A Flask-Based Web Application To Predict Co2 Emission In Vehicles Using Ml Techniques

[1]Dr. K. Mohana Prasad
Professor, Department of CSE
Sathyabama Institute of Science and Technology
Chennai ,600119,
MohanaPrasad1983@gmail.com,

[2]Aravind K
Professor, Department of CSE
Sathyabama Institute of Science and Technology
Chennai ,600119
arvindkamalay02@gmail.com,

[3]Arjun Singh Maru
Professor, Department of CSE
Sathyabama Institute of Science and Technology
Chennai ,600119
arjunsingh29102001@gmail.com

***Abstract* — Carbon Dioxide and other gases absorb sunlight and solar rays that have previously been mirroring off the earth's surface as they accumulate in the atmosphere, causing global temperatures to rise. As a result, air pollution has advanced in several inhaling disorders and cardiac diseases among humans. Air pollution is also causing many effects on the international world by affecting soil fertility, air quality, and water quality. Car Pollution also forces the animals to abandon their habitat and move to a new place.**

Passenger Vehicles also emit other gas pollutants, including nitrogen dioxide, carbon monoxide, and formaldehyde, that harm the global environment. Noise levels from vehicles due to the increasing city traffic also cause many hearing problems and psychological ill-health. One of the most challenging parts of the energy transition is lowering CO2 emissions from the transportation sector. Data is the critical element that enables algorithm training the most. With data, machine learning is more manageable for AI systems to perform. We use regression models for the prediction of the emission of CO2 from cars. The data, such as car details and features collected from the literature resources, are given as input to the ML model. The ML model predicts the amount of CO2 emitted from the car. The Road Transport Authority staff notifies the registered car owner to service the car if the estimated CO2 emission level is within the threshold value. This application program produces better outcomes by using many features of the car like fuel consumption, fuel transmission, engine size etc., as input and offers 24x7 service availability around the clock through internet connection to predict CO2 emission level.

Keywords. Carbon dioxide, Machine Learning, Flask, Random Forest Regression, Dataset.

I. INTRODUCTION

Our private vehicles primarily cause global warming, mainly climatic change and air pollution. Collectively, cars produce about 15% of the world's total carbon dioxide (CO_2) emissions. One liter of gasoline produces around 2.5 kilograms of CO_2 from a car's exhaust pipeline. Therefore, the average passenger car emits about 252 grams of CO_2 per Kilometre from its exhaust pipeline. The study says that a median gasoline car on the highway has a fuel efficiency of around 10 km per liter and steers around 18,500 km per year. Therefore, a typical passenger car emits about 2420 grams of CO_2 from one liter of gasoline. Since 2000, the Governments of respective countries have taken the initiative to reduce the CO_2 emission by 35% from brand-new cars sold using New European Driving Cycle (NEDC) test procedure. In 2000, the CO_2 emission from new cars was around 172 g/km. Today they are still 120 g/km and unlikely to reach 100g/km until 2025.

Passenger cars produced carbon dioxide emissions from 2.2 billion metric tons in 2000 to a peak of 3.2 billion metric tons in 2020. An average passenger car releases about 4.7 metric tons of CO_2 gas annually. Considerably, the reduction of CO_2 from vehicles will be challenging. Still, the data that is already accessible may be utilized to extract the controlled and uncontrolled features of cars that influence the CO_2 properties. The model, fuel type, transmission, and many other factors all contribute to the CO_2 emission level from passenger cars. Therefore, we should take as many pertinent features as possible to improve our predicted accuracy of CO_2 from vehicles—the lesser independent parts results in acquiring high prediction accuracy. By using the attributes of a car, we can develop a machine-learning model for CO_2 emissions from that particular car. We put a minimum threshold value of CO_2; if the vehicle exceeds the threshold value, Road Transportation Authority (RTA) staff notifies the owner to service his car.

