CO2 EMISSION PREDICTION FROM VEHICLES

Submitted by

ABHINAND K (223001)

ABHIRAM B NAIR (223002)

ABHISHNAV M (223003)

ADARSH P S (223004)

In partial fulfilment of the requirements for the award of Master of Science

in Computer Science with Specialization in Data Analytics

Of

School of Digital Sciences

Kerala University of Digital Sciences, Innovation, and Technology

(Digital University Kerala)

Technocity Campus, Thiruvananthapuram, Kerala – 695317

July 2023

BONAFIDE CERTIFICATE

This is to certify that the project report entitled "**CO2 Emission Prediction from Vehicles"** submitted by

Name	**Reg No:**
ABHINAND K	**223001**
ABHIRAM B NAIR	**223002**
ABHISHNAV M	**223003**
ADARSH P S	**223004**

in partial fulfilment of the requirements for the award of **Master of Science in Computer Science with Specialization in Data Analytics** is a Bonafide record of the work carried out at "**Digital University Kerala"** under our supervision.

Supervisor	Course Coordinator
NAME	NAME
DESIGNATION	DESIGNATION
DUK	DUK

Head of Institution

Prof. Saji Gopinath
Vice Chancellor
DUK

DECLARATION

We, Abhinand K, Abhiram B Nair, Abhishnav M, Adarsh P S, *students* of **Master of Science in Computer Science with Specialization in Data Analytics**, hereby declare that this report is substantially the result of my own work, except, where explicitly indicated in the text and has been carried out during the period **March 2023-July 2023**.

Place: TRIVANDRUM

Date:09/09/2023

ACKNOWLEDGEMENT

We wish to extend our heartfelt and profound gratitude **to Dr. T.K Manoj Kumar**, **Associate Professor at Digital University Kerala** in Trivandrum, for his invaluable guidance, expert advice, and unwavering support. It is through his mentorship that we were able to successfully complete this project as a team.

We would also want to express our sincere appreciation to **Prof. Saji Gopinath** for giving us access to a favourable atmosphere, insightful advice, and educational resources that improved my capacity to take on a project of this size

CO_2 EMISSION PREDICTION FROM VEHICLES

TABLE OF CONTENTS

ABSTRACT

The rapid increase in the carbon dioxide emission (CO2) emerged as one of the critical global challenges leading to drastic climatic changes and environmental disruption. As we know one of the major sources for the carbon dioxide is from the combustion of fuels in vehicles. In this study, we are trying to analyse the carbon dioxide emission from various vehicles and finding the underlying internal factors that affects.

The study begins by collecting data of carbon dioxide emission from various vehicles including wide variety of factors like Fuel type, Engine size, Cylinders and Transmission type, Fuel Consumption etc. Data is prepared, analysed and visualized to find the various relationships that connects to carbon dioxide emission and identifying trends over them. Trained the data with various regression models like Lazzo, Ridge, Linear, Decision Tree etc. Furthermore, in this project we created a predictor model which predicts the amount of carbon dioxide that emits from vehicles by taking some input values like Engine size, Fuel type, Fuel Consumption etc.

From our study we found that the fuel type, engine size, fuel consumption, transmission type affects the level of carbon dioxide emission from vehicle. These factors affect the CO2 level drastically and visualized the impact that brings on final output. Also build a model which predicts the final CO2 amount with an accuracy of 90 percent.

INTRODUCTION

"Human activities have raised the atmosphere's carbon dioxide content by 50% in less than 200 years"

----- NASA

As the industrialization gained momentum in the past few decades and it results in the use of various natural resource as the new source of energy/fuel. Carbon Dioxide has become a major threat since last three decades as their emission has increased to a large extent. It led to many climatic changes and environmental hazards. Human activities have a major role on this rapid increase in the amount of CO2 value. The major source of carbon dioxide is from the combustion of fuels and mainly from the vehicles as the product of combustion of these fuels is carbon dioxide. Transportation sector particularly road transportation releases a huge amount of CO2 to the atmosphere. In fact, we all know these things and, in our study, we focus on the vehicular characteristics and how often they affect CO2 emission.

"Delhi Imposes Ban on Non-BS VI-Diesel Vehicles Under GRAP To Tackle Air Pollution"

----- TIMES OF INDIA

Air pollution has become a major concern in many parts of the country including the capital city. Not only the engine but also the entire vehicular characteristics influences the final amount of carbon dioxide produced. Fuel consumption, Engine size, Transmission type (Automatic and Manual) etc. affects the final emission quantity. With these all factors in mind, this study not only analyse how these factors affects the Carbon Dioxide emission but also develop a predictive model which predicts the amount of carbon dioxide that is produced from vehicle when the other requirements are provided. Furthermore, this study can lead to identify the specific factor that produces the more emission and can make a restriction or a guideline for that factor to be used in a vehicle.

