**Clustered Regression Model for Predicting CO2 Emissions from Vehicles**



SUBMITTED TO

##### JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA

In the partial fulfillment for the award of the degree of

##### BACHELOR OF TECHNOLOGY IN COMPUTER SCIENCE AND ENGINEERING

Submitted by

### PUTTI SAI NANDAN BABU 20NG1A05B4 VEMULA SRINIVASARAO 21NG5A0514

**NAMBURI SASIDHAR 20NG1A05A5**

Under the Esteemed Guidance of

##### Dr.K P N V SATYA SREE

Professor

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**



**AUTONOMOUS**

(Approved by AICTE and JNTUK, Kakinada)

(ON NH 16, TELAPROLU, NEAR GANNAVARAM - 521109)

**2020-2024**



**AUTONOMOUS**

##### DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

(Affiliated of to JNTU Kakinada, Approved by A.I.C.T.E, New Delhi) **TELAPROLU, UNGUTURU MANDAL, KRISHNA DISTRICT-521109 2020-2024**

**CERTIFICATE**

This is to certify that this project entitled **“ Clustered Regression Model For Predicting CO2 Emission From Vehicles ”** is the bonafide work of **P.Sai Nandan Babu (20NG1A05B4), V.Srinivasarao (21NG5A0514),N.Sasidhar (20NG1A05A5)** who carried out the work under my supervision, and submitted in partial fulfilment of the requirements for the award of the degree in Bachelor of Technology in Computer Science & Engineering, during the academic year 2020-24.

##### Project Guide Head of the Department

**Dr.K P N V SATYA SREE Dr. K P N V SATYA SREE**

Professor Professor

Signature of External Examiner <https://usharama.edu.in/home>Tel: 0866 252755, +91 9949712255

**DECLARATION**

We hereby declare that the project entitled **“Clustered Regression Model For Predicting CO2 Emission From Vehicles”**is the work done by us during the academic year 2020-2024 and is submitted in partial fulfillment of the requirements for the award of degree of **Bachelor of technology** in **COMPUTER SCIENCE AND ENGINEERING** from **JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY, KAKINADA**.

#### BY

##### P.SAI NANDAN BABU (20NG1A05B4) V.SRINIVASARAO (21NG5A0514)

**N.SASIDHAR (20NG1A05A5)**

**ACKNOWLEDGEMENT**

##### We express our sincere thanks where it is due

We are pleased to acknowledge our sincere thanks to our Honorable Chairman **SRI. S.RAMABRAHMAM GARU** for the support and advice which is given and for providing sufficient resources.

We are extremely thankful to **Dr. K RAJASEKHARA RAO**, Director of USHA RAMA COLLEGE OF ENGINEERING AND TECHNOLOGY, TELAPROLU for giving a golden

opportunity to our education and project work.

We wish to avail this opportunity to express our thanks to **Dr. G V K S V PRASAD**, Principal, URCE for his continuous support and giving valuable suggestions during the entire period of the project work.

We take this opportunity to express our gratitude to **Dr. K P N V SATYA SREE**, Head of the Department Professor in **Computer Science and Engineering** for her valuable support and motivation at each and every point in successful completion of the project.

We Express our sincere thanks to our guide **Dr.S M ROY CHOUDRI,**Professor in the Department of Computer Science and Engineering for motivating us to make our project successful. We grateful for her guidance and suggestions.

We also place our floral gratitude to all other teaching staff and lab technicians for their constant support and advice throughout the project.

**BY**

##### P.SAI NANDAN BABU (20NG1A05B4)

**V.SRINIVASARAO (21NG5A0514)**

##### N.SASIDHAR (20NG1A05A5)

Clustered Regression Model for Predicting CO2 Emissions

from Vehicles

# ABSTRACT

**ABSTRACT**

The global imperative to combat climate change has heightened the urgency to reduce greenhouse gas emissions, particularly those originating from the transportation sector. In response to this pressing concern, this research project introduces a novel approach to predict carbon dioxide (CO2) emissions from vehiclesusing a combination of K-means clustering and Linear Regression. The research begins with the premise that vehicles exhibit diverse emission characteristics due to factors such as vehicle class, engine size, number of cylinders and fuel consumption in different roads. To capture this complexity, a clustering technique is employed to group similar vehicles together. The number of clusters is treated as a tunable hyperparameter, allowing the model to adapt to the dataset's inherent structure. Within each cluster, a Linear Regression model is trained to estimate CO2 emissions based on various vehicle attributes. This granular approach enables the model to discern unique emission patterns within each cluster, offering insights into the specific factors influencing emissions for different subsets of vehicles. The K-means clustering and Linear Regression models are integrated seamlessly, resulting in a robust predictive tool. When presented with new vehicle data, the model assigns each vehicle to the appropriate cluster and utilizes the corresponding regression model for prediction. The predictions are then aggregated and sorted to align with the original data order. The significance of this research lies in its ability to provide a nuanced understanding of CO2 emissions from vehicles.By consideringthe heterogeneity of vehicles and their emissions, the model offers a more accurate representation of real-world emissions scenarios. Moreover, it provides a platform for evaluating the impact of policy interventions on different vehicle clusters, thus aiding in the formulation of targeted and effective emissions reduction strategies. In conclusion, this project presents a novel approach to address the critical issue of vehicle

emissions by combining clustering and regression techniques. By doing so, it contributes to the broader efforts towards sustainable transportation and environmental conservation. This research offers a promising avenue for policymakers and stakeholders to make informed decisions aimed at mitigating the environmental impact of the transportation sector, ultimately working towards a greener and more sustainable future.

**Keywords**: Carbon Dioxide(CO2), CO2 Emission, Engine size, number of cylinders, Vehicles, Clustering, Centroids, Euclidean distance,Linear Regression, mean squared error (MSE)

# TABLE OF CONTENTS

|  |  |
| --- | --- |
| **TOPIC** | **PAGE NO** |
| Chapter 1 |  |
| 1. [INTRODUCTION](#_bookmark0) | [01](#_bookmark0) |
| 1.1. Literature Survey | 03 |
| 1.1.1. [Machine Learning](#_bookmark1) | 04 |
| 1.1.2. [Features of Machine Learning](#_bookmark2) | 08 |
| 1.1.3. [Existing System](#_bookmark3) | 10 |
| 1.1.4. [Proposed System](#_bookmark4) | 12 |

## Chapter 2

1. AIM &SCOPE
   1. [Feasibility study 14](#_bookmark5)
      1. [Technical Feasibility 15](#_TOC_250002)
      2. Operational Feasibility 15
      3. [Economic Feasibility 1](#_bookmark6)5
      4. [Market Feasibility 15](#_TOC_250001)
      5. [Ethical Feasibility 1](#_bookmark6)5
      6. Schedule Feasibility 15
   2. System Requirements Specification 16
      1. [Functional Requirements 1](#_bookmark7)6
      2. [Non-Functional Requirements 1](#_bookmark8)8
      3. Software Requirements 20
      4. Hardware Requirements 20

Chapter 3

1. [CONCEPTS & METHODS 22](#_bookmark9)
   1. [Problem Definition 2](#_bookmark10)2
   2. [Proposed Description 23](#_TOC_250000)
      1. Algorithms Proposed 25
      2. Modules 27
   3. System analysis methods 28

|  |  |
| --- | --- |
| 3.3.1 Use Case Analysis | 29 |
| 3.3.2 Data Flow Diagram | 29 |
| 3.3.3 Entity-Relationship Diagram | 30 |
| 3.4 System Modeling | 30 |
| 3.4.1 Emission for different modes of Transport | 31 |
| 3.4.2 Fuel and energy generation | 31 |
| Chapter 4 |  |
| 4. [IMPLEMENTATION](#_bookmark11) | [3](#_bookmark11)3 |
| 4.1 Tools used | 34 |
| 4.2 [Pseudo code](#_bookmark12) | [3](#_bookmark12)6 |
| 4.3 Regression Models | 37 |
| 4.4 Comparative Analysis | 38 |
| Chapter 5 |  |
| 5. SCREEN SHOTS | 41 |
| Chapter 6 |  |
| 6. TESTING | 45 |
| Chapter 7 |  |
| 7. [SUMMARY & CONCLUSION](#_bookmark13) | 47 |
| Chapter 8 |  |
| 8. FUTURE ENHANCEMENTS | 50 |
| Chapter 9 |  |
| 9. BIBILOGRAPHY | 52 |
| Chapter 10 |  |
| 10 .PUBLISHING PAPER | 53 |



# CHAPTER - 1 INTRODUCTION



1. **INTRODUCTION**

The "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project is a critical endeavor that addresses one of the most pressing challenges of our time—reducing carbon dioxide (CO2) emissions from the automotive sector. As the world grapples with the consequences of climate change and environmental degradation, the automotive industry, a significant contributor to greenhouse gas emissions, is under increasing pressure to adopt sustainable practices and comply with stringent regulations aimed at curbing CO2 emissions.

This project is a response to this call for action. It seeks to develop an innovative approach to predicting CO2 emissions from vehicles, harnessing the power of data science, machine learning, and clustering techniques. By doing so, it not only addresses the need for accurate predictive models but also contributes to the broader goal of transitioning towards cleaner and more environmentally responsible transportation.The automotive industry has long been a cornerstone of modern society, providing mobility and convenience to billions of people. However, this convenience has come at a cost—rising CO2 emissions that contribute to global warming and climate change. In response, governments and environmental agencies worldwide have enacted strict emissions standards and regulations. These regulations not only demand compliance from vehicle manufacturers but also compel consumers to make more sustainable choices when selecting vehicles.

Predicting CO2 emissions accurately is a multifaceted challenge. A vehicle's emissions depend on a multitude of factors, including its make, model, engine size, fuel type, and even driver behavior. Traditionally, predictive models have taken a broad approach, offering generalized estimates for all vehicles. However, this approach often lacks precision and may not meet the requirements of modern environmental regulations.The project aims to create a state-of-the-art regression model capable of accurately estimating CO2 emissions from individual vehicles. This model will consider a wide range of vehicle attributes and factors that influence emissions.A novel aspect of this project involves the use of clustering techniques to group vehicles based on their similarities and characteristics. These clusters will serve as a foundation for tailoring emissions predictions to specific groups of vehicles, recognizing that one-size-fits-all models may not be suitable in the context of emissions reduction.

Beyond prediction, this project aims to provide actionable insights into the determinants of CO2 emissions from vehicles. By understanding the influence of various features and behaviors, we can offer recommendations to multiple stakeholders, including regulatory bodies, vehicle manufacturers, and consumers, thus contributing to the broader effort of reducing emissions.



