Industrial Engineering Journal
ISSN: 0970-2555
Volume : 53, Issue 9, September : 2024

C02 EMISSION RATING BY VEHICLE USING DATA SCIENCE

[1]Kamani laya
Mtech. Department of computer science and engineering
Vaagdevi college of engineering, bollikunta,warangal,telangana.
Email: layakamani@gmail.com

[2] Dr. Sravan kumar B
Assist .Prof Department of computer science and engineering
Vaagdevi college of engineering, bollikunta warangal,telangana.
Email : sravan.researcher@gmail.com

[3]Dr .E. Balakrishna
Assoc .Prof Department of computer science and engineering
Vaagdevi college of engineering, bollikunta warangal,telangana.
Email : balakrishnakits@gmail.com

[4]Dr.N.Satyavathi
Assoc .Prof Department of computer science and engineering
Vaagdevi college of engineering, bollikunta warangal,telangana.
Email : satyanadendla15@gmail.com

***Abstract*—Carbon dioxide and other gases absorb sunlight, leading to rising global temperatures and air pollution. This pollution affects soil fertility, air quality, and water quality. Car pollution forces animals to leave their habitats, emitting pollutants like nitrogen dioxide, carbon monoxide, and formaldehyde. Increased city traffic also causes hearing problems and psychological ill-health. To reduce CO2 emissions, data is crucial for algorithm training and machine learning. Regression models are used to predict car emissions, with Road Transport Authority staff notifying car owners if emissions are within a threshold.**

Keywords—Co2, ANN, vehicles, data science

I. INTRODUCTION

Our private vehicles have become one of the most significant contributors to global warming, primarily by driving climate change and air pollution. The extent of their impact is startling, with cars collectively responsible for about 15% of the world's total carbon dioxide (CO2) emissions. Every time we fill up a vehicle's tank, we contribute to this growing problem: one liter of gasoline combusted in a car engine produces around 2.5 kilograms of CO2, released directly into the atmosphere through the car's exhaust. For an average passenger car, this results in approximately 252 grams of CO2 emitted per kilometer driven. Given the widespread use of passenger cars and the average driving distance, the cumulative emissions are substantial. The typical car on a highway, with a fuel efficiency of around 10 kilometers per liter and an annual mileage of 18,500 kilometers, emits an alarming 2420 grams of CO2 for every liter of gasoline consumed. This adds up to around 4.7 metric tons of CO2 emissions per vehicle each year, highlighting the considerable role that individual car usage plays in the broader context of global emissions. Recognizing the severe environmental impact, governments worldwide have made efforts to curb CO2 emissions from vehicles. Since 2000, numerous countries have implemented stringent regulations aiming to reduce emissions from brand-new cars by 35%, using standardized testing methods such as the New European Driving Cycle (NEDC). Back in 2000, the average CO2 emission from new cars was about 172 grams per kilometer. While advancements in vehicle technology and stricter regulations have helped reduce this figure to approximately 120 grams per kilometer today, the target of reaching 100 grams per kilometer by 2025 remains a formidable challenge. Despite these efforts, the total CO2 emissions from passenger cars have continued to rise, peaking at 3.2 billion metric tons in 2020, up from 2.2 billion metric tons in 2000. This upward trend underscores the complexity of reducing CO2 emissions in the transportation sector. The challenge is exacerbated by the growing global vehicle fleet, increased driving distances, and the continued reliance on fossil fuels. As we move forward, it becomes increasingly important to leverage data science and machine learning to tackle this issue more effectively. Understanding and mitigating vehicular CO2 emissions require a detailed analysis of a wide range of factors. Variables such as the make and model of the car, fuel type, engine size, transmission, and even driving behavior can significantly influence a vehicle's CO2 output. By gathering and analyzing extensive data on these attributes, we can develop sophisticated machine learning models capable of predicting CO2 emissions with greater accuracy. These models can be invaluable for both regulatory bodies and consumers, helping to identify vehicles that exceed acceptable emission levels and guiding purchasing decisions toward more environmentally friendly options. Moreover, these predictive models can serve as a foundation for new regulatory measures. For example, authorities could establish minimum CO2 emission thresholds based on model predictions. If a vehicle exceeds this threshold, the Road Transportation Authority (RTA) could issue a notification to the owner, urging them to service their car or take other corrective actions. Such an approach would not only help reduce overall emissions but also encourage car owners to

