# 1. INTRODUCTION

## 1.1 Overview:

With the development of the world economy, the situation of carbon dioxide emissions control is becoming increasingly serious. CO2 emissions are the primary driver of global climate change. It is widely recognized to avoid the worst impacts of climate change. A typical vehicle emits about 4.6 metric tons of carbon dioxide per year.

This causes global warming and effects the environment. The number can vary based on a vehicle's fuel, fuel economy, and the number of miles driven per year. When gasoline in the vehicle burns, the carbon and hydrogen separate.

The hydrogen combines with oxygen to form water ($H_2O$), and carbon combines with oxygen to form carbon dioxide ($CO_2$). This is how CO2 is emitted from vehicles.

This project will be predicting the amount of CO2 released by the vehicles by exploring the inputs of the vehicle such as engine-size, fuel-consumption etc.

## 1.2 Purpose:

The goal of the project is to develop an emission predictor which predicts the amount of the CO2 emitted by a vehicle with certain characteristics. The model will be trained and tested using Machine Learning Techniques.

The predictor is built by using Flask applications. HTML and CSS are used for creating and designing the webpage.

# 2. LITERATURE SURVEY

## 2.1 Existing problem:

CO2 emissions are highly renewable, qualitative and quantitative. Exposure to high $CO_2$ levels in vehicles results in unpleasant feeling, fatigue, drowsiness or lethargy among the drivers and passengers. CO2 emitted from vehicles can be reduced little by air filtration.

The CO2 which is emitted from vehicles causes climate change by trapping heat, and also leads to respiratory diseases in the human beings. Extreme weather, food supply disruptions, and increased wildfires are few effects of climate change caused by CO2 emissions.

Carbon emissions include a number of different chemicals and particulates that are produced when fuel is burned in an engine. Some of the major substances found in a car's exhaust include carbon dioxide, ozone, and carbon monoxide. Other chemicals often found in exhaust gasses include benzene and nitrogen oxides.

Many of these chemicals serve an important purpose in different parts of the atmosphere, but they can have bad consequences when human beings inhale them directly. Mitigation of Carbon Dioxide emission is the challenge of the future in order to stabilize global warming. The factors affecting CO2 emissions of vehicles include the GDP, population, urbanization rate, transportation development level, transportation energy intensity, energy consumption structure, and industrial structure.

The level of traffic development is a comprehensive indicator, and there exist some differences in how to measure it quantitatively. The classical regression analysis requires the independent variables to be linearly independent.

The carbon emissions are calculated according to the total amount of smoke. The amount of CO2 emission from the transport sector (including cars) accounts for about 20% of total CO2 emissions. Accordingly, from the viewpoint of preventing global warming, reducing that proportion is a key issue.

In regard to CO2 emissions from cars, fuel economy standards are getting tougher all over the world, so improving the fuel economy of cars is strongly desired. It is considered that the fuel economy of engines will be further improved by boosting engine efficiency and by hybridization or electrification of cars.

Improving fuel economy by improving “driving operation” (i.e. the operation in which a car is driven) and by smoothing traffic flows will come into the picture in the near future.

Vehicle emissions have a number of harmful effects on human health. Exposure to car emissions increases the risk of getting certain cancers. Carbon traps heat in the atmosphere, preventing it from escaping from the earth. This has led to a warmer earth over the past century, increasing the odds of severe weather patterns, droughts, and other problems.

This problem can be rectified by using Machine Learning Techniques by predicting the amount of CO2 emitted from a vehicle by giving few inputs or features of the vehicle. By doing this prediction, the amount of CO2 from the vehicles can be known and reduced by doing few measures.

**2.2 Proposed solution:**

Significantly reducing CO2 emissions from cars will not be easy, but the available data can be used to extract the features, know the behavior of cars, and try to reduce the emissions. Machine Learning techniques can be used in this regard. The solution to the above discussed problems is to build a predictor with predicts the amount of CO2 released by a vehicle. The vehicle CO2 emission model is derived based on the theory of vehicle dynamics. When the CO2 emission predictor predicts the amount of CO2 released, then the car might be modified or replaced with parts which cause less amount of CO2 to be released.

In this project, we take a dataset which contains features of the vehicles like:

- Type of car
- Car class
- Engine size
- Cylinder size
- Transmission of the car
- Fuel Type
- Fuel Consumption
- Combine Fuel Consumption
- Fuel Consumption highway

Now using machine learning techniques like Linear Regression, we prepare the model. The following steps are to be performed to the dataset:

- Download the dataset: Machine Learning depends heavily on data, without data, it is impossible for a machine to learn. It is the most crucial aspect that makes algorithm training possible.
- Pre-process the data:

➢ contains handling the null values
➢ handling the categorical values if the dataset contains
➢ normalizing the data wherever required
➢ identifying the independent and dependent variables in the dataset
➢ Splitting the dataset into Train and Test sets. These tests are used while calculating accuracy for the dataset and predicting the output
    - i.e. the CO2 emitted value from the vehicle.

