2023 2nd International Conference for Innovation in Technology (INOCON)
Bangalore, India. Mar 3-5, 2023

CO_2 Emission Rating by Vehicles Using Data Science

Dr. Anitha Julian
Department of Computer Science Engineering
Saveetha Engineering College
Chennai, TamilNadu, India
anithajulia@saveetha.ac.in

Karthick S
Department of Computer Science Engineering
Saveetha Engineering College
Chennai, TamilNadu, India
karthicksadha24@gmail.com

Chandrakishore T
Department of Computer Science Engineering
Saveetha Engineering College
Chennai, TamilNadu, India
kishorethirumal07@gmail.com

Chandramohan A
Department of Computer Science Engineering
Saveetha Engineering College
Chennai, TamilNadu, India
chandramohan1022001@gmail.com

2023 2nd International Conference for Innovation in Technology (INOCON) | 979-8-3503-2092-3/23/$31.00 ©2023 IEEE | DOI: 10.1109/INOCON57975.2023.10101272

***Abstract* — The usage of private transportation is a significant contributor to the exacerbation of global warming. When a gallon of gasoline is burned in a car's engine, it produces approximately 24 pounds of greenhouse gases, which contribute to about 20% of total emissions. Most of these emissions, over 19 pounds, are released directly from the car's tailpipe as heat-trapping pollutants. However, the number of emissions produced during the fuel's extraction, manufacture, and delivery processes is relatively small in comparison. On average, gasoline-powered vehicles that are commonly used on roads around the world have a fuel efficiency of 22 miles per gallon and travel 11,500 miles per year. For every gallon of fuel used, these vehicles produce about 8,887 grams of carbon dioxide. In 1998, the auto industry made a voluntary pledge to cut emissions from new cars by 25 percent by 2008. At that time, new cars' CO_2 emissions on the road were roughly 203 gram per kilometer. They are currently hovering at 170 gram per kilometer and won't likely drop to 140g/km until around 2020. The amount of carbon dioxide emitted by a vehicle can vary depending on factors such as the type of gasoline used, the vehicle's fuel efficiency, and the distance it travels in a year. The projected accuracy decreases as the number of controlled and uncontrolled effect variables that affect the properties of CO_2 increases. Nevertheless, by taking into account the controllable effect factors and their interactions, a few experimental designs have been proposed. The Road and Transport Authority will seize that specific car if the model we developed to anticipate gas emission from cars exceeds the threshold. The model uses the properties of the car to specify if the car has exceeded the threshold value of CO_2. One excellent method for forecasting the CO_2 emission rating is supervised machine learning.**

Keywords — CO_2, accuracy, global warming, emission.

I. INTRODUCTION

The issue of global warming is a problem that affects all nations, as affirmed by the Intergovernmental Panel on Climate Change, which claims that human activities, including the increase of greenhouse gas emissions, are the main cause of the planet's warming by over 95%. Greenhouse gases, including carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), hydrofluorocarbons (HFC), and perfluorocarbons (PFC), are classified as acidic gases that can upset the balance between the Earth and the atmosphere. Among these gases, CO_2 is the main contributor to global warming. According to the International Energy Agency (IEA), global carbon dioxide emissions from fossil fuel combustion reached a record high of around 33.1 billion tons in 2019. Of this total, around 8 billion tons were emitted from the use of fossil fuels for power generation, transportation, and heating. This is a significant contributor to climate change and global warming, as carbon dioxide is a greenhouse gas that traps heat in the Earth's atmosphere, leading to rising temperatures and other negative impacts on the environment and human societies. Carbon dioxide emissions are a consequence of burning water (H_2O) and carbon monoxide (CO), also known as carbon dioxide (CO_2), a greenhouse gas.

II. EXISTING SYSTEM

The accurate modeling of vehicle emissions has been a long-standing goal in transportation research, especially with the increasing importance of evaluating the impact of intelligent transportation systems on the environment. However, current car emission models suffer from either oversimplified assumptions or excessive complexity and prior knowledge requirements, leading to low accuracy. To overcome these limitations, a new microscopic emission model based on deep learning was proposed in this study, providing highly precise CO_2 emission estimates for individual vehicles. To align the observed driving condition data with the measured emission data, which were initially out of sync, we employed the dynamic time wrapping technique.