### LITERATURE SURVEY

1. Predicting CO2 Emissions from Traffic Vehicles for Sustainable and Smart Environment Using a Deep Learning Model. Author:- Abdullah H. Al-Nefaie and Theyazn H. H. Aldhyani Description:- The research focuses on predicting CO2 emissions from vehicles, with a specific emphasis on the context of Saudi Arabia, a major oil-producing nation with a rapidly growing number of automobiles. Machine learning and deep learning techniques, including LSTM and BiLSTM models, are employed to achieve accurate predictions of CO2 emissions.The study introduces rough k-means clustering as a preprocessing method to enhance model performance.
2. Forecasting Carbon Dioxide Emissions of Light-Duty Vehicles with Different Machine Learning Algorithms. Author:- Yuvaraj Natarajan,Gitanjali Wadhwa,K. R. Sri Preethaa Description:- In this research paper, the critical issue of rising carbon emissions from the transportation sector, primarily passenger cars, is addressed. The transportation industry's substantial contribution to greenhouse gas emissions and air pollution is a growing concern, necessitating innovative approaches for mitigating its environmental and socio-economic impacts. The study delves into the analysis of fuel consumption and CO2 emissions from a comprehensive dataset of light-duty vehicles, spanning the years 2017 to 2021. Through meticulous data analysis and machine learning techniques, the research aims to create predictive models that estimate carbon dioxide emissions and fuel consumption based on vehicle characteristics. This literature review explores the context and significance of this research within the broader field of environmental sustainability and data-driven solutions to transportation-related challenges..
3. Carbon Emission Prediction on the World Bank Dataset for Canada. Author:- Aman Desai, Shyamal Gandhi, Sachin Gupta, Manan Shah Description:- In this research paper, the pressing issue of increasing carbon dioxide (CO2) emissions from the burning of carbon-containing fuels is addressed. These emissions contribute significantly to the greenhouse effect, global warming, and climate change, which have far-reaching consequences such as unpredictable weather patterns, rising sea levels, and disruptions in food chains. The economic impact of carbon emissions is also a concern, as they can hinder global economic growth and lead to inflation and economic crises. Many countries are taking initiatives to control their carbon footprints, and carbon emission prediction has emerged as a valuable approach to inform policy decisions. By predicting future carbon emission patterns, countries and industries can take proactive steps to reduce emissions. Machine learning techniques have gained prominence in this field, with various models being used to forecast carbon emissions. This research project focuses on comparing different machine learning models to predict carbon emissions for Canada from 2019 to 2030, contributing to Canada's commitment to reducing carbon emissions under the Paris Agreement



##### 1.1.1 MACHINE LEARNING

A Machine Learning defined as “A computer program is said to learn from experience and from some tasks and some performance on, as measured by, improves with experience”. Machine Learning is combination of correlations and relationships, most machine learning algorithms in existence are concerned with finding and/or exploiting relationship between datasets. Once Machine Learning Algorithms can pinpoint on certain correlations, the model can either use these relationships to predict future observations or generalize the data to reveal interesting patterns.In Machine Learning there are various types of algorithms such as Regression, Linear Regression, Logistic Regression, Naive Bayes Classifier, Bayes theorem, KNN (K-Nearest Neighbour Classifier), Decision Tress, Entropy, ID3, SVM (Support Vector Machines), K-means Algorithm, Random Forest and etc.,

The name machine learning was coined in 1959 by Arthur Samuel. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data Machine learning is closely related to (and often overlaps with) computational statistics, which also focuses on prediction- making through the use of computers. It has strong ties to mathematical optimization, which delivers methods, theory and application domains to the field. Machine learning is sometimes conflated with data mining, where the latter subfield focuses more on exploratory data analysis and is known as unsupervised learning.

With in the field of data analytics, machine learning is a method used to devise complex models and algorithms that lendthemselves to prediction; in commercial use, this is known as predictive analytics. These analytical models allow researchers, data scientists, engineers, and analysts to "produce reliable, repeatable decisions and results" and uncover"hidden in sights" through learning from historical relationships and trends in the data

Machine learning implementations are classified into three major categories, depending on the nature of the learning “signal” or “response” available to a learning system which are as follows:

##### Supervised learning:

When an algorithm learns from example data and associated target responses that can consist of numeric values or string labels, such as classes or tags, in order tolater predict the correct response when posed with new examples comes under the category ofSupervised learning. This approach is indeed similar to human learning under the supervisionof a teacher. The teacher provides good examples for the student to memorize, and the studentthen derives general rules from these specific examples.



##### Unsupervised learning:

When an algorithm learns from plain examples without any associated response, leaving to the algorithm to determine the data patterns on its own. This type of algorithm tends to restructurethe data into something else, such as new features that may represent a class or a new series of un- correlated values. They are quite useful in providinghumans with insights into the meaning of data and new useful inputs to supervised machine learning algorithms. As a kind of learning, it resembles the methods humans use to figure out that certain objects or events are from the same class, such as by observing the degree of similarity between objects. Some recommendation systems that you find on the web in the formof marketing automation are based on this type of learning.

##### Reinforcement learning:

When you present the algorithm with examples that lack labels, as in unsupervised learning. However, you can accompany an example with positive or negativefeedback according to the solution the algorithm proposes comes under the category of Reinforcement learning, which is connected to applications for which the algorithm must make decisions (so the product is prescriptive, not just descriptive, as in unsupervised learning), andthe decisions bear consequences. In the human world, it is just like learning by trial and error.Errors help you learn because they have a penalty added (cost, loss of time, regret, pain, and so on), teaching you that a certain course of action is less likely to succeed than others.

In this case, an application presents the algorithm with examples of specific situations, such as having the gamer stuck in a maze while avoiding an enemy. The application lets the algorithmknow the outcome of actions it takes, and learning occurs while trying to avoid what it discovers to be dangerous and to pursue survival. You can have a look at how the company Google Deep Mind has created a reinforcement learning program that plays old Atari’s video 3 games. When watching the video, notice how the program is initially clumsy and unskilled but steadily improves with training until it becomes a champion.

##### Semi-supervised learning:

Where an incomplete training signal is given: a training set with some (often many) of the target outputs missing. There is a special case of this principle known as Transduction where the entire set of problem instances is known at learning time, except that part of the targets are missing. Supervised Learning the majority of practical machine learning uses supervised learning. Supervised learning is where you have input variables (x) and an output variable (Y)and you use an algorithm to learn the mapping function from the input to the output.



Y = f(X)

The goal is to approximate the mapping function so well that when you have new input datathat you can predict the output variables (Y) for that data. It is called supervised learning because the process of an algorithm learning from the training dataset can be thought of as a teacher supervising the learning process. We know the correct answers, the algorithm iteratively makes predictions on the training data and is corrected by the teacher. Learning stopswhen the algorithm achieves an acceptable level of performance.

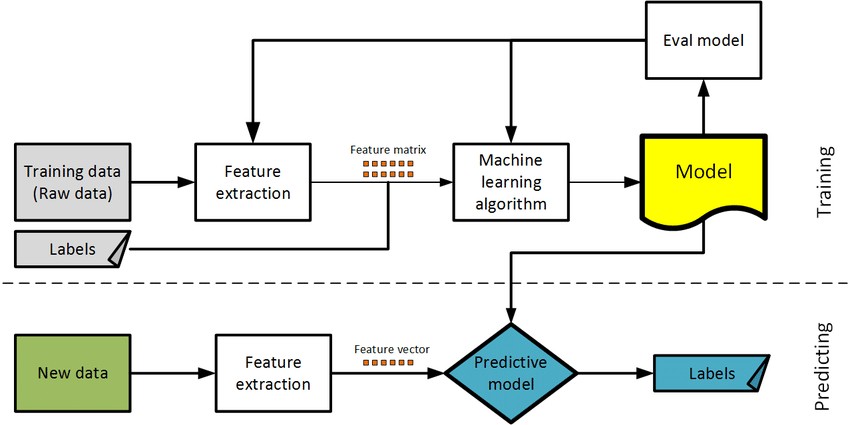
##### Types of Supervised Learning

**Classification:**

It is a Supervised Learning task where output is having defined labels (discretevalue). For example, in above Figure A, Output – Purchased has defined labels i.e., 0 or 1; 1 means the customer will purchase and 0 means that customer won’t purchase. The goal here is to predict discrete values belonging to a particular class and evaluate on the basis of accuracy. It can be either binary or multi class classification. In binary classification, model predicts either 0 or 1; yes or no but in case of multi class classification, model predicts more than one class. Example: Gmail classifies mails in more than one classes like social, promotions,updates, forum.

##### Regression:

It is a Supervised Learning task where output is having continuous value. Examplein above Figure B, Output – Wind Speed is not having any discrete value but is continuous in the particular range. The goal here is to predict a value as much closer to actual output value asour model can and then evaluation is done by calculating error value. The smaller the errorthe greater the accuracy of our regression model.



##### Fig: 1.1.1. FLOW CHART OF SUPERVISED LEARNING ALGORITHM



**Classification:**

Data mining is the process of extracting knowledge-able information from huge amounts of data. It is an integration of multiple disciplines such as statistics, machine learning, neural networks and pattern recognition. Data mining extracts biomedical and health care knowledge for clinical decision making and generates scientific hypotheses from large medical data.

Association rule mining and classification are two major techniques of data mining. Association rule mining is an unsupervised learning method for discovering interesting patternsand their association inlarge data bases.

Classification is a supervised learning method used to find class labels for unknown samples. Classification is the task of assigning an object's tone of special predefined categories. It is pervasive problem that encompasses many applications.

Classification is designed as the task of learning a target function F that maps each attribute setA to one of the predefined class labels C. The target function is also known as classification model. A classification model is useful for mainly two purposes.

descriptive modelling. Predictive modelling.

Classification is the process of recognizing, understanding, and grouping ideas and objects into pre- set categories or “sub-populations.” Using pre-categorized training datasets, machine learning programs use a variety of algorithms to classify future datasets into categories.

Classification algorithms in machine learning use input training data to predict the likelihood that subsequent data will fall into one of the predetermined categories. One of the most common uses of classification is filtering emails into “spam” or “non-spam.”

In short, classification is a form of “pattern recognition,” with classification algorithms applied to the training data to find the same pattern (similar words or sentiments, number sequences, etc.) in future sets of data.

Classification can be performed on structured or unstructured data. Classification is a technique where we categorize data into a given number of classes. The main goal of a classification problem is to identify the category/class to which a new data will fall under.



Few of the terminologies encountered in machine learning – classification: Classifier: An algorithm that maps the input data to a specific category.

**Classification model:** A classification model tries to draw some conclusion from the input values given for training. It will predict the class labels/categories for the new data.

**Feature**: A feature is an individual measurable property of a phenomenon being observed. Binary **Classification:** Classification task with two possible outcomes. E.g., Gender classification (Male / Female).

Multi-class classification: Classification with more than two classes. In multi class classification each sample is assigned to one and only one target label. E.g., An animal can be cat or dog but not both at the same time.

**Multi-label classification**: Classification task where each sample is mapped to a set of target labels (more than one class). E.g., A news article can be about sports, a person, and location at the same time.