maintain their vehicles in a way that minimizes environmental impact. The integration of data science in managing CO2 emissions represents a significant step forward in our ability to combat climate change. By harnessing the power of machine learning and predictive analytics, we can develop more targeted and effective strategies for reducing the carbon footprint of passenger vehicles. As the global community continues to grapple with the challenges of climate change, the role of innovative technologies in reducing vehicular emissions will become increasingly critical. Ultimately, these efforts will contribute to a more sustainable future, where the environmental impact of transportation is minimized, and the health of our planet is preserved for future generations.

II. LITERATURE SURVEY

Over the past decades, advancements in cellular communication technologies have significantly expanded the accessibility and portability of web application programs. These advancements have, in turn, fueled the demand for machine learning-based prediction models, particularly in areas where real-time data processing is critical, such as in the automotive industry. As vehicles remain one of the largest sources of CO2 emissions globally, there has been an increasing focus on developing accurate prediction models to monitor and mitigate these emissions. With the continuous evolution of deep learning and machine learning techniques, methodologies that leverage OnBoard Diagnostic II (OBD-II) sensor data are emerging as powerful tools for predicting automobile CO2 emission levels.The literature reveals several important contributions in this domain, reflecting the growing interest in improving CO2 emission prediction accuracy. For example, the study referenced in [3] focused on predicting CO2 emission levels in urban areas using ensemble learning techniques and gradient boosting algorithms. The authors used a variety of vehicle-specific data points, such as model type, speed, transmission type, and cylinder size, as inputs to their model. By employing gradient boosting regression—a method well-suited for capturing complex non-linear relationships—the researchers were able to create a model that provided reasonable predictions of CO2 emissions. They recognized, however, that the accuracy of these predictions could be enhanced by incorporating additional factors, such as a vehicle's historical CO2 emission data. This insight points to the potential for further refining prediction models by considering a broader array of attributes.Another relevant study, cited as [1], utilized a regression model to predict CO2 emission levels based on historical environmental data sourced from the World Bank, spanning from 1960 to 2014. The dataset included a variety of factors contributing to CO2 emissions, such as biological processes, natural gas consumption, deforestation, and waste management practices. While this study was successful in modeling CO2 emissions on a global scale, the authors identified an opportunity to apply similar regression techniques specifically to predict CO2 emissions from vehicles. This suggests a clear research direction for developing more specialized models that focus on the nuances of vehicular emissions.In the context of automotive CO2 emission prediction, the existing models often rely on a limited set of vehicle attributes, such as transmission type and cylinder size. Although these factors are crucial, relying exclusively on them can lead to less accurate predictions. CO2 emissions are influenced by a multitude of variables, including the vehicle's fuel consumption rate, engine efficiency, driving patterns, load conditions, and even maintenance practices. Thus, there is a need to broaden the scope of attributes considered in predictive models to capture the complexity of factors affecting vehicular CO2 emissions.

Responding to these challenges, recent research has sought to enhance CO2 emission prediction accuracy by integrating a wider range of vehicle attributes into the models. For instance, a study by another group of researchers proposed a machine learning model that includes not only basic vehicle specifications but also real-time data collected from OBD-II sensors, which monitor various engine parameters. This model incorporates features such as fuel consumption, vehicle class, and specific emission data, which collectively provide a more comprehensive understanding of the factors influencing CO2 missions.

