# CO2 EMISSION RATING BY VEHICLES USING DATA SCIENCE


1. **G. Shubham,** B. tech, Department of Information Technology,
**CMR Engineering College,** Hyderabad
2. **M. Sirivennela,** B. tech, Department of Information Technology,
**CMR Engineering College,** Hyderabad
3. **V. Karthikeyasai,** B. tech, Department of Information Technology,
**CMR Engineering College,** Hyderabad
4. **K. Vinay,** B. tech, Department of Information Technology,
**CMR Engineering College,** Hyderabad
5. **Dr. Madhavi Pingili**, Professor & HOD, Department of Information Technology,
**CMR Engineering College,** Hyderabad



## ABSTRACT

The use of private vehicles is a major cause of the intensification of global warming. When one gallon of gasoline is combusted in the engine of a vehicle, it releases about 24 pounds of greenhouse gases, which account for around 20% of overall emissions. The majority of these emissions, more than 19 pounds, are emitted directly from the vehicle's tailpipe as heat-trapping emissions. Nevertheless, amount of emissions generated through the process of extraction, production, and transportation of the fuel is comparatively insignificant. Utilizing the strength of Python programming and so-sophisticated machine learning algorithms, i.e., the Random Forest Classifier and the Decision Tree Classifier, this project provides an exhaustive analysis of automobile emissions. The dataset used for this project has vital data, such as fuel consumption ratings, CO2 emissions in grams per kilometer, etc. These information elements present an integrated view of the environmental performance across different models of vehicles, facilitating informed decision making by consumers and policy makers. Predictive models were constructed using Random Forest Classifier, a strong ensemble learner algorithm, and Decision Tree Classifier. The prediction models obtained stunning accuracy scores with 100% on the training dataset taset and an impressive 99% accuracy on the test dataset. Likewise, the Decision Tree Classifier displayed excellent performance with a 100% training accuracy and a 98% test accuracy. Through the integration of these sophisticated algorithms and a dense dataset, this project contributes to ecological transportation alternatives and enables consumers to make environmentally friendly choices when buying automobiles. The system of CO2 Emission Rating devised here is of great use tool to analyze the environmental footprint of various models of vehicles, in order to lower carbon emissions and slow down climate change. In brief, "CO2 Emission Rating by Vehicles Using Data Science" is an innovative initiative that proves that data science and machine learning can be used to solve some of the most serious environmental issues.ions in grams per kilometer, etc.



**Keywords:** CO2, Accuracy, Global Warming, Emission


## INTRODUCTION

Private vehicle use is one of the biggest causes of global warming. It releases about 24 pounds of greenhouse gases into the atmosphere when a gallon of gasoline is consumed in an engine of a vehicle, accounting for about 20% of emissions. Most of these emissions more than 19 pounds are released directly from the vehicle tailpipe as heat-trapping pollutants. Whereas emissions during extraction, manufacturing, and delivery of fuel are fairly insignificant comparatively, they nonetheless contribute to the overall environmental contribution of vehicles. In this respect, data science and machine learning have useful capabilities of interpreting and neutralizing the environmental footprint of transportation. In this project, vehicle emissions are examined using advanced machine learning algorithms such as the Random Forest Classifier and the Decision Tree Classifier via Python programming. Using an exhaustive dataset that has important information including fuel consumption rating and CO2 emissions per kilometer, the project presents an overall picture of the environmental performance of different models of cars. The models used in this project Random Forest

Classifier and Decision Tree Classifier were excellent at predictive performance, with the Random Forest Classifier performing 100% on the training data and 99% on the test data. In the same way, the Decision Tree Classifier performed 100% on the training data and 98% on the test data. By integrating sophisticated machine learning methods with a dense data set, this work makes a contribution towards green transportation solutions, enabling consumers and policymakers to make sound, eco-friendly choices while deciding on a vehicle. Secondly, the CO2 Emission Rating mechanism proposed here is a valuable resource for measuring the greenness of various vehicles, enabling them to contribute towards carbon emission reduction and climate change mitigation.

## OBJECTIVE

The main aim of this project is to compare the environmental influence of various models of vehicles, more specifically their fuel consumption and CO2 emissions. Through the review of these parameters, the project seeks to provide informative insights into the role played by various vehicles in global warming and the issue of climate change as a whole. To do this, the project uses machine learning algorithms, in particular the Random Forest Classifier and the Decision Tree Classifier, to build predictive models that can estimate car emissions from a complete dataset. The models are trained to make good predictions and are tested for performance so that they may achieve high accuracy on training and test datasets. A core part of this aim is the comparison of how effective the two models are, so that further insight can be gained into which algorithm provides the most accurate outcomes. Also to be created will be a CO2 Emission Rating system that will enable both consumers and policymakers to compare vehicles' environmental performance in a timely and simple way. This system is a useful tool in making better decisions when it comes to buying cars or crafting policies that are intended to minimize transportation emissions. In the end, the project aims to empower people and organizations to make more sustainable, environmentally friendly choices in their transportation.