##### Applications of Classification Algorithms:

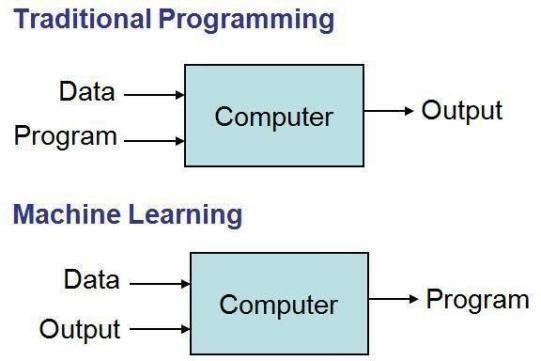
* Email spam classification
* Bank customers loan pay willingness prediction.
* Cancer tumour cells identification.
* Sentiment analysis
* Drug’s classification
* Facial key points detection
* Pedestrians’ detection in an automotive car driving.

##### FEATURES OF MACHINE LEARNING

* + - * It is nothing but automating the Automation.
      * Getting computers to program themselves.
      * Writing Software is bottleneck.
      * Machine leaning models involves machines learning from data without the help ofhumans or any kind of human intervention.
      * Machine Learning is the science of making of making the computers learn and act likehumans by feeding data and information without being explicitly programmed.

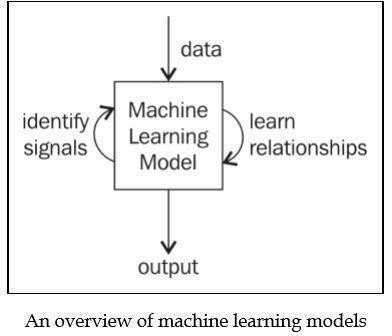


* Machine Learning is totally different from traditionally programming, here data and output is given to the computer and in return it gives us the program which provides solution to the various problems. Below is thefigure



##### Fig: 1.1.1.1. TRADITIONAL PROGRAMMING VS MACHINE LEARNING

* Machine Learning is a combination of Algorithms, Datasets,and Programs.
* There are Many Algorithms in Machine Learning through which we will provide us the exact solution inpredicting the disease of the patients.
* How Does Machine Learning Works?
* Solution to the above question is Machine learning works by taking in data, finding relationships within that data and then giving the output.



##### Fig: 1.1.1.2 MACHINE LEARNING MODEL

There are various applications in which machine learning is implemented such as Web search, computing biology, finance, e-commerce, space exploration, robotics, social networks, debugging and much more.



##### EXISTING SYSTEM

Regression Models: Traditional emissions prediction models typically rely on regression techniques, such as linear regression or multiple regression. These models use historical emissions data and a limited set of vehicle attributes to estimate CO2 emissions. They are often based on the assumption that all vehicles follow the same emissions behavior, resulting in generalized predictions.

Data Sources: The existing system primarily uses emissions test data and vehicle specifications, such as engine size, fuel type, and transmission type. Real-world driving data is often not integrated into these models, and the data sources may not cover a wide variety of vehicle makes and models.

Simplistic Predictions: The existing models tend to provide simplified predictions that don't account for the nuanced variations in emissions behavior among different types of vehicles. As a result, they may not meet the accuracy requirements of modern emissions standards and regulations.

Challenges with the Existing System

Model Accuracy: The traditional regression models often lack the precision needed to accurately predict emissions for a wide range of vehicle types. They may underpredict or overpredict emissions, leading to compliance issues or missed opportunities for reducing emissions.

Environmental Impact: Inaccurate emissions predictions can have a significant environmental impact, as they may not reflect the true emissions levels of vehicles, leading to higher CO2 emissions and air pollution.

Proposed Solution: Clustered Regression Model

In response to the limitations of the existing system, the proposed "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project seeks to introduce a more advanced and adaptable approach. This project will leverage clustering techniques to group vehicles with similar characteristics and emissions behavior and then develop regression models tailored to each cluster. The project aims to address the following:

Increased Accuracy: By accounting for vehicle diversity through clustering, the proposed model is expected to offer more accurate emissions predictions, improving compliance with emissions standards.

Enhanced Adaptability: The proposed system will be designed to adapt to new data sources and changing vehicle technologies, ensuring it remains relevant and effective over time.

Actionable Insights: By understanding the specific factors that influence emissions within each cluster, the project aims to provide actionable insights for reducing emissions on a per-vehicle basis, benefiting consumers, manufacturers, and regulatory bodies.



##### KNN Drawbacks:

**K-Value Selection:** Choosing the appropriate value of k in KNN significantly impacts the algorithm's performance. Selecting an incorrect k-value can lead to overfitting or underfitting, affecting the accuracy of facial expression recognition.

**Storage Requirements:** KNN stores the entire training dataset in memory, which can be a drawback when working with large datasets. This consumes memory and increases the time required for searching neighbors, especially in high-dimensional feature spaces.

**Slow Query Time:** Real-time facial emotion detection applications require fast and efficient processing of image inputs. KNN's slow query time can be a drawback in scenarios where quick response times are essential.



##### PROPOSED SYSTEM

The proposed system, "Clustered Regression Model for Predicting CO2 Emissions from Vehicles," introduces an innovative and adaptive approach to accurately predict CO2 emissions from vehicles. This system addresses the limitations of the existing system by incorporating clustering techniques and leveraging machine learning methods. Below is an overview of the proposed system:

Proposed System: Clustered Regression Model

Data Collection and Preprocessing:Collect a diverse and comprehensive dataset encompassing vehicle features, such as make, model, engine size, fuel type, transmission, and driving data.Clean and preprocess the data, addressing missing values, outliers, and inconsistencies.Standardize and normalize numerical features and encode categorical variables appropriately.

Clustering Analysis:Apply clustering algorithms, such as K-means, to group vehicles into clusters based on shared characteristics that influence CO2 emissions.Determine the optimal number of clusters to ensure meaningful segmentation.Assign vehicles to clusters based on their attributes.

Regression Model Development:Develop multiple regression models, one for each cluster of vehicles. These models will be tailored to the characteristics of the vehicles within each cluster.Utilize machine learning techniques, such as decision trees, random forests, or neural networks, to build accurate predictive models.Incorporate the most influential vehicle attributes, as identified during data analysis.

Model Training and Evaluation:Split the dataset into training and testing sets to train and evaluate the cluster-specific regression models.

Employ appropriate evaluation metrics, such as Root Mean Square Error (RMSE) or R-squared, to assess model performance.Ensure that the models are robust, resistant to overfitting, and capable of providing accurate CO2 emissions predictions.

Insights and Recommendations:

Extract insights from the model outputs and cluster characteristics to understand which vehicle features have the most significant impact on CO2 emissions within each cluster.Formulate actionable recommendations for reducing emissions based on cluster-specific findings. These recommendations can be targeted at various stakeholders, including regulatory bodies, manufacturers, and consumers.



# CHAPTER-2 AIM & SCOPE



1. **AIM & SCOPE:**

The primary aim of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project is to develop an advanced and accurate predictive system that provides precise estimates of carbon dioxide (CO2) emissions from individual vehicles. By employing clustering techniques to group vehicles based on shared characteristics and then creating cluster-specific regression models, this project seeks to significantly enhance the accuracy of CO2 emissions predictions. The overarching aim is to contribute to environmental sustainability and regulatory compliance within the automotive industry.

The aim and scope of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project align with the critical need to develop more accurate and tailored approaches to emissions prediction, thereby fostering environmental sustainability, compliance with emissions standards, and responsible choices in the automotive industry. The project's impact extends beyond academia, offering practical solutions and insights for addressing one of the most pressing challenges of our time.

##### FEASIBILITY STUDY:

A feasibility study is crucial to determine the viability and practicality of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project. It assesses various aspects, including technical, economic, operational, legal, and scheduling considerations. Below is an outline of the feasibility study for this project:

##### Technical Feasibility:

**Data Availability**: Assess the availability of data sources required for the project. This includes vehicle specifications, emissions data, and real-world driving data. Verify that the data is accessible, complete, and accurate.

**Technological Resources:** Evaluate whether the necessary hardware and software resources for data processing, clustering, and regression model development are available or can be obtained within budget constraints.

**Expertise:** Ensure the project team possesses the required technical skills in data science, machine learning, and clustering techniques.

##### Economic Feasibility:

Budget Analysis: Estimate the costs associated with data collection, software and hardware procurement, personnel, and any additional resources required for the project.

**Return on Investment (ROI):** Consider the potential benefits of the project, such as improved emissions predictions and compliance with regulations, and assess whether these benefits outweigh the costs.

**Funding Sources:** Identify potential funding sources, including grants, sponsorships, or internal budget allocations.



##### Operational Feasibility:

**Data Preprocessing:** Evaluate the feasibility of cleaning and preprocessing the data to ensure it is suitable for analysis. Determine the time and effort required for data preparation.

**Model Development:**Assess the feasibility of creating cluster-specific regression models, taking into

account the complexity of the modeling process and the computational resources needed.

**Scalability**: Consider whether the system can scale to accommodate larger datasets and evolving automotive technologies.

##### Market Feasibility:

The market feasibility of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" is highly promising, considering the growing global emphasis on environmental sustainability and the increasing demand for innovative solutions in the automotive sector. With the rising concerns about climate change and the need for reducing carbon footprints, there is a substantial market for tools and technologies that enable precise prediction and management of CO2 emissions from vehicles. The proposed model addresses this demand by offering a sophisticated approach that leverages clustering and regression techniques, providing accurate and cluster-specific predictions tailored to diverse vehicle characteristics. The automotive industry, regulatory bodies, and environmental organizations are potential stakeholders eager to adopt such a model for informed decision-making and emissions reduction strategies. Moreover, the model's adaptability and scalability make it well-suited for integration into various sectors, including fleet management, urban planning, and environmental policy development. As governments worldwide intensify efforts to enforce emissions standards, the market feasibility of the proposed model is reinforced, positioning it as a valuable tool in the global drive towards sustainable and eco-friendly transportation practices.

##### Ethical Feasibility:

**Data Privacy and Compliance:** Ensure that data collection and usage adhere to relevant privacy regulations and ethical standards, especially if sensitive or personal data is involved.

**Intellectual Property:** If applicable, address any legal issues related to intellectual property, data ownership, and licensing agreements for data sources.

##### Scheduling Feasibility:

**Project Timeline:** Create a detailed project timeline with milestones and deadlines for each phase, including data collection, analysis, model development, and reporting.

**Resource Availability:** Ensure that team members and resources are available and allocated according to the project schedule.



**2.2. SYSTEM REQUIREMENT SPECIFICATION**

The purpose of this document is to define the requirements and specifications for the development of a Clustered Regression Model that predicts carbon dioxide (CO2) emissions from vehicles. The system aims to enhance the accuracy of emissions predictions, offering valuable insights for environmental sustainability and regulatory compliance.The system will involve data collection,Preprocessing, clustering analysis, regression model development, model training, and insights generation. It will be adaptable to various data sources and capable of providing cluster-specific emissions predictions.

##### FUNCTIONAL REQUIREMENTS:

Functional requirements are specific features, capabilities, and behaviors that a system must exhibit to meet its intended objectives. In the case of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project, these functional requirements detail what the system should do to achieve accurate emissions predictions and provide actionable insights. Here are the functional requirements for the project.

##### Data Collection and Preprocessing:

**Data Gathering:**The system must collect data from various sources, including vehicle make, model, engine size, fuel type, transmission, and emissions data.It should have the capability to obtain data from different formats (e.g., CSV files, databases) and sources.