III. PROPOSED SYSTEM

Data was initially acquired for the study using online tools. like a few distinctive characteristics. Before training the model, the dataset must be preprocessed after being collected from multiple sources. Reading the obtained dataset is the first step in the data preprocessing process, which follows with data cleaning. Certain duplicate features are present in the datasets after data cleansing, but the attributes are not taken into account when predicting CO_2 emissions. Therefore, we must remove unnecessary attributes and datasets with some missing information. To improve accuracy, we must remove these missing values or replace them with undesirable nan values. With the aid of statistical algorithms and machine learning, it is feasible to forecast outcomes based on historical data

IV. LITERATURE SURVEY

A literature review is a critical analysis of published works, such as books, articles, and other academic sources, related to a specific research question or topic. The purpose of a literature review is to provide a comprehensive overview of the existing knowledge and research in a particular field, identify gaps or inconsistencies in the literature, and suggest directions for future research. It usually precedes a research proposal and can be a simple list of references. The structure of a literature review typically involves synthesis and summarization.

A. Review of Literature Survey

According to [1] [2], The concentration of carbon dioxide CO_2 in the Earth's atmosphere has indeed increased from pre-industrial levels of around 280 parts per million (ppm) to current levels of around 410 ppm, primarily as a result of human activities such as burning fossil fuels, deforestation, and land use changes. This increase in atmospheric CO_2 has contributed to the greenhouse effect, which causes the Earth's surface to warm by trapping heat from the sun. The most affluent nations are the largest contributors of greenhouse gas emissions, and there is a notable association between gaseous emissions and the gross domestic product (GDP) of countries The relationship between technological advancements, local and international regulations, and their impact on complex systems is nonlinear and dynamic. Soft computing methods have proven to be effective in analyzing such interactions, as they can offer a concise approach to multi-variable parameters without the need for explicit knowledge of fundamental system properties. The proposed work presents a novel transfer learning approach, which involves training models on data from both developed and emerging nations to estimate the per capita Gross Domestic Product of diverse countries based on their CO_2 emissions.

The aim of this study is to accurately measure and quantify carbon emissions, promote energy efficiency, and contribute to the government's carbon peak policy goals in the mid- and long-term. The study begins by examining the carbon emissions and its peak vales using a proposed methodology. Moreover, the study utilizes a Long Short-Term Memory Neural Network (LSTM) to construct a model capable of predicting carbon emissions for any future research endeavors. By taking into account crucial variables such as industry investment, labor productivity, and carbon emissions intensity that impact carbon emissions within the " region of study, the research was able to devise suitable training and prediction models that effectively estimate carbon emissions data with high accuracy. The study suggests that the SVR model is more effective in predicting complex non-linear situations when compared to LSTM training. By utilizing deep learning models, the region can now estimate carbon emission statistics, expanding the application of deep learning technology in this field. This research serves as a reference for predicting carbon emission information and holds significance for relevant fields.

As per the authors of [3], who have analyzed the alarming pace of climate change, accurately estimating fuel consumption and emissions is vital to comprehend the consequences of materials and stringent emission regulations. The research conducted an analytical and predictive analysis using a dataset provided by the Government of Canada. The dataset included 4973 light-duty vehicles that were observed between 2017 and 2021.This study offers evidence-based suggestions, based on statistical data analysis, for producers and vehicle users to decrease their environmental impact.

The authors of [4] evaluated the Random Forest and SVM models to measure the amount of CO_2 emissions. The increase in CO_2 emissions is mainly attributed to energy consumption, such as the use of coal and electricity. The study aimed to monitor CO_2 emissions from the industrial use of coal and electricity. To train and test the model, data on electricity and energy were collected, with 60% used as training data and 40% as test data. The trial-and-error method was used to determine the model's specifications. The model with the smallest error measured by RMSE was considered the most accurate in estimating CO_2 emissions.

The author of [11] has demonstrated comparative techniques, including active contour model- and generic algorithm-based road segmentation from remote sensing LISS data. a method for extracting road damage that is based on object-oriented change detection with vector data. Based on these investigations, we can draw conclusions about the CO_2 emissions rating for several types of roads, including national, state, and city streets. According to the study of Recognizer the Image Caption Generator employing CNN and LSTM [12], which understands the demands of both natural language processing and image processing, it is possible to recognize the relationship of the image in English. By doing so, it will analyze the state of the roads and forecast the rate of CO_2 emissions from heavy cars.