LITERATURE REVIEW

I. Title: "A Machine Learning Approach for Predicting Carbon Dioxide Levels"

This paper by Smith et al. explores the application of machine learning techniques to predict CO2 levels. The authors collected historical CO2 data from various sources and used regression models, including linear regression and decision trees. They found that decision tree-based models outperformed traditional statistical methods in terms of accuracy. However, the study could benefit from a more extensive dataset and evaluation of additional machine learning algorithms.

II. Title: "Time Series Forecasting of CO2 Emissions Using Recurrent Neural Networks"

In this paper by Chen and Wang, the authors employ recurrent neural networks (RNNs) for time series forecasting of CO2 emissions. They utilized a large dataset of historical CO2 emissions and weather data. The results show that RNNs can capture temporal dependencies effectively, leading to improved prediction accuracy compared to traditional methods. However, the paper does not delve into the interpretability of RNNs.

III. Title: "Hybrid Forecasting Model for Urban CO2 Concentrations Combining ARIMA and Long Short-Term Memory Networks"

This study by Kim et al. presents a hybrid forecasting model that combines autoregressive integrated moving average (ARIMA) and Long Short-Term Memory (LSTM) networks for urban CO2 concentration prediction. The hybrid approach leverages the strengths of both methods, with ARIMA handling short-term patterns and LSTM capturing long-term dependencies. The results demonstrate enhanced prediction accuracy compared to using either method individually, making it a promising approach for urban CO2 prediction.

IV. Title: "Predicting Carbon Dioxide Emissions from Power Plants Using Support Vector Machines"

In this paper by Li and Liu, the authors focus on predicting CO2 emissions specifically from power plants using support vector machines (SVMs). They collected data on power plant characteristics and operational parameters. SVMs were applied to build predictive models. The study found that SVMs provided accurate predictions and could help power plants optimize their emissions. However, the paper could benefit from considering the integration of additional features, such as environmental factors.

These four papers collectively demonstrate the application of various machine learning techniques for CO2 prediction. They highlight the importance of data quality, model selection, and hybrid approaches for achieving accurate and robust predictions. Incorporating insights from these studies into your own ML project on CO2 prediction can help inform your methodology and approach.

DATA ANALYSIS

ABOUT DATASET

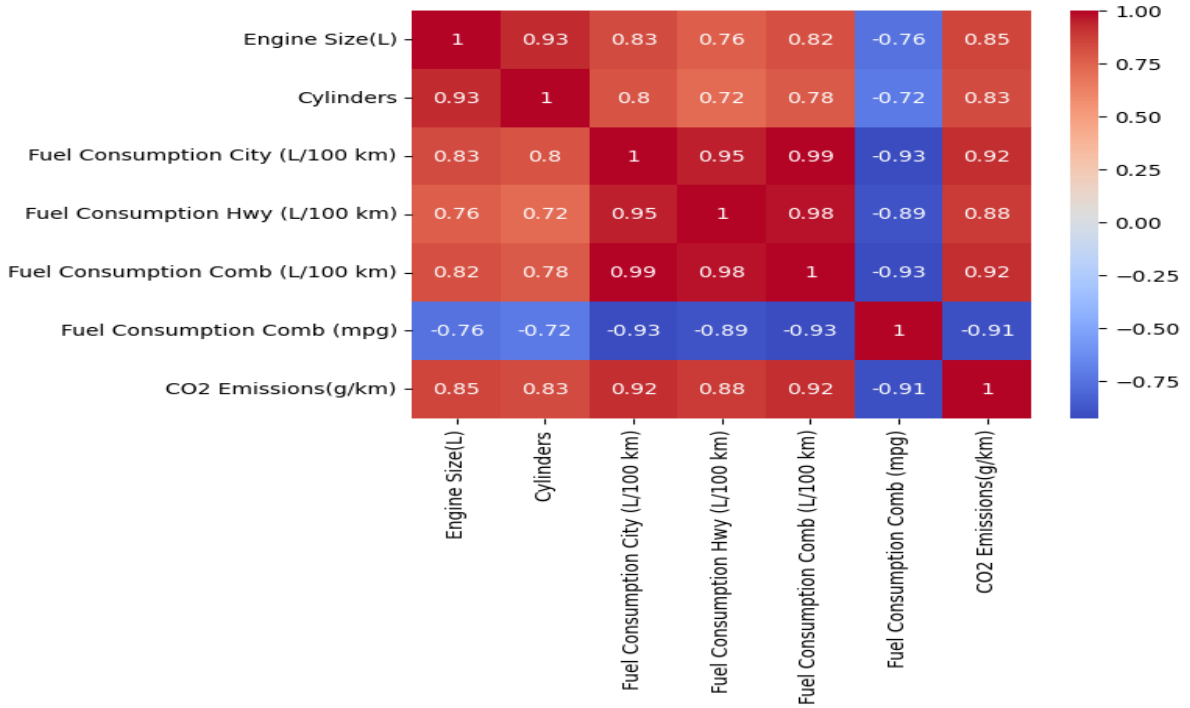
The dataset utilized in this project plays a pivotal role in uncovering patterns related to carbon dioxide emissions and their relationship with various features. This dataset provides insight into how vehicle emissions of carbon dioxide vary in conjunction with other factors. It has been sourced from the official open data website of the Canadian Government, ensuring its credibility. Spanning a period of 7 years, this dataset comprises 7385 rows and encompasses 12 distinct features or columns. It's worth noting that there are instances of missing values and duplicated entries within the dataset. The data has been acquired and organized from the following official Canadian Government link.