- Analyze the pre-processed data: involves the dropping few columns we don't require for
    - Our prediction, understanding the data type and summary of the features, observing numerical
    - and categorical values and converting into numerical values if any required.
- Train and test the machine with pre-processed data: Data is to be split into dependent and independent variables. The dependent variable is nothing but output in the dataset and the independent variable is all inputs in the dataset. The data can be divided into train and test sets by considering 80% of the data for training and 20% for testing. The split can be done by passing an argument "random state".
- Save the model and its dependencies: Pickle is used for serializing and de- serializing Python object structures called as flattening. Serialization refers to the process of converting an object in memory to a byte stream that can be stored on disk or sent over a network. Later on, this character stream can then be retrieved and de-serialized back to a Python object.
- Build a Web application using flask that integrates with the model built.
- Static: this folder contains the images and the CSS styles for the webpage we designed.
- Templates: this folder contains the .html files we created for building our web page.
- Training: this folder contains the .ipynb file we created and trained and tested the dataset and performed data visualization techniques.
- The main folder contains the .pkl file i.e. the pickle file.

After training and testing the model, the model should be dumped and saved. This file should be placed

in the main folder.

Some of the functions used in Flask are:

- app – our flask application name
- model – it will contain the model which we build
- @app.route() – it is a decorator that can redirect to different functions.

    i. one for routing to the home page (‘/’)
    ii. route to the prediction page (‘/Prediction’)
    iii. route to the same home page itself (‘/home1’)
    iv. routing to the result page (‘/predict’)

- render_template () – it is used for render our html page from the templates folder

  predict () – is taking the values form the prediction page and storing it into a variable and then we are creating a Data Frame along with the values and 9 independent features and finally we are predicting the values using or loaded model which we build and storing the output in a variable and returning it to the result page.
- app.run(debug=False) – for running our app.

The code for the Flask application is provided in the appendix.

Using the web applications, we build three html pages:

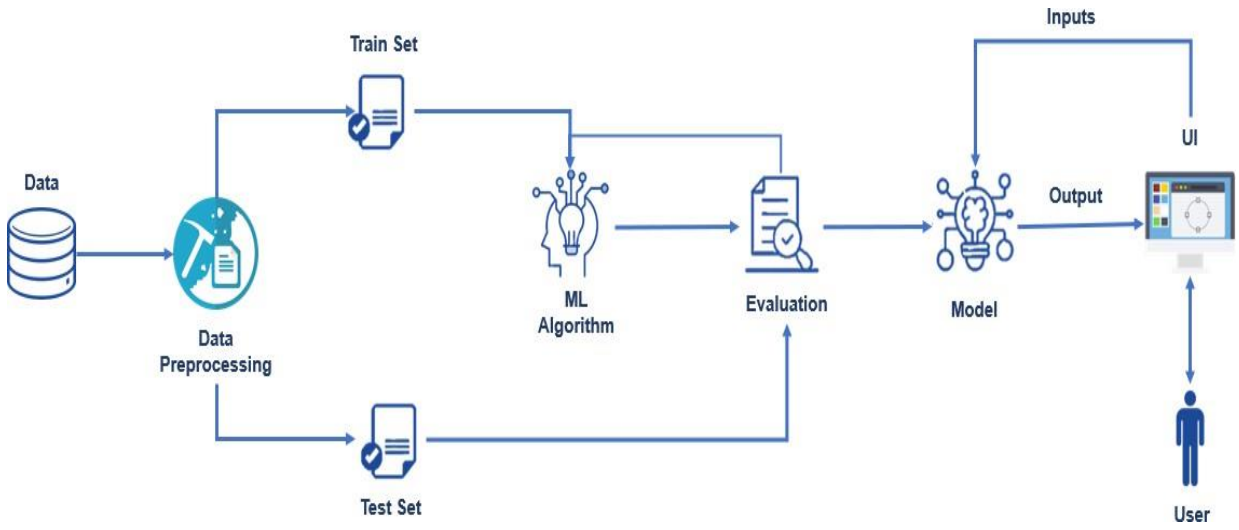
The home page: contains the introduction of the project. This page navigates to the prediction page where we give inputs and predict our model.