V. DESIGN ARCHITECTURE.

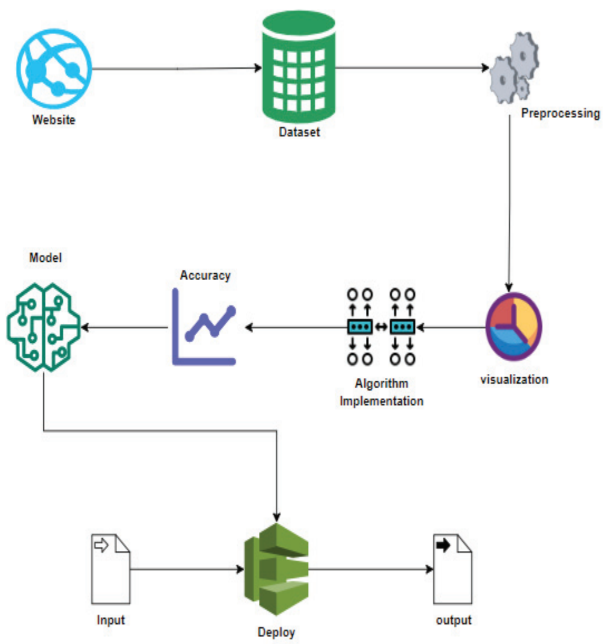


Fig 1. System Architecture

A. Entity Relationship Diagram (ERD)

An entity relationship diagram (ERD) or entity relationship model is a visual depiction of an information system that illustrates the connections between individuals, objects, places, ideas, or occurrences.

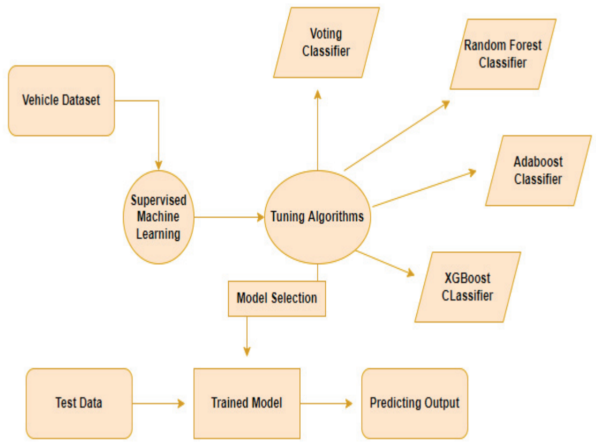


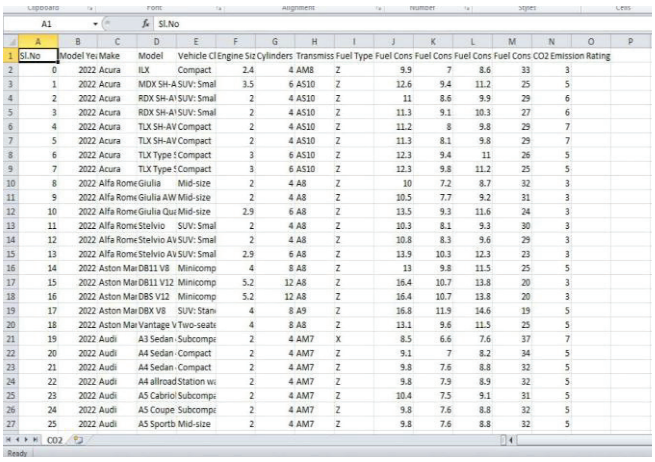
Fig 2. ERD diagram

VI. METHODOLOGY

Preparing the dataset: This dataset includes 950 records of features that were taken from vehicles and used to calculate the vehicles' pollution ratings.

A. Data Pre-processing

In machine learning (ML), validation procedures are utilized to estimate the error rate of a model, which is deemed as a reliable approximation of the dataset's actual error rate. When dealing with large datasets that accurately represent the population, validation approaches may not be necessary. However, it is crucial to thoroughly examine the data, including identifying duplicate or missing values and verifying data types, such as whether a variable is a float or an integer. This subset of data is used to evaluate the model's fit to a training dataset while adjusting hyperparameters.