II. LITERATURE REVIEW

Over the last decades, an increase in cellular communication technologies implicit that web application programs can be extensively accessible and portable. This is the primary incentive for the demand for Machine Learning established prediction models. Further, with advancements in deep learning and machine learning, we can use the On-Board Diagnostic II sensor data-based methodology for

A. K. Visvam Devadoss et al. (eds.), *Proceedings of the 6th International Conference on Intelligent Computing (ICIC-6 2023)*, Advances in Computer Science Research 107,
https://doi.org/10.2991/978-94-6463-250-7_12

automobile CO2 emission levels. A brief literature review of some of the related work done follows:

The authors in [3] have used ensemble learning techniques and gradient boosting algorithms to develop the urban area's co2 emission level prediction model. Further, they have used several details of the car, like the model, speed, transmission, and cylinder size, as input data for the model. For building the model, the authors have used gradient boosting regression to predict CO2 emission. Future research indicated focusing on car fuel consumption as another attribute to make the system more accurate in all situations. In this work, we have attempted to add the fuel consumption of the vehicle attribute as input to the model and prediction the amount of CO2 emitted from the particular vehicle.

In [1], authors have used a regression model to predict CO2 emission levels based on historical data from the world bank from 1960 to 2014. The dataset includes living organisms, fermentation, natural gas, trees, and solid waste to predict CO2 emission. Future research focused on using a regression model algorithm to predict CO2 emission from vehicles. In this work, we have proposed to predict CO2 emissions from cars by using regression models.

The above literature review depicts that the present CO2 estimation and machine learning models depend mainly on vehicle transmissions and cylinder size as input data to the model. Therefore, the prediction accuracy could be higher because fewer vehicle attributes are used as the input data for the prediction. On the other hand, the proposed model uses more vehicle features like vehicle consumption and regression models for the CO2 emission level prediction. In addition, the model considers diverse parameters for the prediction, like a model, vehicle class, fuel consumption, etc. Further, the model that is created is integrated into a web application. So that non-technical people may use it quickly and easily to determine the quantity of CO2 emission emitted from their cars.

III. METHODOLOGY

A. Data Accumulation

The dataset is acquired through online platform resources like Kaggle and datahub.io to initiate the project. The dataset comprises 679 rows and 12 columns of attributes. The attributes consist of data like vehicle class, engine size, cylinder, transmission, etc... The data collected is stored in an excel document as a comma-separated value (.csv format). The machine learning model will be trained and tested based on the dataset.

B. Data Visualization

Data visualization is the process of displaying a specific data collection graphically. It helps identify patterns, correlations, and trends that could be missed in text-based data. Like with any approach, understanding your data and its relationships is essential to building a machine-learning model. In truth, even the most sophisticated machine learning algorithms must improve when dealing with poorly presented and understood data. Therefore, we use libraries like seaborn, pyplot, and matplotlib for visualizing data in a graphical format.

Let's see the connection between the dependent and independent attributes of our dataset by using a seaborn bar plot plotting the graph.s

The following visualization represents the relationship between vehicle class and CO2 emissions. Again, we can see a slight variation in vehicle class over CO2 emission.

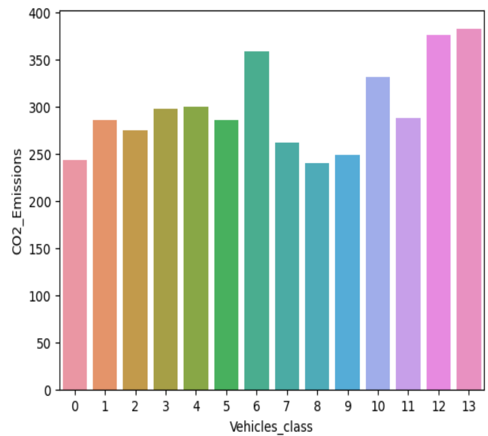

FIGURE 1. Barplot Graph between CO2 Emission and Vehicles Class

The following visualization represents the relationship between making and CO2 emissions. Again, we can see a slight variation in make over the CO2 emission.