https://open.canada.ca/data/en/dataset/98f1a129-f628-4ce4-b24d-6f16bf24dd64#wb-auto-6

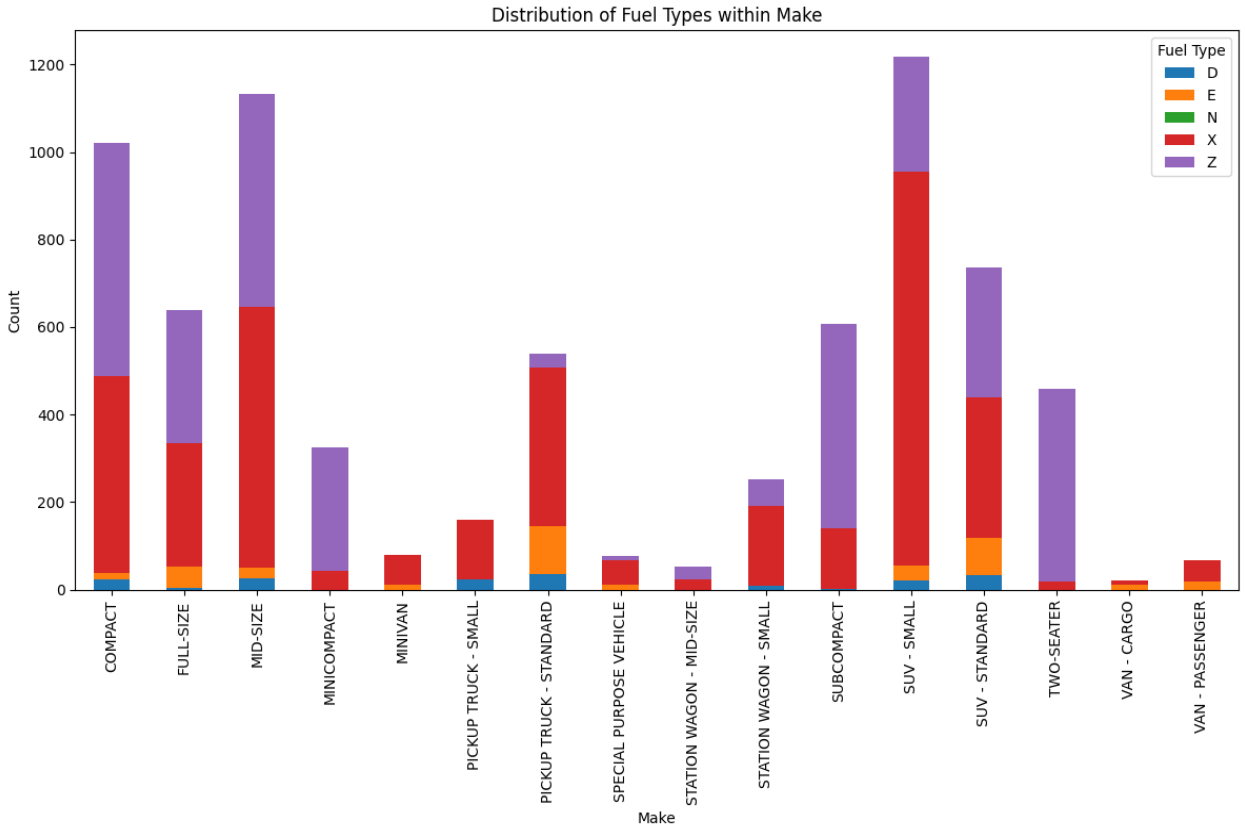
EXPLORATORY DATA ANALYTICS

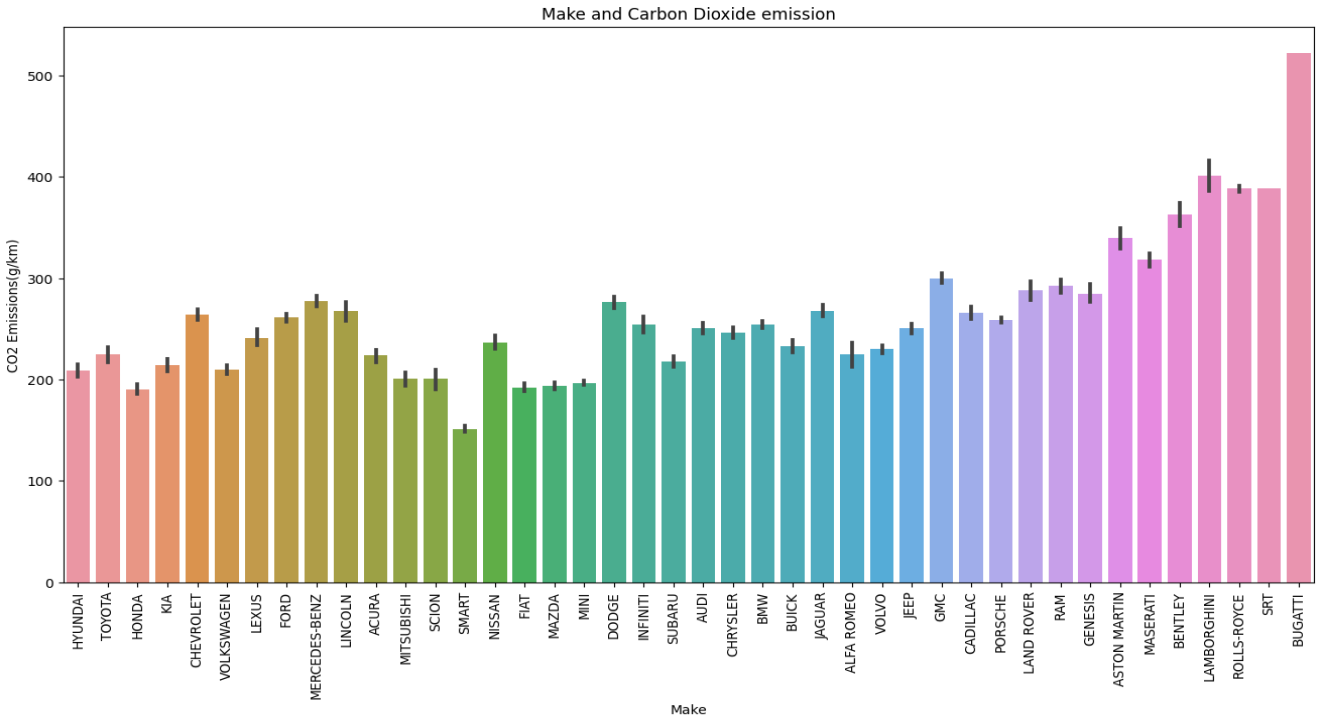
We have done various data cleaning steps like removing outliers, filling the null values etc. and done visualization among the features and helps in finding the underlying patterns and relation between them.

First of all, we draw a heatmap for checking the correlation between features. From the analysis, it is clear that engine size and cylinders are