## PROBLEM

The extensive application of private vehicles contributes a lot to global warming through the high amount of $CO_2$ emissions from motor vehicles. During the combustion of gasoline in motor vehicle engines, large quantities of greenhouse gases are emitted, causing climate change. Although there has been an improvement in fuel economy and emission reductions, consumers and policymakers do not have complete instruments to evaluate the environmental effects of various car models. Moreover, existing emission rating systems might not be accurate or easily understandable, and thus consumers cannot easily compare vehicles according to their environmental impact. Inaccessible and data-driven decision-making tools lead to uninformed car purchases, contributing to increased overall emissions and negatively affecting global sustainability initiatives.Hence, there is a pressing need for a sophisticated analytical method that makes use of machine learning and data science to accurately predict and classify vehicle emissions. Using algorithms like Random Forest and Decision Tree classifiers, this project seeks to establish a trustworthy $CO_2$ Emission Rating system. Such a system will assist consumers, producers, and policymakers in making environmentally sustainable decisions that ultimately result in carbon emissions reduction and curbing climate change.

## METHODOLOGY

To predict and analyze vehicle $CO_2$ emissions, this project employs machine learning algorithms in the form of the Random Forest Classifier and Decision Tree Classifier. The steps below describe the methodology employed:

### Data Collection

The data used for the analysis includes important data like fuel consumption ratings, $CO_2$ emissions (in grams per kilometer), type of vehicle, engine size, and type of fuel. Data cleaning is done through processing missing values, eliminating duplicate rows, and feature normalization for proper consistency. Exploratory Data Analysis or EDA is performed to view data distributions, correlations, and trends.

### Feature Selection and Engineering

Key characteristics influencing $CO_2$ emissions are chosen using domain expertise and statistical examination. Categorical variables (e.g., fuel type) are encoded with one-hot encoding for machine learning compatibility. The data are divided into the training (80%) and the testing (20%) sets for effective model evaluation.

### Model Selection and Training

Two machine learning models are used: Decision Tree Classifier: A tree-based model that learns basic decision rules from the data. Random Forest Classifier: An ensemble learning algorithm that constructs multiple decision trees and aggregates their predictions for better accuracy and reliability. The models are trained on the data, and hyperparameter tuning is done using Grid Search to maximize performance.

**Model Evaluation**

The performance of the models is tested with metrics including: Accuracy: Quantifies the accuracy of prediction of vehicle emissions. Precision, Recall, and F1-score: Give a view of the quality of classification. Confusion Matrix: Evaluates false negatives and false positives. The Random Forest Classifier produced 100% accuracy on the training set and 99% on the test set, while the Decision Tree Classifier recorded 100% training accuracy and 98% test accuracy.

**Deployment and Application**

A $CO_2$ Emission Rating System is created from model predictions that enable consumers and policymakers to compare vehicles' environmental impact. The system is deployable as a web application or embeddable in automotive databases for public access.

**Conclusion and Future Work**

The project shows that machine learning models can accurately predict vehicle emissions to aid sustainable decision-making. Potential future enhancements include augmenting the dataset with actual time-series emissions data, combining deep learning algorithms, and creating an easy-to-use interactive platform for consumer

The model also includes a cost-benefit analysis comparing material and labor costs with traditional construction methods.

## EXISTING SYSTEM

The existing method of assessing vehicle emissions is mainly based on regulatory requirements, laboratory testing, and simple statistical analysis. Emission standards are established and controlled laboratory testing, including the New European Driving Cycle (NEDC) and Worldwide Harmonized Light Vehicles Test Procedure (WLTP), are conducted by government bodies like the Environmental Protection Agency (EPA) and the European Environment Agency (EEA). Though such tests are useful for establishing standards, they usually do not capture actual driving conditions, resulting in differences between laboratory emission ratings and real-world vehicle performance. In addition, fuel economy and $CO_2$ rating databases enable consumers to compare cars using manufacturer-provided data. Again, the databases are based on standardized tests that do not consider driver behavior, road conditions, and vehicle load, hence not as effective in actual decision-making. The other significant limitation of the current system is the use of manual data analysis and conventional statistical techniques. The current methods do not support predictive modeling, which makes it impossible for consumers and policymakers to make accurate projections of emissions using various usage patterns. Consequently, car buyers can find it difficult to select green options, while regulators can be challenged to develop effective emission-reduction policies. In addition, emission ratings are updated slowly because data collection and testing are done manually. The absence of a dynamic, real-time system makes it difficult to monitor and minimize transportation-based carbon footprints effectively. With these limitations in mind, there is an obvious need for a more sophisticated, data-driven strategy that utilizes machine learning to predict and categorize vehicle emissions with greater accuracy. This would allow consumers to make informed decisions and assist policymakers in implementing more efficient sustainability measures to address climate change