1. The prediction page: contains the inputs of the vehicles i.e. the features of the cars. This page navigates to the final page where the CO2 emitted value is shown.
2. The final page: contains the value of the predicted CO2 emitted by a vehicle.

By doing all these applications we can predict the amount of CO2 emitted by a vehicle and decrease the adverse effects caused by CO2 emission. This is the solution for the above discussed problem.

# 3. THEORITICAL ANALYSIS

## 3.1 Block diagram:

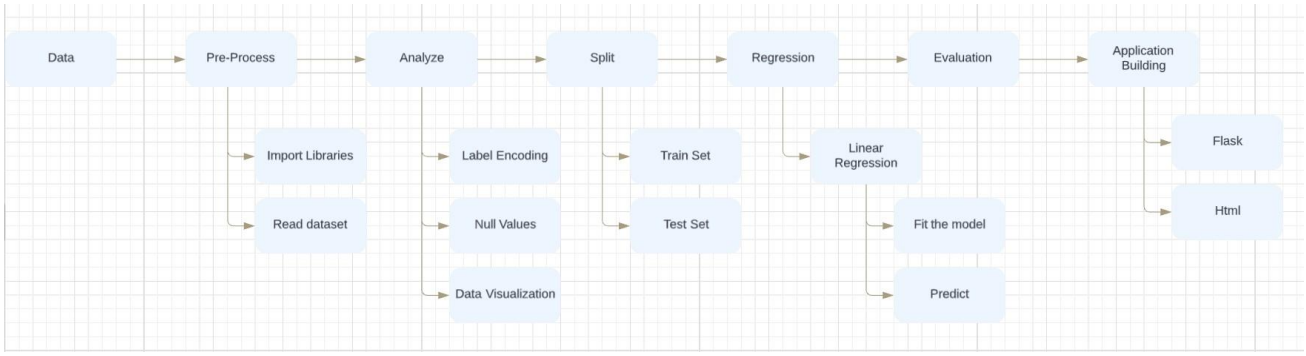



The above diagram is a representation of how the Using ML algorithm, the project works. The data is pre- processed. The data is divided into train and test sets. Using ML algorithm, the data is evaluated. The ML algorithms include regression and classification and, in this project, linear regression is used as the output is in continuous data.

The model is then used for prediction. When the user gives the inputs to the model, the model predicts the amount of CO2 released by the vehicle and gives the output to the user. The model uses Linear Regression and fits the input train and output train into the regression.

For the output to be displayed, web applications are built using Flask. The first page will be the introduction of the project and this page navigates the user to the input pages where the user can give the features of the vehicle and predict the CO2 emission. This page navigates the user by one click on the predict button to the final page where the user will be able to see the predicted value depicted by the model which is obtained by the Machine Learning Algorithms.

### 3.2 Software designing:




The above picture is a design of how the project works.

- The data is loaded and pre-processed.
- While pre-processing the data, the required libraries are imported, the dataset is read.
- The next step is to analyze the dataset. Analyzing the dataset includes:
    - Handling the null values i.e. checking if the dataset contains any null values or not, if contained replacing the null values with a specific value in the dataset.
    - Performing data visualization techniques like plotting graphs between the input or the output to understand the dataset. Few data visualization techniques involve bar-graph, bar-plot etc.
    - Finding the correlation between the independent variables. Correlation is a statistical relationship between two variables and it could be positive or negative. The correlation can be found by using heat map.
        i. Positive: both variables move in the same direction
        ii. Negative: both variables are inversely proportional i.e. when the value of one variable increases, the values of the other one decreases.
    - Label Encoding: this converts the categorical values to numerical values.
- The dataset is to be split into train and test sets. 80% of the data is trained and 20% of the data is tested.

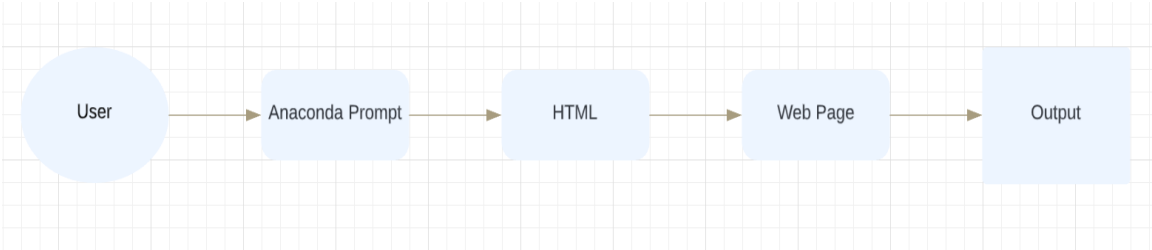
- Next to the dataset the Machine learning algorithms are to be performed. Since the dataset is continuous dataset Regression is used. Using Linear Regression, the model is fitted.