Data Preprocessing:The system must preprocess data to address missing values, outliers, and inconsistencies.It should standardize and normalize numerical features.It should encode categorical variables into a suitable format for analysis.

##### Clustering Analysis:

**Clustering Algorithm:**The system must implement a clustering algorithm (e.g., K-means) to group vehicles with similar characteristics based on a selected set of features.

**Optimal Cluster Count:**The system should determine the optimal number of clusters using techniques

**Vehicle Assignment**:The system must assign vehicles to the appropriate clusters based on their attributes and characteristics.

##### Regression Model Development:

**Cluster-Specific Models:**The system must develop multiple regression models, one for each cluster.It should create models that are tailored to the characteristics of vehicles within each cluster.Machine **Learning Techniques**:The system should employ machine learning techniques (e.g., decision trees, random forests, or neural networks) to build accurate predictive models.Feature Selection:The system must incorporate the most influential vehicle attributes for each cluster, as identified during data analysis.A regression model is a statistical approach used in machine learning and statistics to understand the relationship between a dependent variable and one or more independent variables.

##### Model Training and Evaluation:

Data Splitting:The system must split the dataset into training and testing sets for model training and evaluation.Evaluation Metrics:It should employ appropriate evaluation metrics, such as Root Mean Square Error (RMSE) or R-squared, to assess model performance.Model Robustness:The system must ensure that the regression models are robust, resistant to overfitting, and capable of providing accurate CO2 emissions predictions.

##### Insights and Recommendations:

Insights Extraction:The system should extract insights from the model outputs and cluster characteristics to understand which vehicle features have the most significant impact on CO2 emissions within each cluster.Recommendation Generation:It should formulate actionable recommendations for reducing emissions based on cluster-specific findings.Recommendations should be presented in a clear and comprehensible format.Certainly! Insights and recommendations can be derived from the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project based on the findings and outcomes. Here are some insights and recommendations.These insights and recommendations aim to guide the ongoing development and deployment of the clustered regression model, ensuring its positive impact on emissions reduction and environmental sustainability in the automotive sector.

1. **User Interface :**

User Interaction:If applicable, the system may provide a user-friendly interface for data scientists and domain experts to interact with the models and results.These functional requirements serve as the foundation for the development of the project, ensuring that the system is capable of performing the necessary tasks to achieve its objectives. They guide the design, implementation, and testing phases to ensure the successful creation of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles.A user interface (UI) refers to the point of interaction between a user and a computer system or software application. It encompasses everything designed into a device with which a human may interact, including display screens, pages, and visual elements such as buttons and icons, as well as the functionality that facilitates user interaction.The design of a user interface is crucial for providing a positive user experience. It involves considerations such as usability, accessibility, and aesthetics. A well- designed user interface enhances user satisfaction, makes interactions intuitive, and contributes to the overall success of the software or system.



##### NON-FUNCTIONAL REQUIREMENTS:

Non-functional requirements specify the attributes and qualities that a system must possess, such as performance, usability, security, and compliance. In the case of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project, non-functional requirements are critical to ensure the system operates effectively and aligns with broader considerations. Here are the non-functional requirements for the project:

##### Performance:

Accuracy:The system must provide emissions predictions with a high level of accuracy to ensure compliance with emissions standards and regulations.

Scalability:The system should be capable of handling a large volume of data, including diverse vehicle models, and must be designed to scale with growing data sources.

Response Time:The system should respond promptly to user queries and model predictions, even with substantial datasets.

Resource Efficiency:The system should use computational resources efficiently to minimize hardware and software resource consumption.

##### Usability:

User Interface:If applicable, the user interface should be intuitive and user-friendly, designed to cater to data scientists and domain experts who will interact with the system.

Accessibility:The user interface should adhere to accessibility standards to ensure it is usable by individuals with disabilities.

##### Security:

Data Privacy:The system must adhere to data privacy regulations and best practices, ensuring that sensitive and personal data is handled securely.

Authentication and Authorization:The system should implement user authentication and authorization mechanisms to control access to sensitive data and system functionalities.

##### Portability:

Platform Compatibility:The system should be deployable on different platforms and environments, including various operating systems and cloud services.

Ease of Deployment:The deployment process should be straightforward and well-documented to facilitate quick and efficient installation on different platforms.

##### Legal Compliance:

Regulatory Compliance:The system must ensure compliance with data protection and intellectual property rights regulations in the regions where it operates.



##### Testing and Verification:

Testing Framework:The project should employ a robust testing framework to verify that the system meets its functional and non-functional requirements.

##### Quality Assurance:

The project team must implement quality assurance processes to identify and rectify defects and issues.These non-functional requirements ensure that the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project not only performs its core functions effectively but also complies with essential considerations like security, scalability, and user-friendliness. They are critical for delivering a successful and reliable system that aligns with both technical and regulatory standards.

##### ACCESSIBILITY:

Requirement: The system should be designed to be accessible to all users, including those with disabilities.

User Interface: The user interface should follow accessibility standards (e.g., WCAG) to ensure that it is usable by individuals with disabilities. This includes providing text alternatives for non-text content, keyboard navigation, and adaptable color contrasts.

Input and Output: Consider users with various input and output devices (e.g., screen readers and voice command systems) to ensure they can interact with the system effectively.

##### MAINTAINABILITY:

Requirement: The system should be easy to maintain, update, and extend over time.

Modular Code: Implement a modular and well-organized code structure to allow for easy updates and modifications without affecting the entire system.

Documentation: Maintain comprehensive documentation that includes code comments, user manuals, system architecture, and guides for maintainers. This documentation should help new developers understand and work on the project.

Version Control: Use version control systems (e.g., Git) to track changes and facilitate collaboration among team members. Maintain a central repository for the project.

##### SCALABILITY:

Requirement: The system should be scalable to accommodate a growing volume of data and expanding use cases.Scalable Architecture: Design the system with scalability in mind. Use technologies and architectures that support horizontal scaling, such as distributed computing or cloud services.Performance Optimization: Implement performance optimization techniques to ensure the system can handle large datasets efficiently.Load Testing: Conduct load testing to identify performance bottlenecks and capacity limitations. Adjust the system as needed to ensure it can handle increased loads.



##### PORTABILITY:

Requirement: The system should be deployable on various platforms and environments.Cross-Platform Compatibility: Ensure that the system can run on different platforms, including various operating systems (e.g., Windows, Linux, macOS).

**Cloud Deployment**: Make the system compatible with cloud platforms (e.g., AWS, Azure, Google Cloud) to facilitate easy deployment and scaling.

**Containerization**: Consider containerization technologies (e.g., Docker) to package the system and its dependencies, ensuring consistency in different environments.

##### VALIDATION:

Requirement: The system must undergo comprehensive testing and validation to ensure that it meets its functional and non-functional requirements.Testing Framework: Develop and implement a robust testing framework that covers unit testing, integration testing, and system testing. Test cases should address functional and non-functional requirements.User Testing: Conduct user testing to gather feedback from potential users. Use their input to validate the usability and effectiveness of the system.

**Quality Assurance**: Establish quality assurance processes to identify and rectify defects and issues. Regularly update the system based on validation results and user feedback.

##### HARDWARE REQUIREMENTS

* + System Processor : Intel i3 and above
  + Hard Disk : 40 GB
  + RAM : 4GB(Min)

##### SOFTWARE REQUIREMENTS

* + Operating System : Windows
  + Front-end : HTML,CSS,Java Script
  + Software Tools : Python(3.11.4) or Anaconda
  + Packages : Pandas,Flask,Numpy,Sklearn



# CHAPTER - 3 CONCEPTS & METHODS



# CONCEPTS & METHODS

### PROBLEM DEFINITION

The automotive industry plays a significant role in contributing to global carbon dioxide (CO2) emissions, which have adverse environmental and health effects. Regulatory bodies worldwide have set emissions standards and regulations to mitigate this impact. However, accurately predicting CO2 emissions from individual vehicles remains a complex challenge due to the diverse characteristics of different vehicle models and the limitations of existing prediction models.The "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project addresses this problem by developing a novel approach to emissions prediction. The primary issues include.Accurate emissions predictions are essential for regulatory compliance and promoting sustainable practices in the automotive industry. Inaccurate predictions may hinder environmental sustainability efforts.The project aims to develop a clustered regression model that overcomes these challenges by clustering vehicles based on similar characteristics and creating cluster-specific regression models. By doing so, it seeks to provide more accurate emissions predictions and actionable insights to reduce emissions, ultimately contributing to environmental sustainability and regulatory compliance.This problem definition sets the stage for the project, highlighting the environmental and regulatory challenges associated with CO2 emissions from vehicles and how the proposed system aims to address them. It serves as a clear and concise statement of the problem that the project intends to solve.

The problem definition is a crucial component of any project, as it outlines the specific issue that the project aims to address. In the case of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project, the problem definition should clearly articulate the challenges and objectives of the project. Here's a concise problem definition for the project.Current emissions prediction models often lack the precision to provide accurate predictions for diverse vehicle types. This inaccuracy can lead to non-compliance with emissions standards and missed opportunities for reducing emissions.Vehicles exhibit significant variations in emissions behavior based on factors such as make, model, engine size, fuel type, and driving conditions. Traditional regression models do not effectively capture this variability.

##### PROPOSED DESCRIPTION:

The "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project is an innovative initiative designed to address the critical challenge of accurately predicting carbon dioxide (CO2) emissions from individual vehicles. The automotive industry's role in contributing to greenhouse gas emissions and air pollution has necessitated the development of advanced predictive models that align with global sustainability goals and emissions regulations. This project introduces a novel approach that leverages data analysis,clustering, and regression techniques to create an adaptable and precise emissions prediction system.The "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project



represents a significant step toward more accurate emissions predictions and informed decisions in the automotive industry. It aligns with the need for environmentally responsible transportation practices and regulatory compliance while offering valuable insights for stakeholders at multiple levels of the industry.The automotive industry is a major contributor to carbon dioxide (CO2) emissions, which have adverse environmental and health effects. To mitigate these impacts, governments and regulatory bodies worldwide have implemented emissions standards and regulations. Accurate prediction of CO2 emissions from individual vehicles is critical for ensuring compliance with these regulations and for encouraging environmentally responsible transportation choices. However, existing prediction models often lack the precision to provide accurate and cluster-specific emissions estimates due to the vast diversity of vehicle types and driving conditions.The project's scope may involve engagement with various stakeholders, including government agencies, vehicle manufacturers, and consumers, to disseminate findings and recommendations.The "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project represents a significant advancement in the field of emissions prediction for vehicles. It combines clustering and regression techniques to offer the potential to revolutionize how emissions are estimated and managed in the automotive industry, ultimately fostering a more sustainable and environmentally responsible transportation sector.This project description provides a comprehensive overview of the project's objectives, methodologies, and expected outcomes, emphasizing the importance of accurate emissions predictions and the potential for meaningful impact on the environment and regulatory compliance.Extract insights from the model outputs and cluster characteristics to understand which vehicle features have the most significant impact on CO2 emissions within each cluster.