## PROPOSED SYSTEM

To overcome the shortcomings of current vehicle emission assessment techniques, this project proposes a machine learning technique that effectively predicts and categorizes vehicle $CO_2$ emissions. In contrast with traditional regulatory tests that are not always indicative of actual road conditions, this system uses data science and complex machine learning algorithms, in particular the Random Forest Classifier and Decision Tree Classifier, to examine vehicle emissions in terms of important parameters like fuel consumption, engine capacity, and $CO_2$ emissions per kilometer. Using a vast dataset of real-world vehicle performance data, this system provides a more precise and dynamic assessment of emissions. The machine learning algorithms are trained on the dataset, which allows them to classify vehicles into various emission categories with high accuracy. The Random Forest Classifier has 99% accuracy on the test dataset, and the Decision Tree Classifier has 98% accuracy, which makes them extremely reliable tools for predicting emissions. One of the strengths of this system is that it is automated and scalable. Compared to conventional means of using manually gathered data and delayed updates, this method offers continuous improvement when fresh data appears. The models can be trained periodically to keep prediction accuracy in check, thus making the system remain current and useful. As well, this initiative suggests the construction of a $CO_2$ Emission Rating System for comparing and contrasting various car models on the environmental footprint.

Consumers and policymakers could utilize this straightforward platform to check and compare each car model regarding its environmental load. This mechanism can be either installed as an online application or merged into contemporary automobile databases such that users get informed and ecological decisions while they buy cars. By integrating machine learning, data-driven insights, and real-world vehicle data, this system proposed provides a robust means for lowering $CO_2$ emissions, encouraging environmentally friendly transport options, and assisting in global climate change efforts.