highly related to each other. Also, there is an interesting fact that Fuel Consumption Comb(mpg) have a negative correlation with all the other features. Like as the Fuel Consumption Comb(mpg) increases CO2 Emissions decreases.

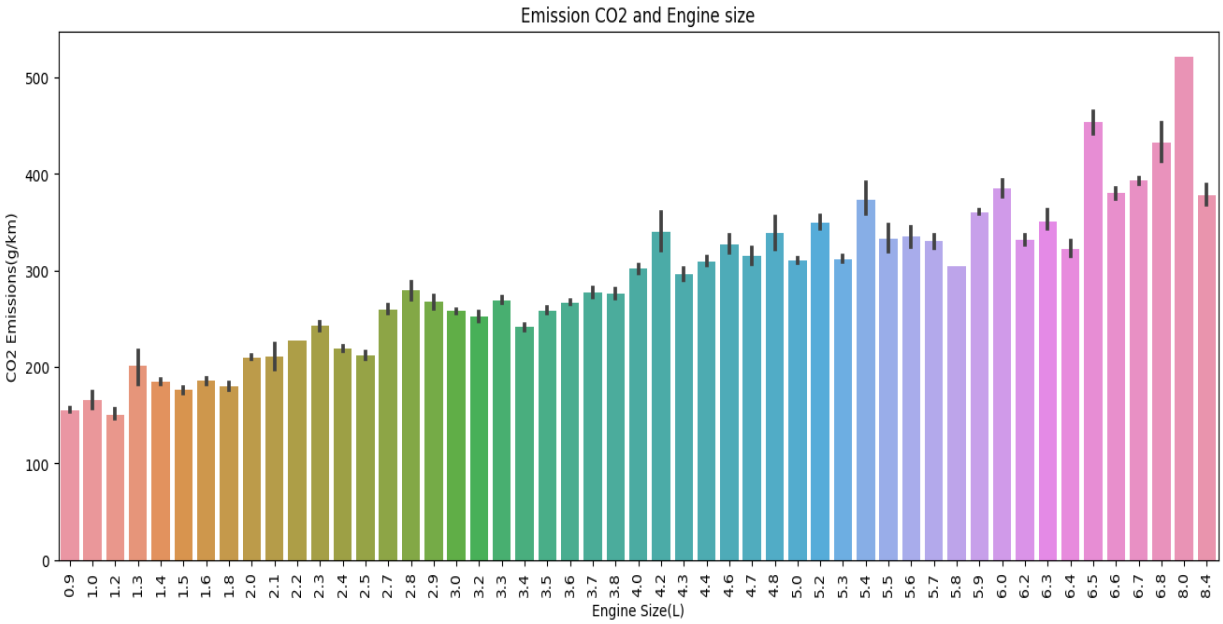


A Stacked bar chart is constructed with the Vehicle class and the fuel type. This helps to find which type of fuel is commonly in a specified vehicles class. Most the vehicles belong to the SUV class and mid-size. Regular gasoline and premium gasoline are mostly used in vehicles.