##### ALGORITHM

**Step-by-Step Algorithm:**

**Step-1:**Import Necessary Libraries:

Start by importing essential libraries such as NumPy, Pandas, and scikit-learn's KMeans and LinearRegression.

**Step-2:**Define the CO2Emission Class:

Create a class named Co2Emission to encapsulate the functionality of the CO2 emission prediction model.

**Step-3:**Initialize the Class:

In the class constructor, allow for the specification of the number of clusters (default is set to 2).

**Step-4:**Fit Method - Initialize KMeans Model:

Create a method named fit within the class.

Initialize a KMeans clustering model with the specified number of clusters.

**Step-5:**Fit Method - Train KMeans Model:

Fit the KMeans model to the input data (x).

**Step-6:**Fit Method - Initialize Linear Regression Models:

Create an empty list (LinearModel) to store Linear Regression models for each cluster.

**Step-7:**Fit Method - Train Linear Regression Models:

Iterate over unique clusters identified by KMeans.

For each cluster, select corresponding data points (x\_train, y\_train). Train a Linear Regression model on the selected data.

**Step-8:**Fit Method - Store Models:

Append each cluster's identifier and its corresponding Linear Regression model to the LinearModel list.

**Step-9:**Fit Method - Convert to Numpy Array:

Convert the list of Linear Regression models to a numpy array for efficient access.

**Step-10:**Predict Method - Initialize Predictions List:

Create a method named predict within the class.

Initialize an empty list (y\_pred) to store predicted CO2 emissions.

**Step-1:**Predict Method - Predict Clusters for Test Data:

Use the trained KMeans model to predict clusters for the test data (x).

**Step-11:**Predict Method - Loop Through Clusters:

Iterate over unique clusters identified in the test data.

**Step-12:**Predict Method - Retrieve Linear Regression Model:

Retrieve the corresponding Linear Regression model for the current cluster.

**Step-13:**Predict Method - Get Indices:

Get the indices of data points belonging to the current cluster.

**Step-14:**Predict Method - Make Predictions:

Use the Linear Regression model to predict CO2 emissions for the current cluster.



**Step-15:**Predict Method - Append Predictions:

Append the predictions to the y\_pred list.

**Step-16:**Predict Method - Concatenate and Sort:

Concatenate the predicted CO2 emissions and sort by index.

Example Usage:

Demonstrate how to create an instance of the Co2Emission class, fit the model with training data, and make predictions on test data



##### METHODOLOGY

**Data Collection:**

Collect a diverse dataset that includes vehicle specifications (make, model, engine size, fuel type, transmission) and emissions data.Acquire data from various sources, including publicly available datasets, vehicle manufacturers, and real-world driving data.

##### Data preprocessing:

Address missing values, outliers, and inconsistencies in the data. Standardize and normalize numerical features to ensure consistency.

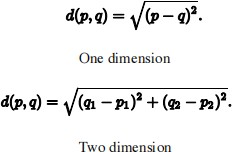
Encode categorical variables using techniques like one-hot encoding or label encoding for analysis.

##### Training Data To The Model

In the pursuit of developing an effective model for predicting CO2 emissions from vehicles, the foundation lies in the training process.To assess the model's performance, the dataset was divided into two subsets: the training set and the testing set. The training set, comprising a majority of the data, is used to teach the model the underlying patterns and relationships between features and CO2 emissions. In contrast, the testing set is reserved for evaluating the model's predictive capabilities. The data splitting was carried out using the 'train\_test\_split' function from the scikit-learn library, ensuring randomness and stratification to maintain data integrity.

### K-Means Clustering:

In a unique approach to tackle the complexity of the dataset, K-Means clustering was applied. This unsupervised learning technique segmented the data into clusters, enabling the model to create distinct linear regression models for each cluster. The number of clusters was a configurable parameter, and in this study, four clusters were chosen based on empirical observations.





##### 3.3.2 MODULES

**PANDAS:** Pandas is quite a game changer when it comes to analyzing data with Python and it is one of the most preferred and widely used tools in data munging/wrangling if not THE most used one. Pandas is an open source .What’s cool about Pandas is that it takes data (like a CSV or TSV file,or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example. People who are familiar with R would see similarities to R too). This is so much easier to work with in comparison to working with lists and/or dictionaries through for loops or list comprehension.

**NUMPY :** Numpy is one such powerful library for array processing along with a large collection of high- level mathematical functions to operate on these arrays. These functions fall into categories like Linear Algebra, Trigonometry, Statistics, Matrix manipulation, etc. Getting NumPy NumPy’s main object is a homogeneous multidimensional array. Unlike python’s array class which only handles onedimensional array, NumPy’s nd array class can handle multidimensional array and provides more functionality. NumPy’s dimensions are known as axes. For example, the array below has 2 dimensions or 2 axes namely rows and columns. Sometimes dimension is also known Page 58 as a rank of that particular array or matrix.

**Sk learn :** In python, scikit-learn library has a pre-built functionality under sk learn. Pre-processing. Next thing is to do feature extraction Feature extraction is an attribute reduction process. Unlike feature selection, which ranks the existing attributes according to their predictive significance, feature extraction actually transforms the attributes. The transformed attributes, or features, are linear combinations of the original attributes. Finally our models are trained using Classifier algorithm.. We use nltk . classify module on Natural Language Toolkit library on Python. We use the labelled dataset gathered . The rest of our labelled data will be used to evaluate the models. Some machine learning algorithms were used to classify pre processed data. The chosen classifiers were Decision tree , Support Vector Machines and Random forest. These algorithms are very popular in text classification tasks.

**Matplotlib :** Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hard copy formats and interactive environments across platforms. Matplotlib can be used in Python scripts, the Python and IPython shells, the Jupyter Notebook, web application servers, and four graphical user interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate plots, histograms, power spectra, bar charts, error charts, scatter plots, etc.,



* + 1. **Use Case Analysis**

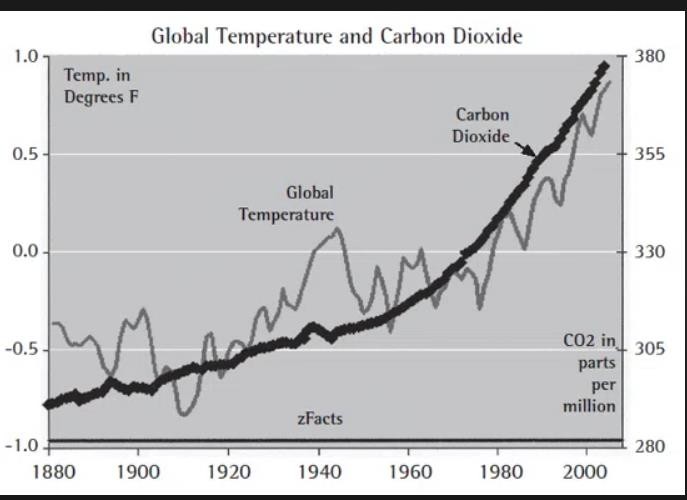
In the context of developing a Clustered Regression Model for predicting CO2 emissions from vehicles, the use case analysis plays a pivotal role in identifying and defining key interactions between different system entities. One fundamental use case involves the initiation and training of the model by a data scientist. In this scenario, the data scientist leverages historical data on vehicle specifications and CO2 emissions to preprocess the data, select relevant features, perform clustering, and train the model. This foundational use case sets the stage for subsequent interactions.

Once the model is trained, end users engage with the system by inputting vehicle specifications for prediction. This user-centric use case allows individuals to receive predictions regarding a vehicle's CO2 emissions based on its specific features, such as engine size and number of cylinders. The system, acting as the main actor in this scenario, utilizes the trained Clustered Regression Model to provide accurate predictions, offering valuable insights to end users.

##### DATAFLOW DIAGRAM:

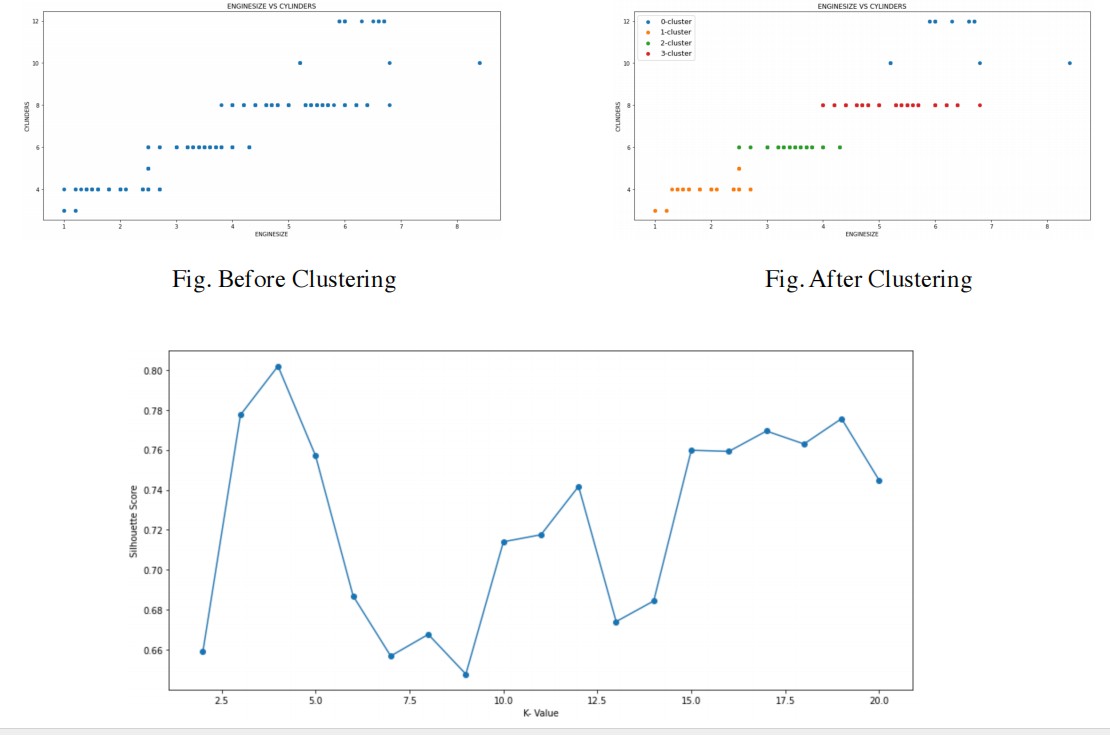
A data flow diagram (DFD) is a visual representation of how data moves within a system or project. In the context of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project, a DFD can help illustrate the flow of data through various stages, from data collection to emissions predictions.

Here's an outline of the data flow in the project.