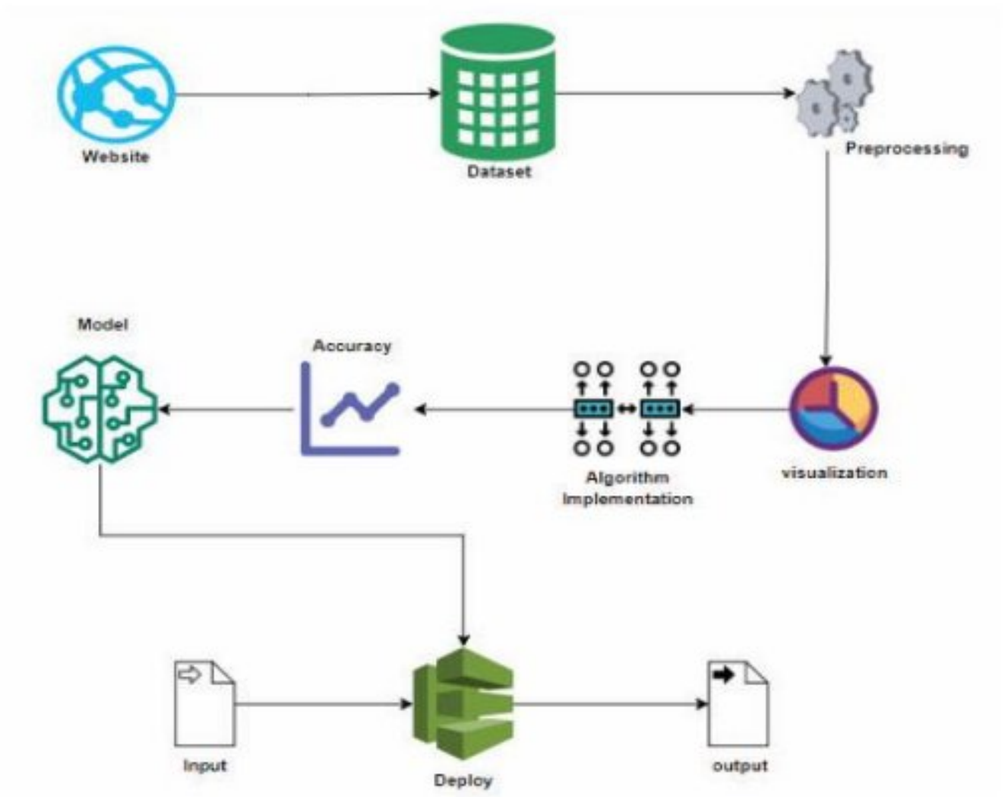
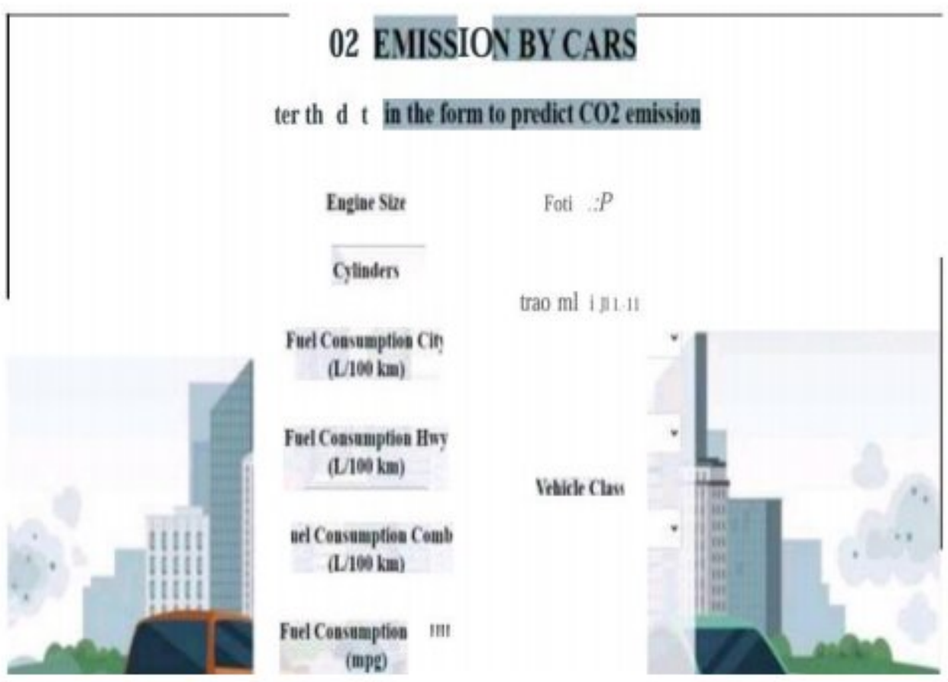
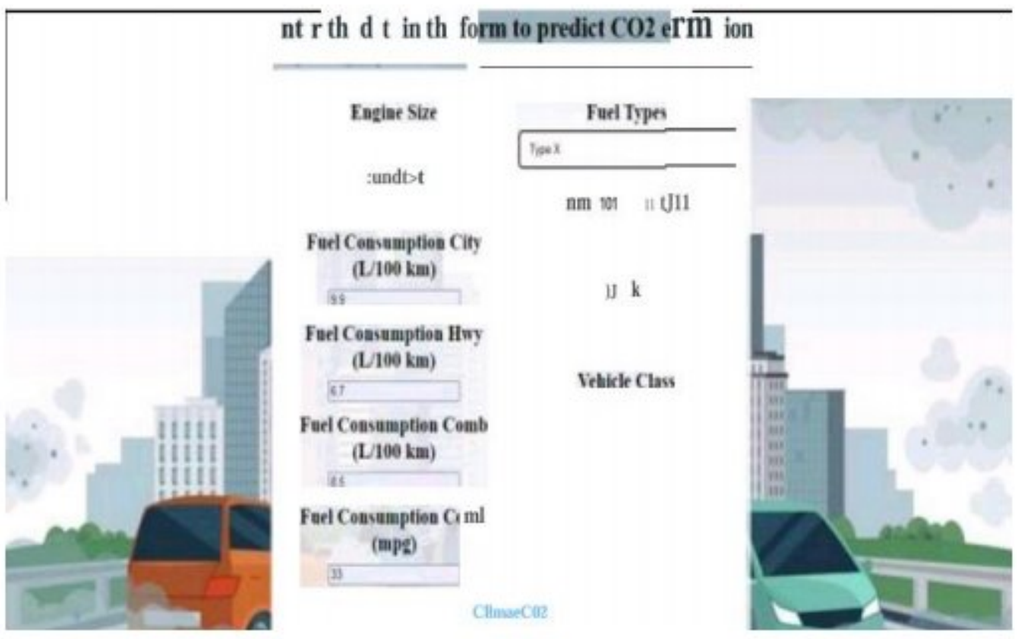