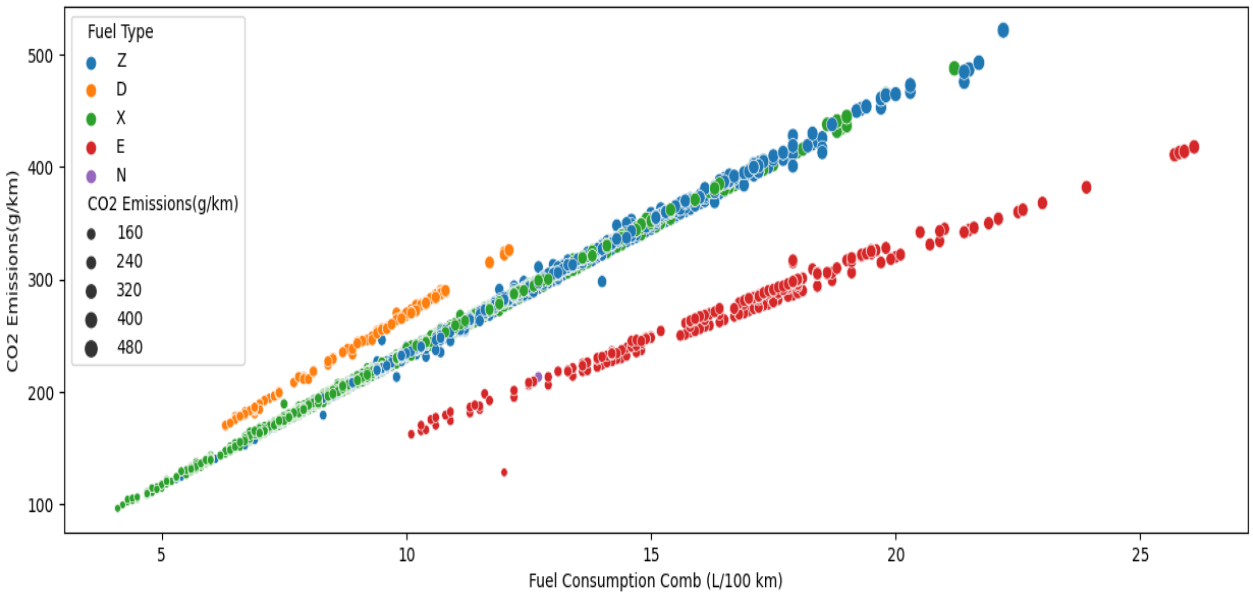




We also find that the vehicles from the brand BUGATTI making more Co2 emission with respect to the other manufactures. It produces an emission of almost 500 g/km while the majority of others shares a range of 170 – 300.