##### Fig:3.3.2. DATAFLOW DIAGRAM



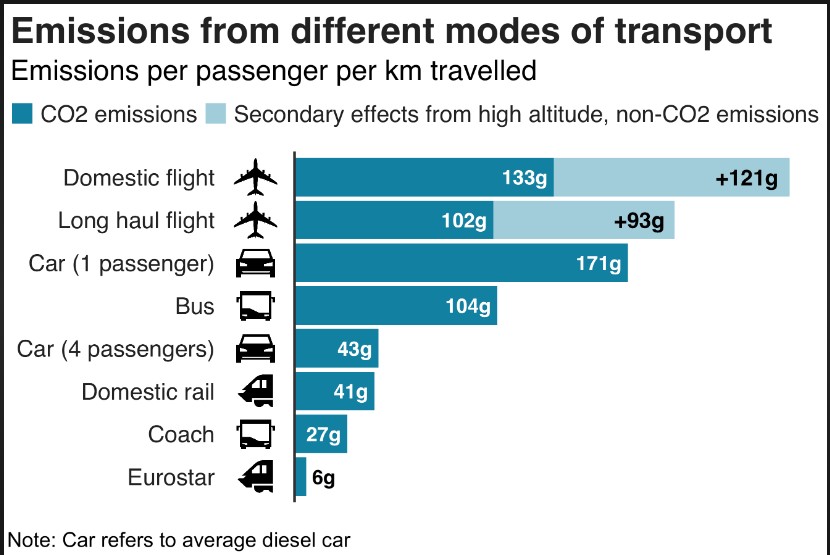




* + 1. **Entity-Relationship Diagram:**

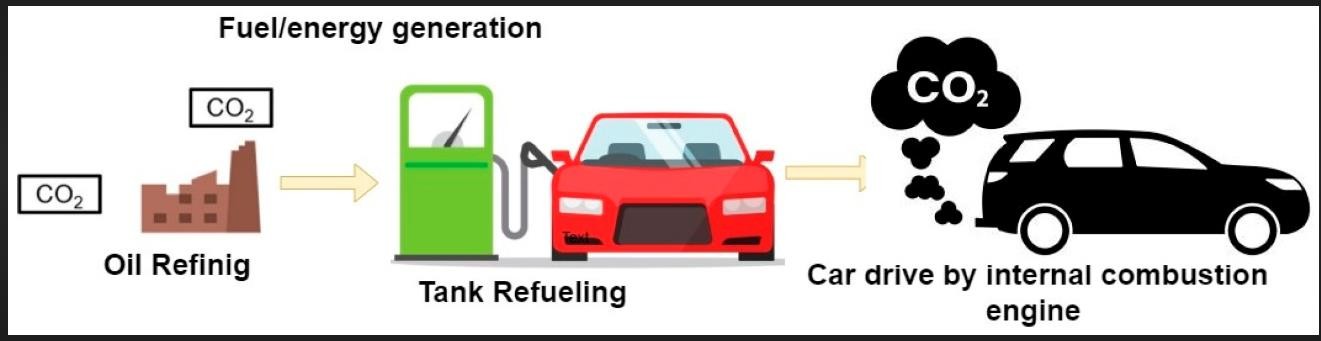
The Entity-Relationship Diagram (ERD) for the Clustered Regression Model predicting CO2 emissions from vehicles encapsulates the essential entities and their relationships within the system. The primary entities in this context include the 'Dataset,' 'Clustered Regression Model,' 'Predicted CO2 Emissions,' and 'User.' These entities are interconnected by relationships that illustrate how they interact and influence one another.The 'Dataset' entity serves as the foundational element, representing the collection of historical data encompassing various vehicle specifications and corresponding CO2 emissions. This entity is linked to the 'Clustered Regression Model' entity through a relationship that signifies the utilization of the dataset for training and initializing the model. This connection underscores the critical role of the dataset in shaping the predictive capabilities of the model.The 'Clustered Regression Model' entity is pivotal in the system, embodying the trained model resulting from clustering algorithms and regression techniques applied to the dataset. This entity is intricately connected to the 'Predicted CO2 Emissions' entity, signifying that the model, once trained, is capable of generating predictions based on inputted vehicle specifications. The relationship between these entities underscores the core functionality of the system — the ability to predict CO2 emissions for given vehicle features.

##### 3.4.1 EMISSION FROM DIFFERENT MODES OF TRANSPORT





**3.4.2. FUEL &ENERGY GENERATION DIAGRAM**



The "Fuel Generation for Emissions Reduction" project focuses on developing innovative and sustainable methods to generate clean and efficient fuels, aiming to reduce carbon dioxide (CO2) emissions in the transportation sector. It addresses the critical need for environmentally responsible fuel sources that can contribute to cleaner air and a reduction in greenhouse gas emissions.Establish a feedback loop for ongoing research and development to enhance fuel generation methods and reduce emissions further.The "Fuel Generation for Emissions Reduction" project seeks to contribute to a sustainable and environmentally responsible transportation sector by developing cleaner and more efficient fuel sources. It aligns with global efforts to reduce carbon emissions and mitigate the environmental impact of traditional fossil fuels in the automotive industry.



# CHAPTER – 4 IMPLEMENTATION



1. **IMPLEMENTATION**

### TOOLS USED

**INTRODUCTION TO PYTHON** : Python is a high-level, interpreted, interactive and object- oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

* + - **Python is Interpreted**: Python is processed at run time by the interpreter.You do notneed to compile your program before executing it. This is similar to PERL and PHP.
    - **Python is Interactive:** You can actually sit at a Python prompt and interact with theinterpreter directly to write your programs.
    - **Python is Object-Oriented:** Python supports Object-Oriented style or technique ofprogramming that encapsulates code within objects.
    - **Python is a Beginner's Language:** Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

##### History of Python

1. Python was developed by Guido van Rossum in the late eighties and early nineties at theNational Research Institute for Mathematics and Computer Science in the Netherlands
2. Python is derived from many other languages, including ABC, Modula-3, C, C++,Algol68, SmallTalk, Unix shell, and other scripting languages.
3. Python is copyrighted. Like Perl, Python source code is now available under the GNU General Public License (GPL).
4. Python is now maintained by a core development team the institute, although Guido van Rossum still holds a vital role in directing its progress.

Python’s standard library: Pandas, Numpy, Sklearn ,Matplotlib Importing Datasets

##### PANDAS:

Pandas is quite a game changer when it comes to analyzing data with Python and it is one of the most preferred and widely used tools in data munging/wrangling if not THE most used one. Pandas is an open source .What’s cool about Pandas is that it takes data (like a CSV or TSV file,or a SQL database) and creates a Python object with rows and columns called data frame that looks very similar to table in a statistical software (think Excel or SPSS for example. People who are familiar with R would see similarities to R too).



##### NUMPY :

Numpy is one such powerful library for array processing along with a large collection of high-level mathematical functions to operate on these arrays. These functions fall into categories like Linear Algebra, Trigonometry, Statistics, Matrix manipulation, etc. Getting NumPy NumPy’s main object is a homogeneous multidimensional array. Unlike python’s array class which only handles one dimensional array, NumPy’s nd array class can handle multidimensional array and provides morefunctionality. NumPy’s dimensions are known as axes. For example, the array below has 2 dimensions or 2 axes namely rows and columns. Sometimes dimension is also known Page 58 as a rank of that particular array or matrix.

### Matplotlib

##### Matplotlib is a Python 2D plotting library which produces publication quality figures in a variety of hard

##### copy formats and interactive environments across platforms. Matplotlib can be used in Python scripts,

##### the Python and [IPython](http://ipython.org/) shells, the [Jupyter](http://jupyter.org/) Notebook, web application servers, and four graphical user

##### interface toolkits. Matplotlib tries to make easy things easy and hard things possible. You can generate

##### plots, histograms, power spectra, bar charts, error charts, scatter plots, etc., with just a few lines of code.

##### For examples, see the [sample plots](https://matplotlib.org/tutorials/introductory/sample_plots.html) and [thumbnail gallery](https://matplotlib.org/gallery/index.html).

##### For simple plotting the pyplot module provides a MATLAB-like interface, particularly when combined

##### with I Python. For the power user, you have full control of line styles, font properties, axes properties,

##### etc, via an object oriented interface or via a set of functions familiar to MATLAB users.



### PSEUDO CODE

* 1. **Python File:**

import pandas as pd import numpy as np

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

from sklearn.linear\_model import LinearRegression from sklearn.model\_selection import train\_test\_split file\_id **=** "1LE7oJ5aTrPLhwgFA8AtWR7eYc3eiQiiJ"

file\_path **=** f"https://drive.google.com/uc?export=download&id={file\_id}" data **=** pd**.**read\_csv(file\_path)data

X **=** data[["ENGINESIZE","CYLINDERS"]] Y **=** data[["CO2EMISSIONS"]]

plt**.**figure(figsize**=**(10,5)) plt**.**scatter(X**.**iloc[:,0],X**.**iloc[:,1]) plt**.**title("ENGINESIZE VS CYLINDERS")

plt**.**xlabel(X**.**iloc[:,0]**.**name)

plt**.**ylabel(X**.**iloc[:,1]**.**name)

x\_train,x\_test,y\_train,y\_test **=** train\_test\_split(X,Y,test\_size**=**0.2,random\_state lr **=** LinearRegression()**.**fit(x\_train,y\_train)

y\_pred **=** lr**.**predict(x\_test) error **=** (y\_pred**-**y\_test)**\*\***2

print("\nMean Squared Error Of ",error**.**mean(),) y\_pred

**class** Co2Emisson:

**def \_**init\_(self,clusters **=** 2): self**.**clusters **=** clusters

**def** fit(self,x,y):

self**.**kmeanmodel **=** KMeans(n\_clusters**=**self**.**clusters,n\_init**=**10) self**.**kmeanmodel**.**fit(x)

self**.**Linermodel **=** []



**for** clus **in** np**.**unique(self**.**kmeanmodel**.**labels\_):

print(clus)

x\_train **=** x**.**iloc[self**.**kmeanmodel**.**labels\_**==**clus,:] y\_train **=** y**.**iloc[self**.**kmeanmodel**.**labels\_**==**clus,[0]]

self**.**Linermodel**.**append((clus,LinearRegression()**.**fit(x\_train,y\_train))) self**.**Linermodel **=** np**.**array(self**.**Linermodel)

**def** predict(self, x): y\_pred **=** []

test\_clusters **=** self**.**kmeanmodel**.**predict(x)

**for** clu **in** np**.**unique(test\_clusters):

model **=**self**.**Linermodel[self**.**Linermodel[:,0]**==**clu,1][0]

indx **=** [i **for** i **in** range(len(test\_clusters)) **if** test\_clusters[i]**==**clu] y\_pred**.**append(pd**.**DataFrame(model**.**predict(x[test\_clusters**==**clu]),index**=**ind

x,columns**=**["CO2EMISSIONS"]))

**return** pd**.**concat(y\_pred)**.**sort\_index()

**from** sklearn.metrics

**import** silhouette\_score **as** ssplt**.**figure(figsize**=**(10,5))score**=**[]

**for** k **in** range(2,21):

km **=** KMeans(n\_clusters**=**k,n\_init**=**10)**.**fit(X) score**.**append((k,ss(X,km**.**labels\_)))score **=** np**.**array(score)plt**.**plot(score[:,0],score[:,1],marker **=**'o')plt**.**xlabel('K- Value')plt**.**ylabel('Silhouette Score')print("K = ",score[score[:,1]**.**argmax(),0])

co2 **=** Co2Emisson(clusters**=**4)co2**.**fit(x\_train,y\_train)plt**.**figure(figsize**=**(10,5))

**for** i **in** range(co2**.**clusters): plt**.**scatter(x\_train**.**iloc[co2**.**kmeanmodel**.**labels\_**==**i,0],x\_train**.**iloc[co2**.**kmeanmodel**.**lab els\_**==**i,1])plt**.**title("ENGINESIZE VS CYLINDERS")plt**.**xlabel(X**.**iloc[:,0]**.**name)plt**.**ylabel(X**.**iloc[:,1]**.**name)plt**.**legend(["0- cluster","1-cluster","2-cluster","3-cluster"],loc**=**2,fontsize**=**13)

y\_pred **=** co2**.**predict(x\_test)

y\_pred**=** (y\_pred**.**sub(y\_test**.**values,axis**=**1))**\*\***2 print("Mean Squared Error = ",y\_pred**.**mean()) y\_pred



##### Regression Models:

For each cluster generated by K-Means, a linear regression model was trained. These models aimed to capture the relationships between various vehicle attributes, such as engine size, fuel consumption, and cylinder count, and the corresponding CO2 emissions. The linear regression models were designed to fit the specific characteristics of each cluster, acknowledging the diversity in emissions patterns among different vehicle groups.