Furthermore, the integration of these advanced models into web-based applications represents a significant step forward in making such tools accessible to a broader audience. A notable example is the development of web applications that allow users to input their vehicle's details and instantly receive an estimate of CO2 emissions. These applications are particularly valuable for non-technical users, such as everyday car owners, who can utilize them to monitor and reduce their vehicle's environmental impact.Another aspect of ongoing research focuses on the real-time monitoring and predictive maintenance of vehicles to reduce CO2 emissions. By using continuous data streams from OBD-II sensors, machine learning models can detect patterns that indicate when a vehicle is likely to exceed acceptable CO2 emission levels. These models can then trigger alerts or recommendations for vehicle maintenance, ensuring that cars are kept in optimal condition to minimize their emissions. Moreover, some researchers have explored the potential of using deep learning models for more complex and dynamic prediction tasks. For example, recurrent neural networks (RNNs) and long short-term memory (LSTM) networks have been proposed to model the temporal dependencies in vehicle emission data. These advanced models can account for variations in driving conditions over time, offering more precise predictions of CO2 emissions based on historical driving data and real-time sensor inputs.The integration of these machine learning models with IoT (Internet of Things) technologies is another emerging trend. By connecting vehicles to a network of sensors and cloud-based analytics platforms, it is possible to create a comprehensive system that not only predicts CO2 emissions but also provides actionable insights for reducing emissions. Such systems could be integrated into smart city infrastructures, where data from thousands of vehicles could be aggregated and analyzed to optimize traffic flows and reduce overall urban emissions.In conclusion, the ongoing advancements in machine learning and deep learning, coupled with the increasing availability of real-time vehicle data through OBD-II sensors and IoT technologies, are

paving the way for more accurate and accessible CO2 emission prediction models. These models are not only becoming more precise but also more user-friendly, thanks to their integration into web-based applications that can be easily used by non-technical individuals. As research in this area continues to evolve, we can expect further improvements in the accuracy of CO2 emission predictions, ultimately contributing to more effective strategies for reducing the environmental impact of vehicles.

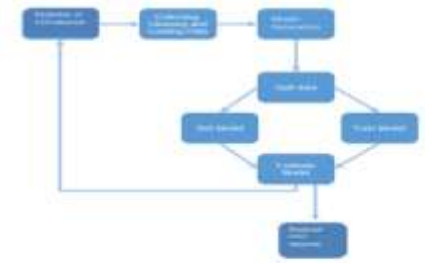
III. EXISTING SYSTEM

Current machine learning models for predicting CO2 emissions in vehicles face several limitations that impact their accuracy and effectiveness. These models often convert time-based data into distance-based metrics, which can introduce errors, especially in complex driving conditions. The accuracy of predictions is further constrained by the quality and relevance of historical data, which may not reflect real-time changes in driving behavior, road conditions, or vehicle maintenance. Additionally, many models oversimplify emissions by neglecting important contextual factors like fuel type, terrain, and load, leading to less precise predictions. The complexity of advanced models also poses challenges, such as the risk of overfitting and the lack of user-friendly interfaces, making them difficult for non-technical users to apply effectively in real-world scenarios.

IV. PROPOSED SYSTEM

In this paper, the author presents a groundbreaking approach for forecasting CO2 emissions in heavy vehicles, leveraging the capabilities of machine learning, specifically Artificial Neural Networks (ANNs). The approach integrates seven critical predictors, meticulously derived from an extensive analysis of vehicle speed and road grade data. This innovative methodology aims to develop a highly sophisticated neural network model that accurately captures the complex interactions between these input variables and CO2 emissions.The study delves into the construction of this neural network model with a focus on precision and reliability, emphasizing the model's ability to process and analyze intricate data patterns. By doing so, the model offers a robust framework for predicting the environmental impact of heavy vehicle operations with greater accuracy.The proposed approach not only advances the understanding of CO2 emission dynamics within the transportation sector but also provides actionable insights for improving sustainability and operational efficiency in heavy vehicle fleets. The model's predictive capabilities are expected to be instrumental in devising strategies for reducing emissions and enhancing the environmental performance of transportation systems.Furthermore, the paper explores the potential of this methodology to inform policy decisions and operational practices, ultimately contributing to the broader goals of environmental conservation and the reduction of greenhouse gas emissions. Through its comprehensive analysis and innovative use of machine learning techniques, the study represents a significant step forward in addressing the challenges associated with CO2 emissions in heavy vehicles.