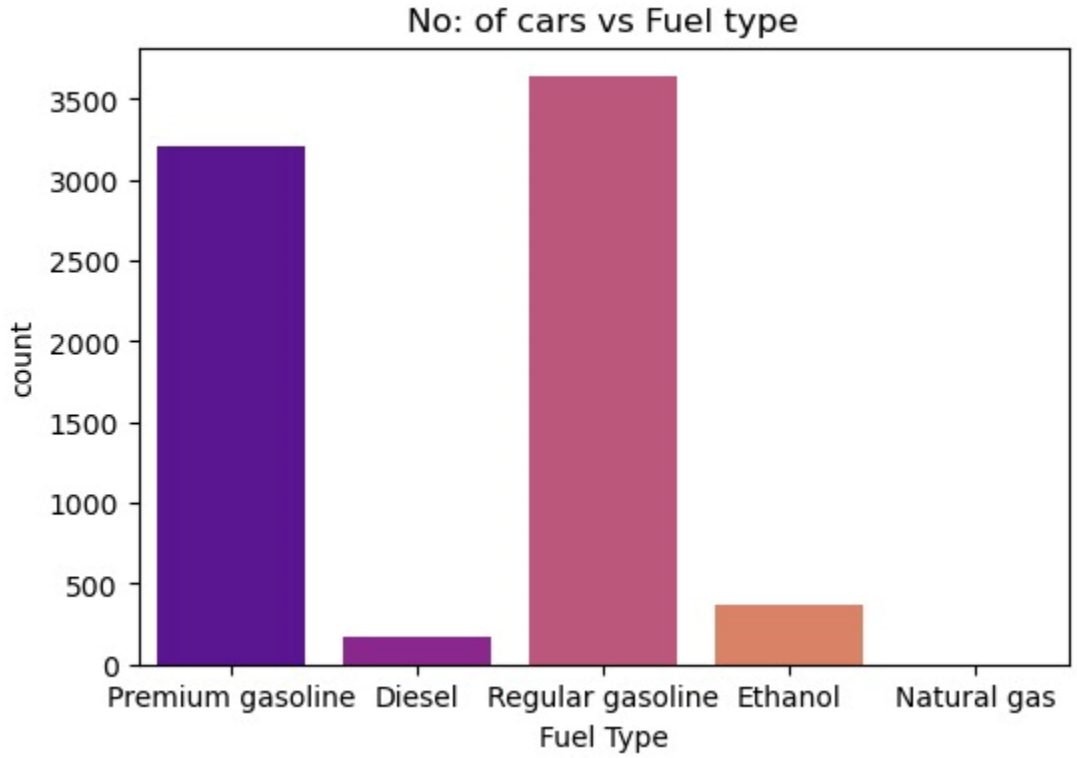


Engine size and CO2 Emission were related to each other. From this bar graph it is clear that as the engine size increases the emission increases. We can see a gradual increase in the emission also draw a linear line for depicting the increase in engine size.

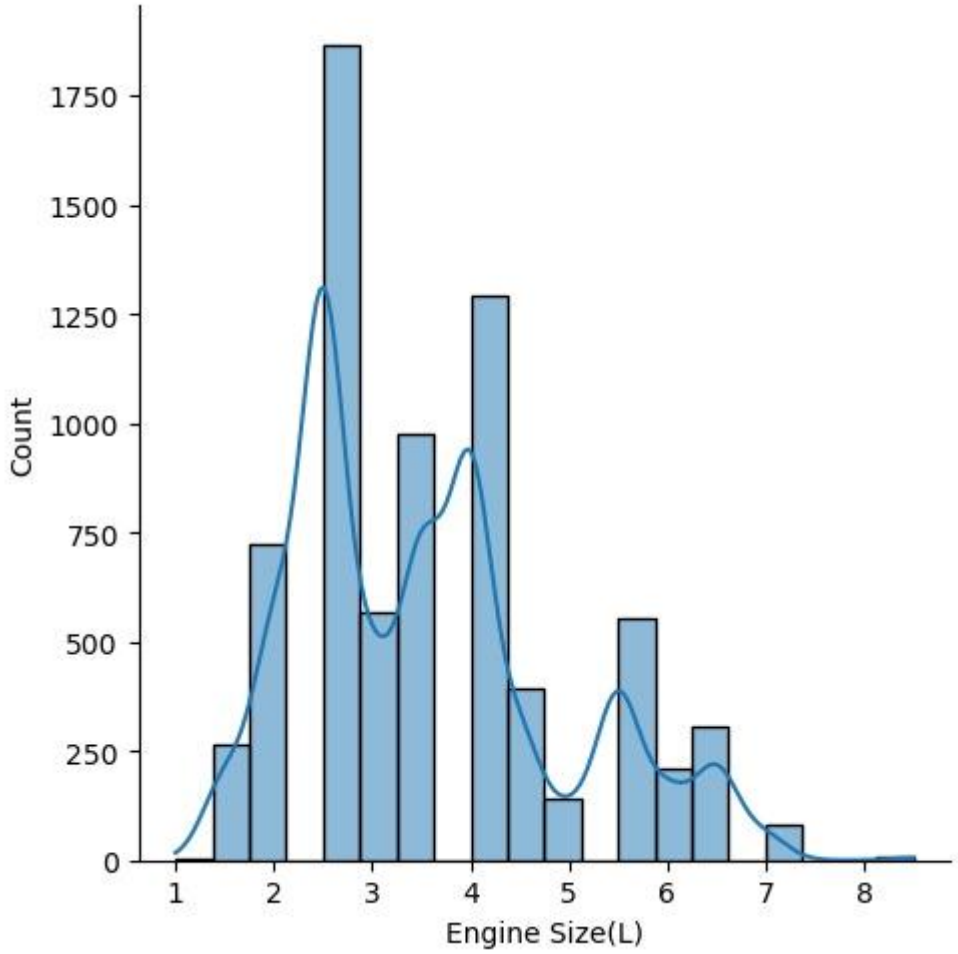


We have fuel combined fuel consumption (L/100km) plotted against CO_2 emissions (g/km) plotted above with the hues adjusted for fuel type. As you can see there are very distinct slopes are evident and appear to be related to the fuel type of each of these vehicles. The slopes of each trend provide some insight regarding the relationships between the increase in fuel consumption and CO_2 emissions. A steeper slope indicating a higher increase in CO_2 emissions for an increase in fuel consumption.