Sl.No	Model Yea	Make	Model	Vehicle Cl	Engine Siz	Cylinders	Transmiss	Fuel Type	Fuel Cons	Fuel Cons	Fuel Cons	Fuel Cons	CO2 Emission Rating
0	2022	Acura	ILX	Compact	2.4	4	AM8	Z	9.9	7	8.6	33	3
1	2022	Acura	MDX SH-A	SUV: Smal	3.5	6	AS10	Z	12.6	9.4	11.2	25	5
2	2022	Acura	RDX SH-A	SUV: Smal	2	4	AS10	Z	11	8.6	9.9	29	6
3	2022	Acura	RDX SH-A	SUV: Smal	2	4	AS10	Z	11.3	9.1	10.3	27	6
4	2022	Acura	TLX SH-AV	Compact	2	4	AS10	Z	11.2	8	9.8	29	7
5	2022	Acura	TLX SH-AV	Compact	2	4	AS10	Z	11.3	8.1	9.8	29	7
6	2022	Acura	TLX Type S	Compact	3	6	AS10	Z	12.3	9.4	11	26	5
7	2022	Acura	TLX Type S	Compact	3	6	AS10	Z	12.3	9.8	11.2	25	5
8	2022	Alfa Rome	Giulia	Mid-size	2	4	A8	Z	10	7.2	8.7	32	3
9	2022	Alfa Rome	Giulia AW	Mid-size	2	4	A8	Z	10.5	7.7	9.2	31	3
10	2022	Alfa Rome	Giulia Qua	Mid-size	2.9	6	A8	Z	13.5	9.3	11.6	24	3
11	2022	Alfa Rome	Stelvio	SUV: Smal	2	4	A8	Z	10.3	8.1	9.3	30	3
12	2022	Alfa Rome	Stelvio AV	SUV: Smal	2	4	A8	Z	10.8	8.3	9.6	29	3
13	2022	Alfa Rome	Stelvio AV	SUV: Smal	2.9	6	A8	Z	13.9	10.3	12.3	23	3
14	2022	Aston Mar	DB11 V8	Minicomp	4	8	A8	Z	13	9.8	11.5	25	5
15	2022	Aston Mar	DB11 V12	Minicomp	5.2	12	A8	Z	16.4	10.7	13.8	20	3
16	2022	Aston Mar	DBS V12	Minicomp	5.2	12	A8	Z	16.4	10.7	13.8	20	3
17	2022	Aston Mar	DBX V8	SUV: Stan	4	8	A9	Z	16.8	11.9	14.6	19	5
18	2022	Aston Mar	Vantage V	Two-seate	4	8	A8	Z	13.1	9.6	11.5	25	5
19	2022	Audi	A3 Sedan	Subcompa	2	4	AM7	X	8.5	6.6	7.6	37	7
20	2022	Audi	A4 Sedan	Compact	2	4	AM7	Z	9.1	7	8.2	34	5
21	2022	Audi	A4 Sedan	Compact	2	4	AM7	Z	9.8	7.6	8.8	32	5
22	2022	Audi	A4 allroad	Station wa	2	4	AM7	Z	9.8	7.9	8.9	32	5
23	2022	Audi	A5 Cabriol	Subcompa	2	4	AM7	Z	10.4	7.5	9.1	31	5
24	2022	Audi	A5 Coupe	Subcompa	2	4	AM7	Z	9.8	7.6	8.8	32	5
25	2022	Audi	A5 Sportb	Mid-size	2	4	AM7	Z	9.8	7.6	8.8	32	5

Fig 3. Raw dataset

B. Data visualization

In the field of statistics of applied and machine learning the ability to visualize data is crucial. While statistics primarily focuses on numerical estimates and data descriptions, data visualization provides a vital set of tools for gaining a qualitative understanding of the data. By examining and exploring a dataset through visualizations, one can identify trends, outliers, and potentially corrupt data. Data visualizations also offer an effective way to communicate critical relationships through vivid and engaging charts and plots, making them more accessible to stakeholders with little subject knowledge. To delve deeper into the subjects of data visualization and exploratory data analysis, I recommend reviewing some of the recommended books on the topic.