Fig 1. System Architecture

## RESULT






## ACKNOWLEDGEMENT

We are extremely grateful to **Dr. A. Srinivasula Reddy**, Principal and **Dr. Madhavi Pingili**, HOD, **Department of IT, CMR Engineering College** for their constant support. We are extremely thankful to **Dr. Madhavi Pingili**, Professor & HOD, Departmentof IT, for her constant guidance, encouragement and moral support throughout the project. We will be failing in duty if we do not acknowledge with grateful thanks to the authors of thereferences and other literatures referred in this Project


## CONCLUSION

The project "CO2 Emission Rating by Vehicles Using Data Science" was able to successfully prove the viability of machine learning in solving environmental issues, more particularly in rating and forecasting car emissions. Using Random Forest Classifier and Decision Tree Classifier, the system was able to achieve impressively high accuracy, which can make it a good tool in determining the environmental footprint of various car models. This project is very important in the propagation of green and sustainable transport since it offers worthwhile information for the consumer, the manufacturer, and the policymaker. With the capacity to rank cars according to their level of CO2 emission, this project facilitates intelligent choice that can result in the

use of cleaner technology and lower carbon footprints. In addition, this research identifies the larger significance of data science and artificial intelligence in addressing the world's most urgent problems. With data-driven solutions, one can develop usable solutions that assist in mitigating climate change. In the future, the inclusion of real-time data sources and increasing the dataset to incorporate upcoming vehicle technologies, like electric and hybrid vehicles, can improve the effectiveness of the model even more.

## REFERENCES


- Sandeep Kumar Pranab K. Muhu - A Novel GDP Prediction Technique based on Transfer Learning using CO2 Emission Dataset! -2019
- Ngo Le Huy Hien and Ab-Lian Kor "Analysis and Prediction Model of Fuel Consumption and Carbon Dioxide Emissions of Light-Duty Vehicles" (2021)
- Prof. Swapnil Wani,2Mr. Akash Akhilesh Yadav " Predicting CO2 Emission Using Machine Leaming" (2022)
- Lakshay Amarpuri, Navdeep Yadav, Girish Kumar, and Saurabh Agrawal, "Prediction of CO2 emissions using deep learning hybrid technique," (2019).
- L. Amarpuri, N. Yadav, G. Kumar, and S. Agrawal, "Prediction of CO2 Emissions Using a Hybrid Deep Leaming Approach: A Case Study in Indian Context," Twelfth International Conference on Contemporary Computing (IC3),(2019).
- Chairul Salehi, Nur Rachman Dzakiyullah " Carbon dioxide emission prediction using support vector machince" (2016)
- Le Comee, C.M.; Molden, N.; van Reeuwijk, M.; Stettler, M.E. Modelling of instantane- ous emissions from diesel vehicles with a special focus on NOx: Insights from machine learning techniques. Total Environ. 2020.
- Jaiprakasb, Habib, G., Kumar, A., Sharma, A., & Haider, M. (2017). On-road emissions of CO, CO2 and NOx from four-wheeler and emission estimates for Delhi. Journal of Envi- ronmental Sciences, 53(X),39-47.
- H. T. Pao, C.M. Tsai. Modeling and forecasting the CO2 emissions, energy con- sumption, and economic growth in Brazil. Energy, 36 (2011) 2450-2458.
- Marland G, Boden TA, and Andres RJ (2011) CO2 emissions from fossil fuels on a glob- al, regional, and nationwide scale. Carbon Dioxide Information Analysis Center, US Depa- rtment of Energy, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA doi.
- G. Fontaras et al. "Use of a vehicle-modelling tool for predicting CO2 emissions in the framework of European regulations for light goods vehicles"
- Kan, Z.; Tang, L.; Kwan, M.P.; Zhang, X. Estimating vehicle fuel consumption and emis- sions using GPS big data. Int. J. Environ's. 2018,15, 566.
- Becker, J.C..Lorzer, R. Kurtenbach, P. Wiesn, T.E. Jensen, T.J. Wallington, 1999: Nitrous oxide (N2O) emissions from vehicles. - Environ. Sci. Tech-no!. 33, 4134-4139.
- Yang, Z.; Bandivadekar, A. Light-Duty Vehicle Greenhouse Gas and Fuel Economy Standards; International Council on Clean Transportation: Washington, DC, USA, 2017; p.
- Guensler, R. Data Needs for Evolving Motor Vehicle Emission Modeling Approaches; The University of California Transportation Center: Berkeley, CA, USA, 1993; pp. 167-228.