As seen in the figure below, the ethanol-based fuel has the lowest slope of the various vehicle fuels and indicates that it produces less CO_2 emissions for a similar fuel consumption. On the other end, diesel has the steepest slope and therefore has the most CO_2 emissions for a given fuel consumption. A note regarding the ethanol-based fuel is that the low end of the fuel consumption is at 10.1L/100km, whereas the low end of fuel consumption for the gasoline and diesel are 4.1 L/100km and 6.2 L/100km, respectively. Corresponding to the higher fuel consumption, the ethanol fuel yields a minimum of 162 g/km CO_2 emissions; however, gasoline and diesel yield CO_2 emissions as low as 96 g/km and 169 g/km, respectively.



The most common fuel used by cars in a country is Regular gasoline. This is because it is the most affordable fuel option. Premium gasoline is the second most popular fuel type, followed by Diesel. These fuels are more expensive than regular gasoline, but they offer better performance or fuel economy for some cars. Ethanol and natural gas are less popular fuel types, but they are becoming more common as they are seen as more sustainable alternatives to gasoline and diesel.



From the above distribution plot,

Smaller engine sizes are more fuel-efficient, which is important to many car buyers. Smaller engine sizes are also less expensive, which is another factor that many car buyers consider. Larger engine sizes offer more power and performance, which is appealing to some drivers. However, larger engine sizes can also be less fuel-efficient and more expensive.

REGRESSION MODELS

LINEAR REGRESSION

Linear regression analysis serves the purpose of predicting one variable's value based on another variable. The variable under consideration for prediction is termed the "dependent variable," while the variable used to make these predictions is known as the "independent variable." This analytical approach involves determining the coefficients of a linear equation, which may encompass multiple independent variables, to optimize the prediction of the dependent variable's value. Linear regression essentially seeks to find a straight line or surface that minimizes the disparities between predicted and actual values. Additionally, there are readily available calculators specifically designed for simple linear regression, employing the "least squares" method to ascertain the most accurate fit for a given dataset of paired data points

LASSO REGRESSION

LASSO regression, an acronym for Least Absolute Shrinkage and Selection Operator, is a widely adopted technique in statistical modelling and machine learning for understanding variable relationships and making predictions. It seeks to strike a delicate equilibrium between model simplicity and predictive accuracy. LASSO accomplishes this by introducing a penalty term into the conventional linear regression model, promoting sparse solutions in which certain coefficients are compelled to reach precisely zero. This distinctive attribute endows LASSO with exceptional utility in feature selection tasks, enabling it to automatically detect and eliminate irrelevant or redundant variables

RIDGE REGRESSION

Ridge regression is a specialized method employed for the analysis of multiple regression data that exhibits multicollinearity. While it serves as a fundamental regularization technique, it may not be as widely adopted due to its perceived complexity. Nevertheless, delving into the principles of ridge regression in the context of R becomes more manageable with a foundational grasp of multiple regression concepts. In ridge regression, the core principles of regression remain consistent, but the approach to determining model coefficients diverges. The primary concept behind ridge regression revolves around fitting a novel line that deviates from the conventional fitting approach.

CO2 EMISSION PREDICTOR

Later divided the dataset in the ratio of 4:1 for splitting them into testing and training dataset. Trained various machine learning models like Linear Regression, Lasso Regression, Ridge Regression using scikit-learn library. Preprocessing steps like Label-Encoding and Scaling were done on the dataset. Relevant features taken for building the model.

	Model	Accuracy Score
1	Ridge	0.895516
0	Linear	0.894979
2	Lasso	0.881676

While comparing the accuracy of these models, it is clear that Ridge Regression model have the highest accuracy. Since due to its higher performance, it is selected for building the model. A CO2 emission prediction is built by taking input values like engine size, transmission type, fuel consumption, make, model etc. The model predicts the output with an accuracy of almost 90 percent. Created a User Interface using streamlit.