V. SYSTEM DESIGN

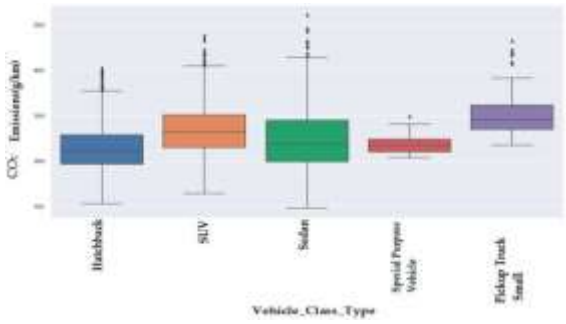


Dataset

This dataset meticulously details the CO2 emission records for a diverse range of vehicles, providing an extensive view of how different vehicle attributes influence emissions. Sourced from the Canadian government's open data portal, the dataset has been updated to reflect the most recent information and covers a substantial period of 7 years. It consists of 12 columns and 7385 rows, each representing various characteristics of the vehicles. The use of acronyms such as `fload64` for floating-point numbers (which can include both positive and negative values separated by a decimal point) and `int64` for integer values helps streamline the data presentation.

The dataset categorizes vehicles into 16 distinct classes, including hatchbacks, sedans, SUVs, and trucks. Each class is defined by specific attributes that contribute to its classification. Hatchbacks are four-door vehicles with a two-box design, where the rear hatch opens upwards, combining the passenger and cargo areas into one continuous space. Sedans, in contrast, also have four doors but feature a three-box design, with a separate trunk compartment. SUVs, or Sports Utility Vehicles, are larger, combining the passenger capacity of minivans with the off-road capabilities and towing power of pickup trucks, resulting in generally higher CO2 emissions due to their larger size and weight.

The analysis of this dataset reveals that vehicle size plays a significant role in CO2 emissions. Larger vehicles, such as SUVs and trucks, typically exhibit higher emissions compared to smaller vehicles like hatchbacks and sedans. This relationship is visually represented in the dataset's box plot, which illustrates the spread of CO2 emissions across different vehicle types.



In the box plot:

- The box itself captures the interquartile range (IQR) of CO2 emissions, with the lower and upper edges representing the 25th and 75th percentiles, respectively.

- The line that bisects the box horizontally denotes the median level of CO2 emissions, offering a central tendency for each vehicle class.

- Whiskers extending from the box indicate the range of CO2 emissions within 1.5 times the IQR from the quartiles.

- Data points beyond these whiskers represent outliers, indicating CO2 emissions that exceed the typical range for each vehicle class.

This detailed visualization underscores the increasing trend of CO2 emissions with vehicle size. For example, the plot demonstrates that SUVs, with their larger dimensions and more powerful engines, tend to have higher emissions compared to sedans or hatchbacks. The extreme values, as shown by the data points beyond the whiskers, highlight cases where certain vehicles produce exceptionally high CO2 emissions, further emphasizing the impact of size and engine characteristics on environmental outcomes.

By analyzing this dataset, researchers and policymakers can gain valuable insights into the relationship between vehicle attributes and CO2 emissions. The data highlights not only the typical emission profiles of various vehicle classes but also the outliers, which may warrant further investigation. Understanding these patterns is crucial for developing strategies to reduce greenhouse gas emissions in the automotive sector, promoting more sustainable vehicle choices, and informing regulations aimed at mitigating environmental impact.

Preprocessing

The term “data preparation” encompasses a range of critical actions required to convert or encode raw data into a format that can be effectively processed and understood by computer algorithms. This foundational step is essential for building accurate and reliable machine learning models. Proper data preparation ensures that the dataset is clean, consistent, and ready for analysis, which directly influences the model's ability to learn and make precise predictions. Key preprocessing tasks include filling in missing values, smoothing noisy data, resolving inconsistencies, and eliminating outliers.In the context of this dataset, which contains 6281 entries and 12 features, meticulous attention has been given to ensure there are no missing or duplicate values, resulting in a clean dataset. Data preparation also often involves integrating data from multiple sources into a comprehensive data storage facility, such as a data warehouse, to provide a unified dataset for analysis.

Feature scaling, or data normalization, is a crucial part of data preparation that standardizes the range of feature values to facilitate more effective model training. This process adjusts the scales of different features to ensure that each contributes equally to the model's performance. One of the most common normalization techniques is min-max scaling, also known as min-max normalization, which rescales feature values to a specific range, typically [0, 1].