C. Algorithm implementation

To ensure optimal performance in machine learning, it's important to systematically compare various algorithms. Scikit-learn is a popular machine learning library in Python that provides a wide range of tools and algorithms for building and evaluating machine learning models. One of the key strengths of scikit-learn is its ease of use and flexibility, which makes it a great tool for developing a test harness that can be applied to a wide range of machine learning problems. By incorporating a range of algorithms, you can conduct thorough comparisons and identify the most effective approach for your needs. Different performance traits will be present in each model. Cross-validation and other resampling methods can be employed to assess the performance of models on new, unfamiliar data and determine their accuracy. These estimations can then be used to identify one or two of the most effective models from the collection you have created. Just as visualizing a dataset through different perspectives can reveal valuable insights, selecting a machine learning model requires a similar approach. In order to select the most optimal one or two models for finalization, it is essential to assess the predicted accuracy of your machine learning techniques through diverse approaches. Utilizing a range of visualization methods that exhibit the mean accuracy, variability, and additional characteristics of the model accuracy distribution is a viable technique for evaluation. You will learn precisely how to achieve that in Python using scikit-learn in the following section. To ensure a fair and unbiased comparison of machine learning algorithms, it's crucial to evaluate each algorithm uniformly on the same data. This can be achieved by utilizing a consistent test harness for testing each algorithm.

1) Performance Metrics to calculate

False Positives (FP) - A situation where a person is expected to make a payment but instead pays late can be described as the actual class being false, despite it being anticipated as true. Similarly, if a predicted class indicates that a passenger would survive, but the actual class indicates that the passenger did not survive, it would also be considered a case the obtained data is false.

False Negatives (FN) - If a defaulter is predicted as a player, meaning that the forecasted class is negative but the actual class is positive, it is considered an instance of misclassification. For instance, if a passenger's survival is indicated in the actual class, but the predicted class suggests their death, it is an example of a positive class being projected as negative.

True Positives (TP) - When a defaulter is presumed to be a nonpayer, it can result in accurate positive predictions, where both the actual and predicted class values align. For instance, if a passenger's survival is predicted and also confirmed in the actual class, it confirms the correctness of the outcome.

True Negatives (TN) - If the predicted class is negative (e.g., a binary classification problem with two possible classes - positive and negative) and the actual class is also negative, it would be considered a true negative (TN) - a case where the model correctly predicted the negative outcome.

True Positive Rate (TPR) = TP / (FN + TP) (1)

False Positive rate (FPR) = FP / (TN + FP) (2)

Accuracy: (TP + TN) / (TN + TP + FN + FP) (3)

The accuracy is a performance measure that is straightforward to grasp, as it denotes the proportion of accurately predicted observations to the total number of observations. Although the model may be accurate, people may consider it the best. While a balanced false positive and false negative rate in the dataset is desirable, precision, which is the proportion of true positive predictions, is also a valuable indicator of performance.

Precision: TP / (TP + FP) (4)

Precision is a performance metric used in binary classification problems that measures the proportion of true positive (TP) predictions among all instances that the model predicted as positive, regardless of whether the actual class was positive or negative. In other words, precision measures how often the model correctly predicted the positive class out of all instances it classified as positive. How many passengers who were declared as having survived the accident actually did? is the question that this measure seeks to address. Low false positive rate correlates with high precision. 0.788 accuracy is a respectable number.

Recall: Precision is a measure of the proportion of true positive predictions among all instances that the model predicted as positive, regardless of the actual class. It does not directly refer to the percentage of actual defaulters that the model predicts correctly.

However, precision can be used to evaluate the performance of a model in identifying a specific class, such as defaulters in a binary classification problem. A high precision indicates that the model is making fewer false positive predictions, which means that the model is correctly identifying more true positives (i.e., actual defaulters) among all instances it classified as positive (i.e., predicted defaulters).

Recall = TP / (FN+TP) (5)

The F_1 Score is the weighted average of precision and recall, taking into account all false positives and false negatives. While accuracy is simpler to understand, F_1 is generally more useful, particularly in cases of imbalanced class distribution. When the expenses of false positives and false negatives are comparable, accuracy is the recommended metric. Conversely, if the costs of false positives and false negatives have notable discrepancies, Precision and Recall should both be taken into account.

