOP Conf Ser Earth Environ Sci, vol. 1167, no. 1, p. 012009, May 2023, doi: 10.1088/1755-1315/1167/1/012009.

[3] S. Ramesh, S. S. I. M, and J. J. Justus, “CO 2 Emission Rating by Vehicles using Supervised Algorithms,” in 2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), IEEE, May 2023, pp. 1–8. doi: 10.1109/ACCAI58221.2023.10200707.

[4] Y. Li, Q. Wu, Y. Zhang, G. Huang, S. Jin, and S. Fang, “Mapping Highway Mobile Carbon Source Emissions Using Traffic Flow Big Data: A Case Study of Guangdong Province, China,” Front Energy Res, vol. 10, Apr. 2022, doi: 10.3389/fenrg.2022.891742.

[5] Kevin Asgaryansyah and Paniran Paniran, “Implementasi Algoritma Regresi Linier untuk Mengukur Tingkat Pengeluaran CO_2 Pada Kendaraan Bermotor,” Mars : Jurnal Teknik Mesin, Industri, Elektro Dan Ilmu Komputer, vol. 2, no. 3, pp. 84–93, Jun. 2024, doi: 10.61132/mars.v2i3.132.

[6] S. Shah, S. Thakar, K. Jain, B. Shah, and S. Dhage, “A Comparative Study of Machine Learning and Deep Learning Techniques for Prediction of CO_2 Emission in Cars,” Nov. 2022.

[7] C. Y. Wijaya, “See the substantial differences between the famous algorithms,” KDnuggets Technical Content Specialist.

[8] Anshul, “Gradient Boosting Algorithm: A Complete Guide for Beginners,” Analytics Vidhya.

[9] A. Tyagi, “What is XGBoost Algorithm?,” Analytics Vidhya.

[10] M. Mądziel, “Liquified Petroleum Gas-Fuelled Vehicle CO_2 Emission Modelling Based on Portable Emission Measurement System, On-Board Diagnostics Data, and Gradient-Boosting Machine Learning,” Energies (Basel), vol. 16, no. 6, p. 2754, Mar. 2023, doi: 10.3390/en16062754.

[11] X. Wang et al., “Research on Vehicle Emission Factors in Wuhan Based on MOVES,” E3S Web of Conferences, vol. 573, p. 01015, Sep. 2024, doi: 10.1051/e3sconf/202457301015.

[12] M. D. Gajbhiye, S. Lakshmanan, N. Kumar, S. Bhattacharya, and S. Nishad, “Effectiveness of India’s Bharat Stage mitigation measures in reducing vehicular emissions,” Transp Res D Transp Environ, vol. 115, p. 103603, Feb. 2023, doi: 10.1016/j.trd.2022.103603.

[13] C. Kong and F. Men, “Carbon emission reduction of automobile industry under carbon neutrality target,” in Advances in Civil Engineering and Environmental Engineering, Volume 2, London: CRC Press, 2023, pp. 511–516. doi: 10.1201/9781003383031-74.

[14] S. Mishra, J. Schuetter, A. Datta-Gupta, and G. Bromhal, “Robust Data-Driven Machine-Learning Models for Subsurface Applications: Are We There Yet?,” Journal of Petroleum Technology, vol. 73, no. 03, pp. 25–30, Mar. 2021, doi: 10.2118/0321-0025-JPT.

[15] S. S. Jarot, “JENIS JENIS MACHINE LEARNING,” BINUS.

[16] M. Crabtree, “Apa itu Machine Learning? Definisi, Jenis, Alat & Lainnya,” datacamp.

[17] H.-H. Chen, “Understanding Gradient Boosting Classifier: Training, Prediction, and the Role of γ_j,” Oct. 2024.

[18] Trivusi, “Gradient Boosting: Pengertian, Cara Kerja, dan Kegunaannya,” Trivusi.

[19] S. Hakkal and A. A. Lahcen, “XGBoost To Enhance Learner Performance Prediction,” Computers and Education: Artificial Intelligence, vol. 7, p. 100254, Dec. 2024, doi: 10.1016/j.caeai.2024.100254.

[20] “Tools Data Science dengan Algoritma XGBoost,” DqLab.

[21] K. Xu, Z. Xia, M. Cheng, and X. Tan, “Carbon price prediction based on multiple decomposition and XGBoost algorithm,” May 15, 2023. doi: 10.21203/rs.3.rs-2788190/v1.

[22] “What is XGBoost?,” NVIDIA.

[23] Y. Fernando, R. Napianto, and R. I. Borman, “Implementasi Algoritma Dempster-Shafer Theory Pada Sistem Pakar Diagnosa Penyakit Psikologis Gangguan Kontrol Impuls,” Information System Research Journal, vol. 2, no. 2, 2022.

[24] D. Podder, “CO_2 Emission by Vehicles,” kaggle.