GRAPHICAL USER INTERFACE

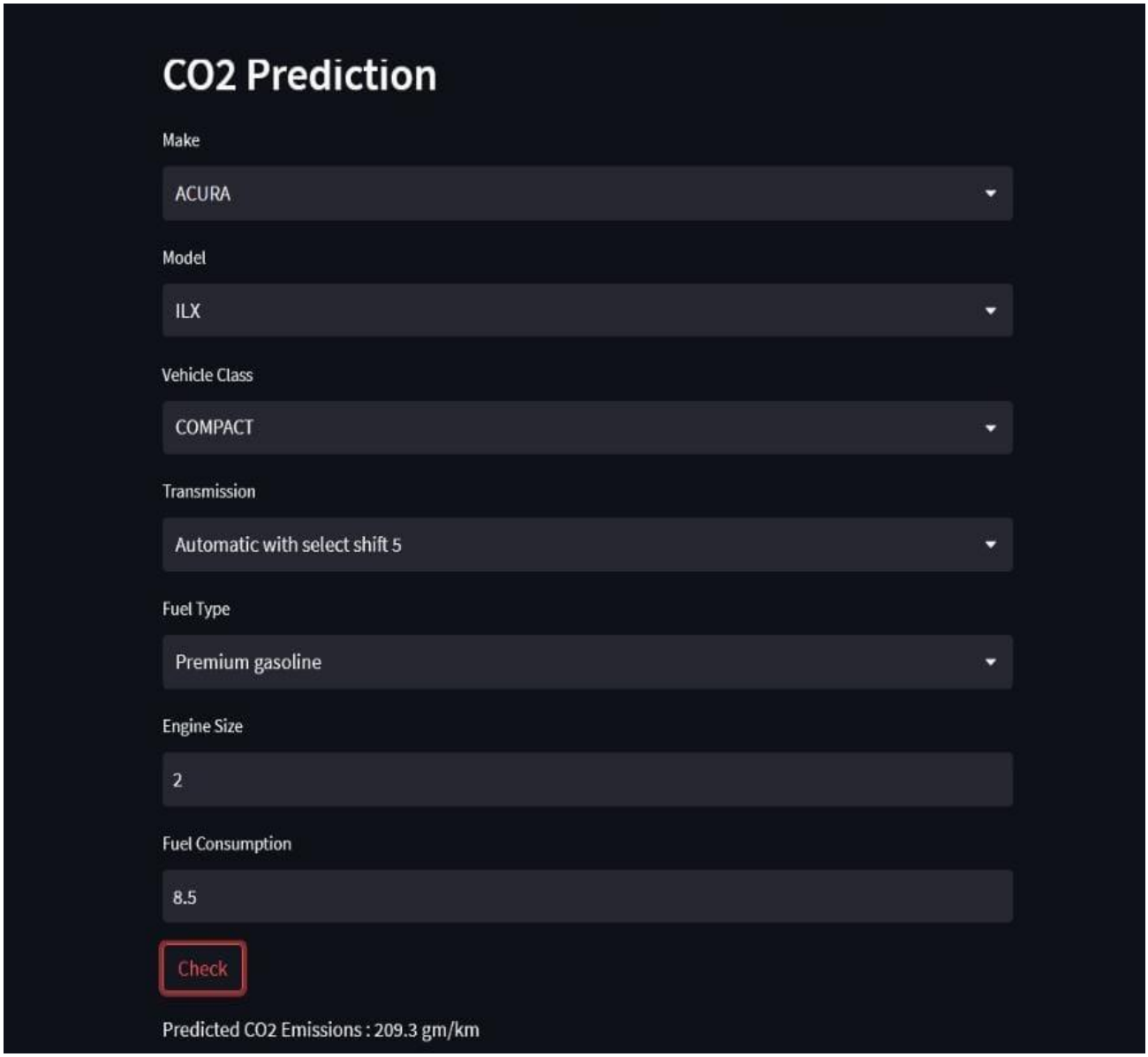


Figure 1: Graphical User Interface for website

CONCLUSION

In a country facing vast problems due to air pollution, this project delved into intricate nexus between CO2 emission from vehicles and the factors that affecting them. Through some meticulous analysis of data and development of a predictive model this work has illuminated the impact of vehicular characteristics in the amount of CO2 emitted.

By the analysis, we found that the factors like engine size, transmission type, Brand etc affect the carbon dioxide emission very drastically. Engine size and CO2 emission have a high positive correlation as the engine size increase the emission also increases. Also, there is a valid point that the vehicles from the brand BUGATTI causes the large CO2 emission. The ethanol-based fuel produces less CO_2 emissions for a similar fuel consumption and diesel has the most CO_2 emissions for a given fuel consumption. Later built a model for predicting the carbon dioxide emission when the other factors are given. The model is trained in lasso regression which acquires an accuracy of almost 90 percent. The outcome of this project profounds implication for urban planning and empowers policy makers to devise targeted interventions that prioritize fuel efficiency and reduced CO2 emissions.

While this project has provided valuable insights, there are avenues for further exploration. Fine-tuning the dataset and expanding the dataset into more geographical area, large spectrum of vehicles and different driving conditions would emphasize to build a better model giving valuable new hidden insights and can be used effectively for implementing a system which have lesser carbon dioxide emission.

REFERENCES

I. https://www.outlookindia.com/national/delhi-imposes-ban-on-non-bs-vi-diesel-vehicles-under-grap-to-tackle-air-pollution-all-you-need-to-know-news-234939

II. https://climate.nasa.gov/vital-signs/carbon-dioxide/#:~:text=Carbon%20dioxide%20in%20the%20atmosphere,in%20less%20than%20200%20years

III. https://www.ibm.com/topics/linear-regression

IV. https://www.engati.com/glossary/ridge-regression

V. https://www.mygreatlearning.com/blog/understanding-of-lasso-regression/

VI. "Machine Learning for CO2 Level Prediction" by Smith et al., published in the journal Energy in 2018.

VII. "Recurrent Neural Networks for CO2 Emission Forecasting" by Chen and Wang, published in the journal Environmental Science & Technology in 2019.

VIII. "A Hybrid Approach for Urban CO2 Concentration Prediction" by Kim et al., published in the journal Sensors in 2020.

IX. "Support Vector Machines for CO2 Emission Prediction from Power Plants" by Li and Liu, published in the journal Energy in 2016.