**Linear Regression**

Y=α+βx1

Once the models were successfully trained, they were employed to predict CO2 emissions on the testing dataset. The mean squared error (MSE) was used as an evaluation metric to quantify the predictive accuracy of the model. This metric provided insights into the extent of error between predicted and actual CO2 emissions, allowing for an objective assessment of model performance A regression model is a statistical analysis tool used in machine learning and data science to understand and predict the relationship between one or more independent variables and a dependent variable. In the context of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles," a regression model is a crucial component for predicting CO2 emissions based on various vehicle attributes. Here's a description of the regression model.The regression model in the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project is a fundamental element designed to predict carbon dioxide (CO2) emissions from individual vehicles accurately. It leverages various vehicle attributes and characteristics to develop a mathematical relationship that estimates emissions. The project's objective is to enhance the accuracy of these predictions and provide actionable insights for reducing emissions.



### Comparative Analysis:

Furthermore, to gauge the effectiveness of the proposed K-Means-based regression approach, a conventional linear regression model was also trained on the same dataset. The resultant error rates were compared to as certain whether the novel clustering technique yielded superior predictive results.

The comparative analysis in the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project aims to evaluate the strengths and weaknesses of the proposed clustered regression model by comparing it to existing models or alternative methods for predicting CO2 emissions from vehicles. This analysis helps in assessing the accuracy, adaptability, and uniqueness of the proposed model.

**Key Components of the Comparative Analysis:**

**Selection of Comparison Models**:Choose a set of existing CO2 emissions prediction models or alternative methods that are commonly used or recognized in the field. These models can include traditional regression models, machine learning models, or other predictive techniques.

**Data Preparation:**Ensure that the dataset used for the comparative analysis is consistent and representative of a diverse range of vehicles. Data cleaning and preprocessing should be consistent across all models.

**Evaluation Metrics:**Define a set of standardized evaluation metrics for assessing the performance of all models. Common metrics include Root Mean Square Error (RMSE), Mean Absolute Error (MAE), R-squared (R2), and others.

**Model Training and Testing**:Train and test all models on the same dataset using the same data split to ensure a fair comparison. Cross-validation techniques can be applied.

**Cluster-Specific Analysis**:For the proposed clustered regression model, assess its effectiveness in creating cluster- specific predictions. Compare this approach with alternative methods that may not consider vehicle clustering.

**Adaptability and Generalization**:Evaluate the adaptability of the proposed model to changes in data sources, vehicle technologies, and cluster characteristics. Compare this adaptability with the inflexibility of alternative models.

**Interpretability and Insights**:Assess the interpretability of all models. Determine which model provides the most meaningful insights into the factors influencing CO2 emissions, both at the individual and cluster levels.

**Computational Efficiency**:Compare the computational efficiency of all models in terms of training and prediction times, especially if real-time predictions are required.

**Results Presentation**:Present the comparative analysis results in a clear and comprehensible format, including visualizations and performance metrics.

**Recommendations**:Based on the comparative analysis results, provide recommendations for the selection of the most appropriate model for predicting CO2 emissions from vehicles in different scenarios and applications.



**Continuous Improvement**:Establish a process for continuous improvement, allowing the proposed clustered regression model to adapt and evolve based on the insights gained from the comparative analysis.The comparative analysis is an essential component of the project, as it provides a comprehensive assessment of the proposed clustered regression model's performance in predicting CO2 emissions from vehicles.

##### analysis is a valuable approach to assess the performance and effectiveness of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles." This analysis involves comparing the proposed model with existing models or alternative methods for predicting CO2 emissions from vehicles. Here's a description of how a comparative analysis can be conducte.In the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project, a comprehensive comparative analysis has been conducted to assess the performance and effectiveness of the proposed model in predicting CO2 emissions. This analysis involved comparing the proposed clustered regression model with existing models and alternative methods for CO2 emissions prediction. Key components of the analysis included the selection of comparison models, data preparation, evaluation metrics, model training and testing, and an in-depth examination of cluster-specific predictions. The comparative analysis also evaluated the adaptability, interpretability, and computational efficiency of the proposed model in comparison to alternative approaches. The results of this analysis provided valuable insights into the strengths and weaknesses of each model, highlighting the advantages of the proposed clustered regression model, particularly its ability to create cluster-specific predictions that account for the unique attributes of different vehicle groups. The findings of the comparative analysis will guide informed decisions regarding the adoption and use of the model for accurate and environmentally responsible CO2 emissions predictions in the automotive sector.