General Formula:

F- Measure = 2TP / (FN + FP+2TP) (6)

F_1-Score Formula:

F_1 Score = 2*(Precision* Recall) / (Precision + Recall) (7)

2) AdaBoost Classifier

AdaBoost is considered to be flexible since it adjusts subsequent weak learners by giving more emphasis on examples that were previously misclassified by earlier classifiers. Compared to other learning algorithms, it may have a lower tendency to encounter overfitting in certain circumstances. AdaBoost (Adaptive Boosting) is an ensemble learning algorithm that combines multiple "weak" learners to create a more powerful "strong" learner. The idea behind AdaBoost is to iteratively train weak models on the same dataset and give more weight to instances that were misclassified in previous iterations, thereby forcing the subsequent models to focus more on the difficult instances. Even if each individual weak learner has only slightly better performance than random guessing, the iterative boosting process can help the final model converge to a powerful learner with high accuracy.

3) XG Boost Algorithm

Gradient boosting is a popular machine learning technique used in supervised learning problems for creating a predictive model. It is a form of ensemble learning that combines multiple weak models, such as decision trees, to create a strong model that can accurately predict a target variable. XGBoost (Extreme Gradient Boosting) is a widely used open-source software library that implements gradient boosting algorithms for classification and regression problems. It is designed to be scalable and efficient, making it a popular choice for large datasets with a high number of features. Regression trees act like weak learners when employing gradient enhanced for regression, and each one transfers an incoming data point to a leaf that carries a consistent score. The regularization of an objective function, consisting of a convex loss function and a model complexity penalty term (L1 and L2), is used in boosting to minimize regression tree functions. To forecast the residuals or errors of previous trees, additional trees are trained iteratively, and the total prediction is obtained by combining all of them. The name "gradient boosting" comes from the use of the gradient descent technique to reduce loss when adding new models.

4) Random Forest Algorithm

Random Forest (RF) is a commonly used ensemble learning technique applied in various tasks such as classification and regression. The RF algorithm generates multiple decision trees during the training phase, and the output is determined by the average of the categories (for classification) or the average prediction (for regression) of all the trees. Overfitting is a common issue with decision trees, but RF can adjust for it. In ensemble learning, several algorithms can be combined or the same technique can be used multiple times to create a more effective prediction model. Random Forest gets its name from combining multiple algorithms of the same type or different decision trees into a "forest" of trees. For both regression and classification tasks, the RF algorithm can be utilized by constructing decision trees and computing the output based on the average predictions of the trees.

- Randomly select a subset of N records (with replacement) from the original dataset. This is called a bootstrap sample.

- Build a decision tree based on the bootstrap sample. At each node of the tree, select a random subset of features to split on.
- Repeat steps 1 and 2 M times to create M decision trees.
- To make a prediction, pass the new instance through all M trees and take the majority vote of the predictions (for classification) or the average prediction (for regression).
- To prevent overfitting, use techniques such as pruning, limiting the depth of the trees, or setting a minimum number of instances required to split a node.

For a regression problem, every tree within a random forest estimates a value for the output variable Y with respect to a new record. The ultimate forecasted value is determined by averaging the predictions of all the trees in the forest In contrast, for a classification problem, each tree predicts the class to which the new data pertains. The class that receives the greatest number of predictions from the trees is selected as the ultimate prediction for the new record.

5) Voting Classifier

A machine learning algorithm known as a voting classifier is trained on a variety of models and forecasts an output class by selecting the class with the greatest probability of being the outcome. The voting classifier aggregates the outputs of all the classifiers that are allowed to participate and estimates the output class with the highest number of votes. The aim is to create a single model that can predict results for all output classes based on the combined majority of the training data used by the models, instead of building separate models for each output class and calculating their accuracy individually. The Voting Classifier supports two different methods of voting.

Hard Voting: Hard voting is a method of voting in which the output class is determined by the classification that receives the highest number of votes from the classifiers, or the classification that each classifier is most likely to predict accurately. For instance, if three classifiers predict the output class and the majority predicts class A (A, A, and B), then the final prediction would be A.

Soft Voting: Soft voting is a type of ensemble learning method used in machine learning classification problems. In soft voting, the output class prediction is made by taking into account the probability estimates for each class produced by each individual classifier in the ensemble. Specifically, each individual classifier produces a set of probabilities for each input instance, indicating the likelihood of that instance belonging to each possible class. The probabilities from all the classifiers are then aggregated, typically by taking the average or weighted average of the probabilities, to produce a final probability estimate for each class. The class with the highest probability is then chosen as the predicted class. Soft voting can be useful in situations where the individual classifiers in the ensemble produce complementary probability estimates, as it can help improve the overall accuracy of the final prediction.