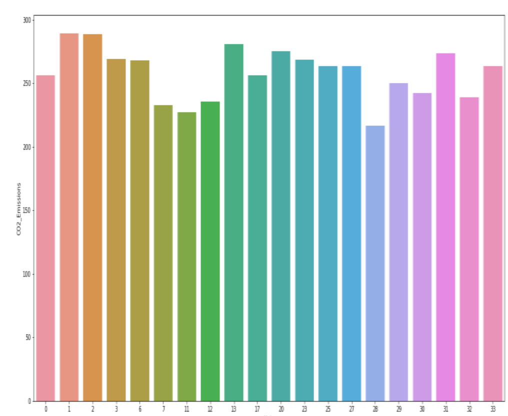

FIGURE 2. Barplot Graph between CO2 Emission and Make

But, finding the relationship between several characteristics is the goal of multivariate analysis. A heatmap from the Seaborn package was utilized in this instance. An arithmetical technique known as a correlation matrix expresses how two attributes are connected. It's essential for expressing direct relations without explicitly declaring cause and effect. The correlation heat map() process is used to visualize. The correlation matrix also helps to find which variables are strongly dependent and weakly reliant on each other. We can quickly identify the strongly associated feature in the image below. The negative correlation is changed to a positive correlation using the Abs() technique.

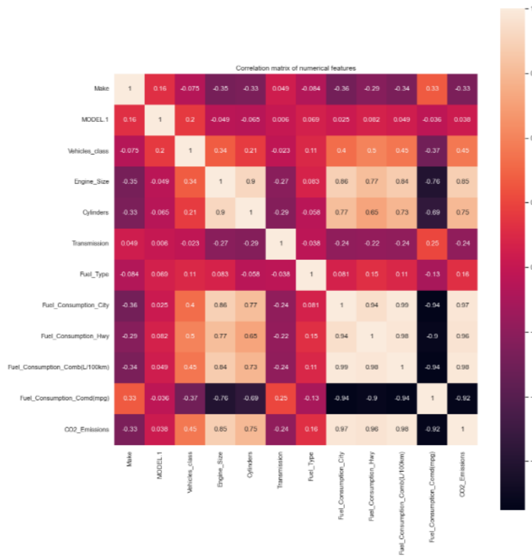

FIGURE 3. Correlation Matrix of Passenger Cars Feature set

We can identify which elements are significantly associated and related to one another every week from the correlation plot shown above for the dataset used to predict CO2 emission.

C. Data Pre-processing

The pre-processing data method transmute raw data into an understandable format that computer systems and machine learning models can comprehend and access. In addition, it is a method for data cleaning and data analysis. In data pre-processing, we handle the null values, take the categorical values if any are present, normalize the data if required, identify the dependent and independent attributes, and split the dataset into train and test.

The dataset may be utilized for developing the model. Yet some data is needed to test the model. Hence, a second set for testing is required. The original data can be tested when a separate test set is unavailable. The dataset is divided into sections for testing, and the remaining portions are used for training. Using the sci-kit-learn library's train_test_spilt, the splitting is carried out. Eighty percent of the accumulated data will be used for training and 20 percent for testing. Supplying the option random state, one may split by selecting random rows.

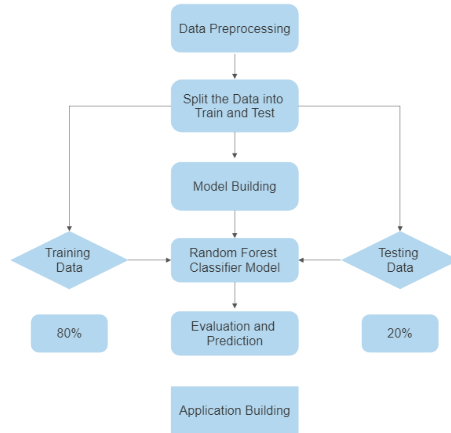

FIGURE 4. Block Diagram of the Application Building

D. Model Building

Depending on the data you will process, you may use various Machine Learning approaches—such as images, audio, text, and numerical numbers. In addition, one can choose the algorithms depending on the objective because our dataset is a regression dataset.

We use a Random Forest Classifier algorithm for the prediction of CO2 emission. We also evaluate the model using metrics like Root Mean Squared Error, R-Squared Score, and Mean Squared Error.