The formula for min-max normalization is:

$$Normalized \quad Value = \frac{X - min(X)}{max(X) - min(X)}$$

In addition to min-max scaling, other normalization techniques such as Z-score normalization (or standardization) might also be applied, depending on the model and data characteristics. Z-score normalization transforms data into a distribution with a mean of 0 and a standard deviation of 1, which is useful for algorithms that assume normally distributed data. The formula for Z-score normalization is:

particularly beneficial for models sensitive to the scale of the input data.

Overall, effective data preparation and normalization are vital steps that enhance the quality of the input data, ensuring that machine learning models are trained on well-processed data and can deliver accurate, reliable results.

VI. RESULTS

CONCLUSION

This paper presented a machine learning model that can be conveniently developed for each heavy vehicle in a fleet. The model relies on seven predictors: number of stops, stop time, average moving speed, characteristic acceleration, aerodynamic speed squared, change in kinetic energy and change in potential energy. The last two predictors are introduced in this paper to help capture the average dynamic behavior of the vehicle. All of the predictors of the model are derived from vehicle speed and road grade. These variables are readily available from telematics devices that are becoming an integral part of connected vehicles. Moreover, the predictors can be easily computed on-board from these two variables. The model predictors are aggregated over a fixed distance traveled (i.e., window) instead of a fixed time interval. This mapping of the input space to the distance domain aligns with the domain of the target output, and produced a machine learning model for co2 emission with an RMSE < 0.015 l/100km. Different model configurations with 1, 2, and 5 km window sizes were evaluated. The results show that the 1 km window has the highest accuracy. This model is able to predict the actual co2 emission on a per 1 km-basis with a CD of 0.91. This performance is closer to that of physics-based models and the proposed model improves upon previous machine learning models that show comparable results only for entire long-distance trips.

REFERENCES

[1] S. Wickramanayake and H. D. Bandara, "Co2 emission prediction of fleet vehicles using machine learning: A comparative study," in Moratuwa Engineering Research Conference (MERCon), 2016. IEEE, 2016, pp. 90–95.

[2] L. Wang, A. Duran, J. Gonder, and K. Kelly, "Modeling heavy/mediumduty co2 emission based on drive cycle properties," SAE Technical Paper, Tech. Rep., 2015.

[3] Fuel Economy and Greenhouse gas exhaust emissions of motor vehicles Subpart B - Fuel Economy and Carbon-Related Exhaust Emission Test Procedures, Code of Federal Regulations Std. 600.111-08, Apr 2014.

[4] SAE International Surface Vehicle Recommended Practice, Co2 emission Test Procedure - Type II, Society of Automotive Engineers Std., 2012.

[5] F. Perrotta, T. Parry, and L. C. Neves, "Application of machine learning for co2 emission modelling of trucks," in Big Data (Big Data), 2017 IEEE International Conference on. IEEE, 2017, pp. 3810–3815.

[6] S. F. Haggis, T. A. Hansen, K. D. Hicks, R. G. Richards, and R. Marx, "In-use evaluation of fuel economy and emissions from coal haul trucks using modified sae j1321 procedures and pems," SAE International Journal of Commercial Vehicles, vol. 1, no. 2008-01-1302, pp. 210–221, 2008.

[7] A. Ivanco, R. Johri, and Z. Filipi, "Assessing the regeneration potential for a refuse truck over a real-world duty cycle," SAE International Journal of Commercial Vehicles, vol. 5, no. 2012-01-1030, pp. 364–370, 2012.

[8] A. A. Zaidi, B. Kulcsr, and H. Wymeersch, "Back-pressure traffic signal control with fixed and adaptive routing for urban vehicular networks," IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 8, pp. 2134–2143, Aug 2016.

[9] J. Zhao, W. Li, J. Wang, and X. Ban, "Dynamic traffic signal timing optimization strategy incorporating various vehicle co2 emission characteristics," IEEE Transactions on Vehicular Technology, vol. 65, no. 6, pp. 3874–3887, June 2016.

[10] G. Ma, M. Ghasemi, and X. Song, "Integrated powertrain energy management and vehicle coordination for multiple connected hybrid electric vehicles," IEEE Transactions on Vehicular Technology, vol. 67, no. 4, pp. 2893–2899, April 2018