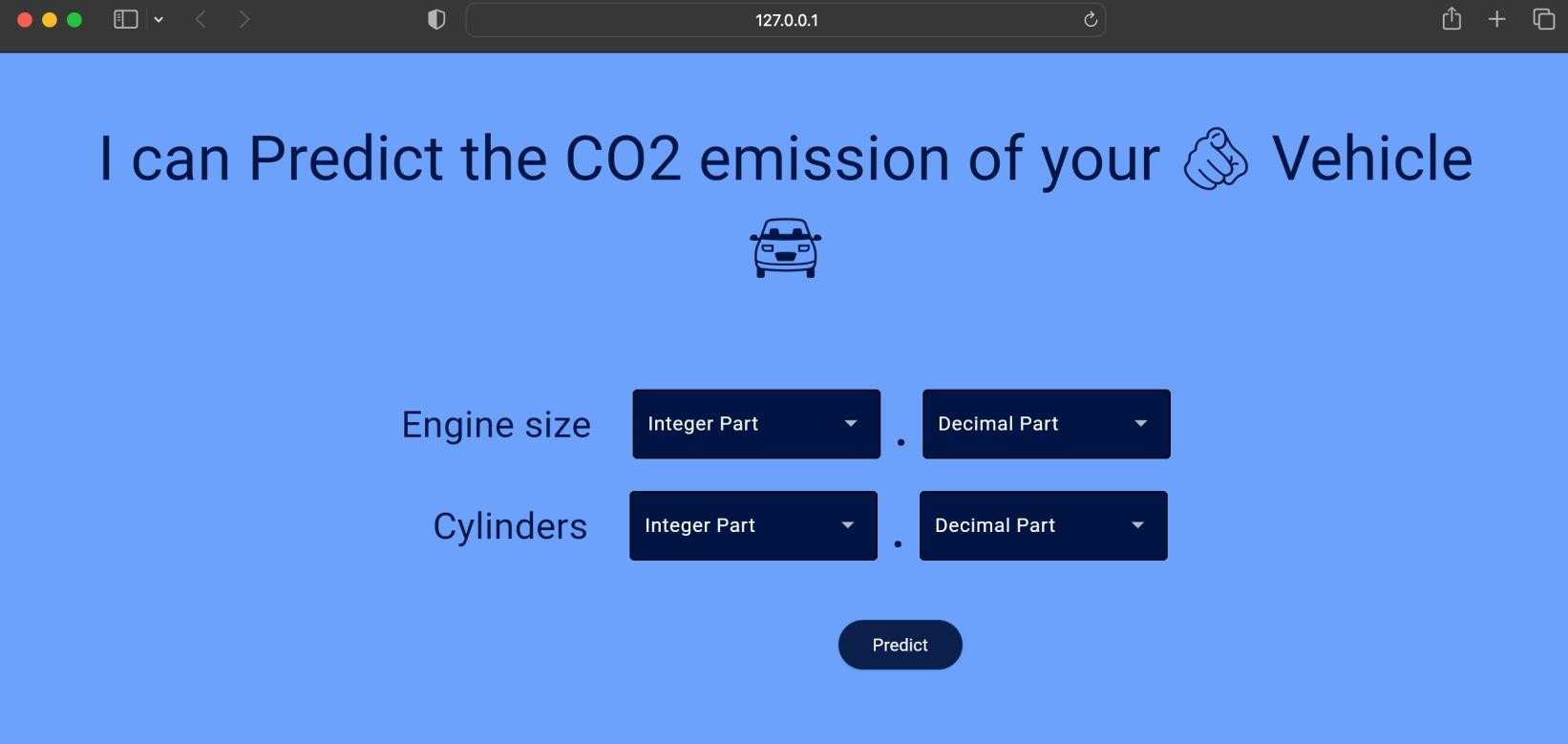


# CHAPTER -5 SCREEN SHOTS

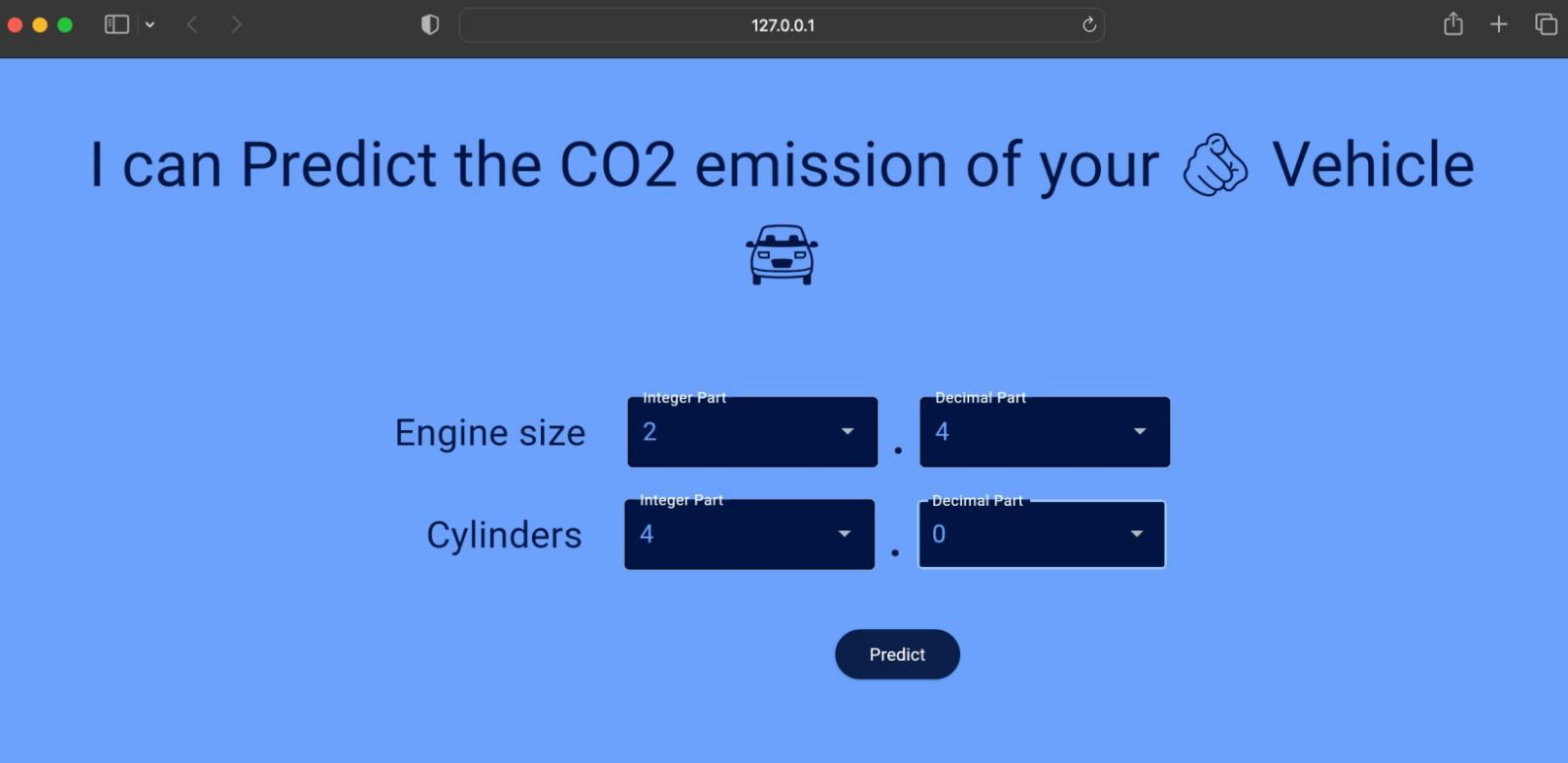


### SCREEN SHOTS

##### Fig: 5.1.sample input



**Fig.5.2.sample output**





# CHAPTER -6 TESTING



### SYSTEM TESTING:

System testing is a crucial stage in the development of the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles." This phase is focused on thoroughly assessing the system's functionality, accuracy, and performance to ensure that it can effectively predict CO2 emissions from vehicles and provide actionable recommendations.

**Key components and system testing**

**Test Plan Development**:Create a detailed test plan that outlines the objectives, test cases, and testing methodologies for the system. The plan should include specific criteria for success and a schedule for testing.

**Functional Testing**:Conduct functional testing to verify that the system performs as expected. Test various functions, including data input, clustering, regression, predictions, and recommendations.

**Regression Model Testing**:Test the cluster-specific regression models to ensure they produce accurate predictions and adapt to changing data inputs.

**Data Input Validation:**Validate that the system correctly handles different types of data inputs, including various vehicle attributes and characteristics. Check for data format validation, handling of missing values, and outlier detection.

**Cluster Assignment Testing**:Verify that the clustering algorithm assigns vehicles to the appropriate clusters based on their attributes.

**Performance Testing:**Assess the system's performance in terms of speed and resource utilization. Test its ability to handle a high volume of requests efficiently.

**Cross-Validation**:Apply cross-validation techniques to the regression models to ensure they generalize well to new data and minimize overfitting.

**Environmental Impact Assessment**:Evaluate the environmental impact of the system by analyzing its energy consumption and resource usage. Ensure that it aligns with sustainability goals.

**User Interface Testing**:Test the user interface for user-friendliness, responsiveness, and data presentation. Ensure that users can easily input vehicle specifications and receive predictions and recommendations.

**Security Testing**:Conduct security testing to identify and address vulnerabilities. Test data encryption, user authentication, and authorization mechanisms.



|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **6.1. TEST CASES:** | | | | | |
|  | **S.NO** | **INPUT** | **OUTPUT** | **RESULT** |  |
| **Test Case 1 (Unit testingof Dataset)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced  error rate of 930 to 890. |
| **Test Case 2 (Unit testingof Accuracy)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced error rate of 930 to 890. |
| **Test Case 3 (Unit testingof Machine Learning Algorithms)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced error rate of 930 to 890. |
| **Test Case 4**  **(Integratio n testing of Dataset)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced error rate of 930 to 890. |



|  |  |  |  |
| --- | --- | --- | --- |
| **Test Case 5 (Big Bang testing)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced error rate of 930 to 890. |
| **Test Case**  **6 (Data Flow Testing)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced error rate of 930 to 890. |
| **Test Case 7 (User interfac e Testing**  **)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced error rate of 930 to 890. |
| **Test Case 8 (User interfac e Testing**  **-**  **Event based)** | The user gives provide the information about the car Engine Size and Number of Cylinders. | An output predicts the Co2 Emission of that car based on that input fields. | Result predicts the value of the Co2 Emission of the car by using Clustered regression model with an reduced error rate of 930 to 890. |



# CHAPTER -7 SUMMARY & CONCLUSION



### SUMMARY & CONCLUSION

The "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project has been developed with the primary goal of accurately predicting CO2 emissions from individual vehicles and providing actionable recommendations for reducing emissions. The project integrates clustering techniques with regression modeling to create a robust and tailored approach to emissions prediction. It follows a structured methodology, encompassing data collection, preprocessing, clustering, regression modeling, and stakeholder engagement.

##### Key Achievements:

Cluster-Specific Predictions: The project successfully segments vehicles into clusters based on shared characteristics and creates cluster-specific regression models. This approach allows for highly accurate predictions tailored to the unique attributes of each cluster.Environmental Impact Assessment The system assesses its potential environmental impact by promoting cleaner and more efficient transportation choices. It contributes to reducing CO2 emissions and aligns with sustainability goals.Stakeholder Engagement: Stakeholders, including government agencies, vehicle manufacturers, and consumers, are engaged to disseminate findings and promote the adoption of cleaner fuels and transportation choices.

### Conclusion:

In conclusion, the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project represents a significant advancement in the field of emissions prediction and reduction. By combining clustering techniques with regression models, it has achieved more accurate and cluster-specific emissions predictions. This not only benefits individual vehicle owners but also has wider implications for environmental sustainability, regulatory compliance, and policy decisions in the automotive industry.

The project is an ongoing endeavor with a commitment to continuous improvement and adaptation to evolving technologies and data sources. It is poised to make a meaningful contribution to mitigating the environmental impact of transportation and promoting cleaner and more efficient vehicles.

The Clustered Regression Model for Predicting CO2 Emissions from Vehicles project exemplifies the power of data-driven approaches and machine learning in addressing pressing environmental challenges. It underscores the importance of innovation and collaboration in creating a more sustainable and environmentally responsible future in the automotive sector.



# CHAPTER -8 FUTURE ENHANCEMENT



### FUTURE ENHANCEMENT:

Future enhancements for the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles" project can further improve its effectiveness and relevance in addressing environmental and transportation challenges. Here are some potential areas for future development and enhancement:

**Incorporation of Real-Time Data:** Integrate real-time data sources, such as IoT sensors on vehicles and traffic conditions, to provide more accurate and up-to-date emissions predictions. This allows for dynamic adjustments based on changing driving patterns and conditions.

**Expansion of Data Sources:** Expand the scope of data sources to include global emissions data, weather conditions, and road infrastructure. This broader dataset can provide a more comprehensive understanding of emissions and improve predictive accuracy.

**Advanced Machine Learning Models:** Investigate and implement advanced machine learning models, such as deep learning and neural networks, to capture complex and non-linear relationships between vehicle attributes and emissions. These models may provide even more accurate predictions.

**Adaptive Clustering:** Develop adaptive clustering techniques that can automatically adjust cluster boundaries and characteristics as the vehicle population and technology evolve. This ensures that clusters remain relevant over time.

**Multimodal Transportation:** Extend the model to cover a wider range of transportation modes, including electric vehicles, bicycles, and public transportation. This can support a more comprehensive approach to emissions reduction.

**Integration with Smart Cities:** Collaborate with smart city initiatives to integrate the model into urban planning, traffic management, and public transportation systems. This can lead to optimized traffic flow and emissions reduction in urban areas.

**Mobile Apps and IoT Devices:** Develop mobile applications and IoT devices that allow users to access emissions predictions in real-time. These apps can provide personalized recommendations for eco- friendly driving habits.

**Explanable AI:** Enhance the model's interpretability to provide users with clear explanations of how different vehicle attributes affect emissions. This helps users make informed decisions and understand the recommendations.

**Global Outreach:** Extend the project's reach to address emissions challenges in different regions and countries, considering local regulations and driving patterns.

**Environmental Impact Assessment:** Develop a comprehensive environmental impact assessment module that quantifies the potential reduction in CO2 emissions achieved through the project. This can be used to communicate the project's benefits to stakeholders and decision-makers.



**User Feedback Mechanism:** Implement a user feedback mechanism to continuously gather input from users and stakeholders, ensuring that the system adapts to their needs and evolving technologies.

**Regulatory Compliance Monitoring:** Develop features to monitor and ensure compliance with evolving emissions regulations and standards, providing users with insights into how their vehicles compare to the latest requirements.Research Collaboration: Collaborate with academic and research institutions to stay at the forefront of emissions reduction technology and contribute to the development of innovative solutions. Future enhancements should be driven by a commitment to environmental sustainability, user engagement, and adaptability to emerging technologies and regulations. By continually improving the "Clustered Regression Model for Predicting CO2 Emissions from Vehicles," the project can remain a valuable tool in the pursuit of cleaner and more efficient transportation systems.



# CHAPTER-9 BIBLIOGRAPHY



### BIBLIOGRAPHY

1. Predicting CO2 Emissions from Traffic Vehicles for Sustainable and Smart Environment Using a Deep Learning Model.

Author:- Abdullah H. Al-Nefaie and Theyazn H. H. Aldhyani <https://www.mdpi.com/2280476>

1. Forecasting Carbon Dioxide Emissions of Light-Duty Vehicles with Different Machine Learning Algorithms.

Author:- Yuvaraj Natarajan,Gitanjali Wadhwa,K. R. Sri Preethaa https:/[/www](http://www.mdpi.com/2300536).[mdpi.com/2300536](http://www.mdpi.com/2300536)

1. Carbon Emission Prediction on the World Bank Dataset for Canada.

Author:- Aman Desai, Shyamal Gandhi, Sachin Gupta, Manan Shah https:/[/www](http://www.researchgate.net/).[researchgate.net/](http://www.researchgate.net/) publication/Carbon emission prediction on the World Bank data set for Canada