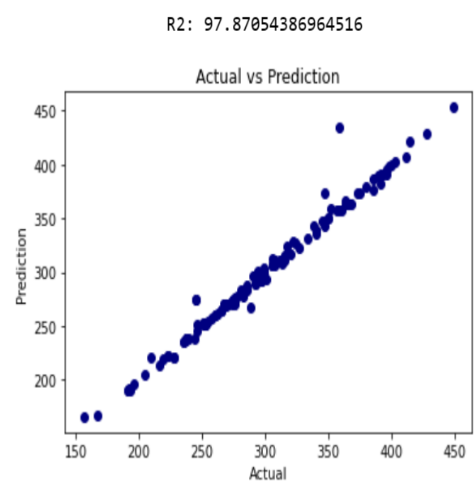

FIGURE 5. Scatterplot of Actual and Predicted CO2 emission

E. Application Building

When the model is created, it will be integrated into a web application so that non-technical people can quickly and easily determine the quantity of CO2 emitted. We use the flask module, a python web framework for building application programs.

Creating an HTML page is essential because that will send the flask application the user's nine input values. Then, the result must be taken from the flask and displayed to the user. In this guided project, we've developed three HTML pages: a landing page, a page for gathering user input, and a page for displaying the results on the screen.

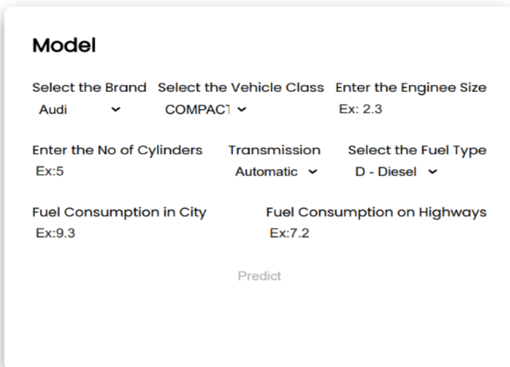

FIGURE 6. Web Application for Prediction of CO2 Emission

IV. RESULTS AND DISCUSSION

Once the model is built using the random forest classifier algorithm, the dataset can be trained to get appropriate predictions. We split the data into train and split. For training the dataset, we use 80 % of the collected data; for testing the accuracy of the model we use 20 %. We use multiple evaluation metrics for evaluating the performance of the model. R2, MAE, RMSE, and MSE valuation metrics are used to acquire the built model's prediction accuracy. 97% of the accuracy is achieved for the built random forest classifier model by using maximum features of a car as input to the model. Brand, Vehicle Class, Engine Size, Cylinders, Transmission, Fuel Type and Fuel Consumption of a car are the parameters utilized for producing 97% accuracy. The built random forest model is saved by importing the pickle library. Finally, the created model is integrated into a web application so that non-technical people may use it quickly and easily to determine the quantity of CO2 emission.

V. CONCLUSION AND FUTURE WORK

The application is fast, protected, portable, easily accessible, and provides 24x7 service availability around the clock through internet connection to predict CO2 emission level. The Road Transport Authority gives car details of a particular car to service the vehicle. In this way, we can reduce the air pollution generated by car tailpipes. In the future, we could upgrade this model by adding more attributes. The project can be upgraded by adding different transport vehicles like trucks and vans into the dataset by using online resources. We can also extend the project using a new On-Board Diagnostic II sensor data established methodology for automobile CO2 emission level monitoring.