Syntax: variable_name = LinearRegression () variable_name=variable_name.fit (x_train, y_train)

- Once the model is trained, the model is ready to predict. We used "predict method" on the model and pass the parameter: x_test. The predicted value:
  i.e. the output will be stored in "y_pred".

- The next step in the project is to evaluate the model. This can be done by calculating:
  - Mean Absolute Error: measure of errors between paired observations expressing the same phenomenon.
  - Mean Squared Error: average of the squared difference between the target value and the value predicted by the regression model.
  - Root Mean Square Error: is the square root of the averaged squared difference between the target value and the value predicted by the model.
    - After evaluating the model, the model is saved and dumped using the pickle file.

- The last step in the design is to run the application. This is done using the Flask application. More detailed information is given in the following picture and description.




The above diagram is the representation of Flask application.

- The user accesses anaconda prompt and navigates to the folder where the files are located.

- Next is to navigate to the local host where one can view the web page.
- It runs on the localhost:5000

# 4. EXPERIMENT INVESTIGATION

Various test cases were developed for each test scenario to check the correctness of the predicted value. Manual testing was also performed for each of the test case where the inputs are given and executed without any tool. The flask application used in the project also gives the value which is similar to the manual tested value. The given table below is few features from the dataset and CO2 emission value is predicted and calculated by manual.

| Test Case | Input | | | | | | | | | Expected Result | Actual Result | Status |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Type | Class | Engine Size | Cylinder Size | Transmission | Fuel Type | Fuel Consumption City | Fuel Consumption Highway | Combine Fuel | | | |
| 1 | Audi | Full Size | 2 | 4 | AV6 | Diesel | 9.9 | 6.7 | 33 | 173.66 | 173.62 | Pass |
| 2 | BMW | Mini van | 3.7 | 6 | M7 | Natural Gas | 6 | 5.8 | 25 | 216.21 | 216.21 | Pass |
| 3 | Fiat | Mid size | 1.8 | 12 | A6 | Ethanol | 11.2 | 7.5 | 28 | 265.68 | 265.68 | Pass |
| 4 | Honda | SUV small | 5.9 | 8 | AM5 | Regular Gasoline | 13.4 | 12.6 | 27 | 299.54 | 299.55 | Pass |
| 5 | Jeep | VAN cargo | 4.7 | 4 | AS8 | Natural Gas | 17.4 | 11.3 | 29 | 260.28 | 260.23 | Pass |
| 6 | Kia | Mini van | 5.9 | 10 | M7 | Diesel | 9.9 | 7.4 | 18 | 283.91 | 283.91 | Pass |
| 7 | Nissan | Full size | 2 | 8 | AS10 | Ethanol | 11.5 | 8.1 | 19 | 266.411 | 266.4 | Pass |
| 8 | Scion | SUV small | 2.4 | 6 | AV7 | Regular Gasoline | 12.8 | 7.5 | 30 | 249.44 | 249.44 | Pass |

- The table contains input of the features and the expected and actual value of the CO2 emitted by the project. The last column judges if the actual and expected result are similar and gives a pass/fail status.
- The model predicts the CO2 value based on the inputs given.
- Pass status indicates that the expected and actual result is similar to each other.
- There are many other features of the vehicles but for instance a few are taken and the CO2 is calculated.

# 5. CONCLUSION

Based on the analysis of vehicle features, and by the proximate analysis data to predict fuel characteristic factor, the calculation method of predicting CO2 emission of vehicles is established. Successfully we were able to predict the CO2 Emissions for the dataset with 86% accuracy by calculating the r2_score of the dataset.

Prediction with high accuracy can give information concerning about CO2 emissions. The available data can be used to extract the features, know the behavior of cars, and try to reduce the emissions. The model is able to extract the CO2 emitted value by all the features and also is able to generate and error when the user tries to give a null value as an input.

# 6. REFFERNCES

[1] Prediction of CO2 Emissions Based on Multiple Linear Regression Analysis- Science Direct.

[2] Prediction of vehicle CO2 emission and its application to eco-routing navigation – Science Direct.

[3] Predicting CO2 Emissions Growth with Decision Tree Modeling | by Mark Mann | Towards Data Science.

[4] Machine Learning based Prediction System for Detecting Air Pollution (ijert.org).

[5] Machine learning approaches for the prediction of materials properties: APL Materials: Vol 8, No 8 (scitation.org).

[6] Prediction of CO2 Emissions Based on Multiple Linear Regression Analysis (sciencedirectassets.com).