D. Deployment

Implementing the model in the Django Framework and forecasting results. In this module, the machine learning model that has been trained is saved in a pickle data format file (.pkl file) and then utilized by our Django framework to enhance the user experience by predicting the CO_2 emissions based on the input data.

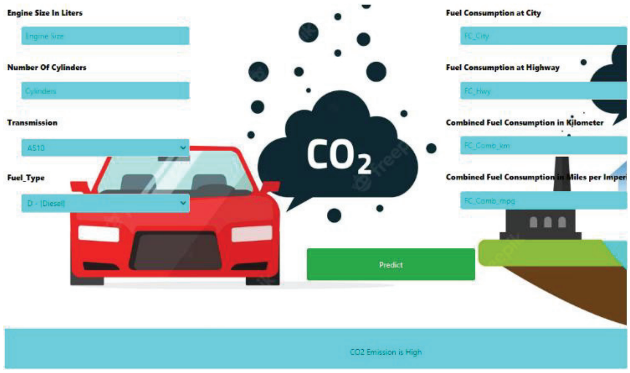


Fig 4. Deployment.

VII. CONCLUSION

The process began with data cleansing and preprocessing followed by analysis of missing values and exploratory analysis and construction and evaluation of model's algorithm with highest accuracy score is selected as the one that performs best on the public test set. This algorithm is then utilized in the program that helps to determine a vehicle CO_2 emission.

REFERENCES

[1] Sandeep Kumar Pranab K. Muhu - A Novel GDP Prediction Technique based on Transfer Learning using CO_2 Emission Dataset1 -2019

[2] Ngo Le Huy Hien and Ah-Lian Kor -Analysis and Prediction Model of Fuel Consumption and Carbon Dioxide Emissions of Light-Duty Vehicles-2021

[3] Huafang Huang 1,2,3, Xiaomao Wu - The Prediction of Carbon Emission Information in Yangtze River Economic Zone by Deep Learning-2021

[4] Prof. Swapnil Wani,2Mr. Akash Akhilesh Yadav - Predicting CO_2 Emission Using Machine Learning, -2022

[5] Chairul Saleh1, Nur Rachman Dzakiyullah2 Carbon dioxide emission prediction using support vector machine-2016

[6] Le Cornec, C.M.; Molden, N.; van Reeuwijk, M.; Stettler, M.E. Modelling of instantaneous emissions from diesel vehicles with a special focus on NOx: Insights from machine learning techniques. Sci. Total Environ. 2020

[7] Abukhalil, T.; AlMahafzah, H.; Alksasbeh, M.; Alqaralleh, B.A. Fuel consumption using OBD-II and support vector machine model. J. Robot. 2020

[8] Jaiprakash, Habib, G., Kumar, A., Sharma, A., & Haider, M. (2017). On-road emissions of CO, CO_2 and NO_X from four-wheeler and emission estimates for Delhi. Journal of Environmental Sciences, 53(X),39–47

[9] MK AN (2020) Role of energy use in the prediction of CO_2 emissions and economic growth in India: evidence from artificial neural networks (ANN). Environ Sci Pollute Res.

[10] Marland G, Boden TA, and Andres RJ (2011) CO_2 emissions from fossil fuels on a global, regional, and nationwide scale. Carbon Dioxide Information Analysis Center, US Department of Energy, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA doi.

[11] Loret, JB Shajilin, et al." recognize the Image Caption Generator Using CNN and LSTM with NumPy functions. "Solid State Technology" 64.2(2021): 2181-2191.

[12] Lakshay Amarpuri, Navdeep Yadav, Girish Kumar, and Saurabh Agrawal, "Prediction of CO_2 emissions using deep learning hybrid technique," IEEE, 2019.

[13] H. T. Pao, C.M. Tsai. Modeling and forecasting the CO_2 emissions, energy consumption, and economic growth in Brazil. Energy, 36 (2011) 2450–2458.

[14] L. Amarpuri, N. Yadav, G. Kumar, and S. Agrawal, "Prediction of CO_2 Emissions Using a Hybrid Deep Learning Approach: A Case Study in Indian Context," Twelfth International Conference on Contemporary Computing (IC3), 2019.