VI .REFERENCES

[1] P. Kadam and S. Vijayumar, "Prediction Model: CO2 Emission Using Machine Learning," *2018 3rd International Conference for Convergence in Technology (I2CT)*, Pune, India, 2018, pp. 1-3, doi: 10.1109/I2CT.2018.8529498.
[2] S. Kangralkar and R. Khanai, "Machine Learning Application for Automotive Emission Prediction," 2021 6th International Conference for Convergence in Technology (I2CT), Maharashtra, India, 2021, pp. 1-5, doi: 10.1109/I2CT51068.2021.9418152.
[3] N. Subramaniam and N. Yusof, "Modelling of CO2 Emission Prediction for Dynamic Vehicle Travel Behavior Using Ensemble Machine Learning Technique," *2021 IEEE 19th Student Conference on Research and Development (SCORED)*, Kota Kinabalu, Malaysia, 2021, pp. 383-387, doi: 10.1109/SCOReD53546.2021. 9652757..
[4] Zeng, Weiliang et al. "Prediction of vehicle CO2 emission and its application to eco-routing navigation*." Transportation Research Part C-emerging Technologies* 68 (2016): 194-214.
[5] Chairul Saleh *et al* 2016 *IOP Conf. Ser.: Mater. Sci. Eng.* 114 012148 **DOI** 10.1088/1757-899X/114/1/012148.
[6] Breiman, Leo, Friedman, J. H., Olshen, R. A. and Stone, C. J. (1984). Classification and regression trees. Monterey, CA: Wadsworth and Brooks/Cole Advanced Books and Software. (Google citation: 37373)
[7] A. K. Agarwal and N. N. Mustafi, "Real-world automotive emissions: Monitoring methodologies and control measures*," Renewable and Sustainable Energy Reviews*, vol. 137, 2021.
[8] James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). An Introduction to Statistical Learning (Vol. 112). New York: Springer. (Google citation: 2673)
[9] Dr. K. Mohana Prasad," Efficient Dynamic Clustering Mechanism through Unfathomed Clustering Techniques," Journal of Advance Research in Dynamical & Control Systems, Vol. 10, 06-Special Issue, 2018, pp. 934-939.
[10] K. Mohana Prasad, Dr. R. Sabitha," Providing cluster categorization of heuristics Technique for increasing accuracy in severe categorization of road accidents" in Proc International Conference on Communication and Signal Processing (ICCSP), ISBN: 978-1-5090-3800-8, 08 February 2018.
[11]Philo Stephy, A., Preethi, C., Mohana Prasad K ,"Analysis of vehicle activities and live streaming using IOT", International Conference on Communication and Signal Processing (ICCSP),DOI: 10.1109/ICCSP.2019.8698001,2019.
[12] K. Mohana Prasad, Dr. R. Sabitha," An Efficient Clustering Formulation from Resemblance in Extant Algorithms, ARPN Journal of Engineering and Applied Sciences, ISSN 1819-6608, VOL. 11, NO. 13, JULY2016, pp. 8278-8283.
[13] Mohana Prasad, K., Sai Nagendra Goru, R., Vamsi, D. "Automated Payroll Using GPS Tracking and Image Capture "IOP Conference Series: Materials Science and Engineering 590 (2019) 012026, IOP Publishing, doi:10.1088/1757-899X/590/1/012026.
[14] Mohana Prasad, K., Dhar, P., Naseem "Industrial Automation with Bidirectional Visitor Counter" IOP Conference Series: Materials Science and Engineering, 590 (2019) 012012, IOP Publishing, doi:10.1088/1757-899X/590/1/012012.
[15] Mohana Prasad, K., Sri Kavya, R., Bhuvaneswari Devi, S. Virtual Fitting Space for Dress Trials", IOP Conference Series: Materials Science and Engineering,590 (2019) 012013, IOP Publishing, doi:10.1088/1757-899X/590/1/012013.
[16] R. Vignesh, K. Mohana Prasad," Cloud-Implementation of E-Healthcare Framework", International Journal of Recent Technology and Engineering (IJRTE), ISSN: 2277-3878, Volume-8 Issue-3, September 2019.
[17] K. Mohana Prasad, Manalina," Effective Clusters Culled Out Through Algorithmic Implementations, ARPN Journal of Engineering and Applied Sciences, ISSN 1819-6608, VOL. 11, NO. 9, MAY 2016, pp. 5574-5579.
[18] K. Mohana Prasad, Jessie Monica. J" Innovative Approach on Algorithmic Implementation for Effective Clusters," International Journal of Applied Engineering Research, ISSN 0973-4562 Volume 10, Number 4 (2015) pp. 10087-10094.
[19] K. Mohanaprasad, A. Cyril Cubert," Fuzzy Means Clusterisation of Land Based on User Priority," International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 9 (2015), pp 8166-8170.

[20] K. Mohanaprasad, Seetharaman," Modified Incremental Affinity Propagation Clustering Based on Client-Server Access Transfer" International Journal of Control Theory and Applications, Volume 9, Number 2 (2016), pp. 1063-1069.