AI (ML DL) April15

Project

On

Prediction of Emission of CO_2 from Cars

Using

Machine learning Algorithms

By Team 23

1) Saran Kumar Reddy Palle

2) Naga Sai Krishna Merugumilli

3) Naganaboina Pavan Kumar

4) Meda Navya Sri

5) Siddula Spandana

Table of Contents

1 **INTRODUCTION**

1.1 Overview

1.2 Purpose

2 **LITERATURE SURVEY**

2.1 Existing problem

2.2 Proposed solution

3 **THEORITICAL ANALYSIS**

3.1 Block diagram

3.2 Hardware / Software designing

4 **EXPERIMENTAL INVESTIGATIONS**

5 **FLOWCHART**

6 **RESULT**

7 **ADVANTAGES & DISADVANTAGES**

8 **APPLICATIONS**

9 **CONCLUSION**

10 **FUTURE SCOPE**

11 **BIBILOGRAPHY**

APPENDIX

A. Source code

1) Introduction

1.1Overview:

Our personal vehicles are a major cause of global warming. Collectively, cars and trucks account for nearly one-fifth of all emissions, emitting around 24 pounds of carbon dioxide and other global-warming gases for every gallon of gas. About five pounds comes from the extraction, production, and delivery of the fuel, while the great bulk of heat-trapping emissions—more than 19 pounds per gallon—comes right out of a car's tailpipe.

A typical passenger vehicle emits about 4.6 metric tons of carbon dioxide per year. This number can vary based on a vehicle's fuel, fuel economy, and the number of miles driven per year. The higher the number of the controlled and uncontrolled effect variables that influence the co_2properties, the lesser the predicted accuracy. Despite this, a few experimental designs have been suggested by considering the controllable effect variables and interaction terms between them [1].

1.2 Purpose:

The effects of car pollution are widespread, affecting air, soil and water quality. Carbon monoxide, another exhaust gas, is particularly dangerous to infants and people suffering from heart disease because it interferes with the blood's ability to transport oxygen. Other car pollutants that harm human health include sulphur dioxide, benzene and formaldehyde. Noise from cars is also harmful, damaging hearing and causing psychological ill-health.

In recent years, the ML methods have become popular as they allow researchers to improve the prediction accuracy of threshold limits about the fuel combustions and CO_2 emissions. The ML methods have been used to increase the prediction accuracy of fuel emissions and the data derived from the literature sources.

Regression models tend to be used for the prediction of emission of CO_2 from cars. These models also demonstrate how the concrete compressive strength depends on the mixing ratios.

Previous studies evaluated the amount of the concrete component materials and compared their results to the published data. In this study, the ML regression methods were compared to predict the compressive strength and slump values of the cube samples. The samples were prepared by accounting for seven simultaneously controllable effect variables in the laboratory. The study aimed to determine the most successful regression method by comparing the decision tree (DT), random forest (RF), support vector machine (SVM).

2) Literature Survey

2.1 Existing Problem

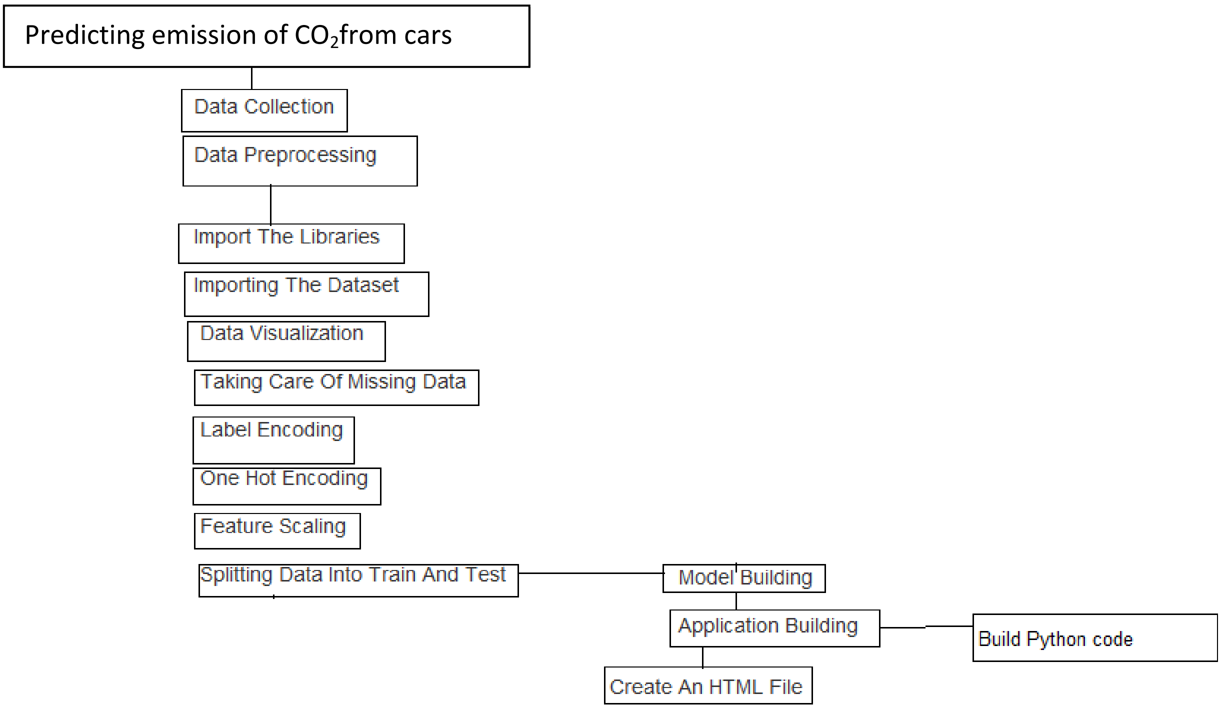
Carbon dioxide and other gases trap heat in the atmosphere, causing global temperatures to rise. Although carbon dioxide has natural sources, such as volcanic eruptions, human activities have caused an increase from 280 parts per million before the development of industry to 370 parts per million today. This consumes a lot of time and requires a lot of labour to prepare different prototypes and test them. Also, this method is prone to human error and one small mistake can cause the wait time to drastically increase.

2.2 Proposed Solution

Fuel combustion emissions can be calculated using the emissions factor of 2.33 kg CO_2e/litre. If your car average 8 L/100 km then you multiply this by 2.33 and divide by 100 to give 186 g CO_2e/km for combustion emissions. Fuel production emissions can be calculated using an emissions factor of 0.43 kg CO_2e/litre. Here in this project we are going to use random forest regression. By using this algorithm depends upon our input data the output of the vehicle are going to predict the co2 emission of that particular car. So that the emission of CO_2 of that particular car is more than the threshold value then that car details should be sent to the particular RTA region head to seize the car.

3) Theoretical Analysis

3.1 Block Diagram



3.2 Hardware / Software Designing Techniques:

The following are the frameworks and packages used to develop this model :

- Python
- Python Web Frame Works
- Python for Data Analysis
- Python For Data Visualization
- Data Pre-processing Techniques
- Machine Learning
- Regression Algorithms.

4) Experimental Investigation :

Reducing transportation emissions is one of the most vital steps in fighting the climate emergency, and solutions to the transportation problem are already available. Our nation needs to shift away from fossil fuel-powered vehicle dependence and toward zero emissions in all transport sectors.

The casting and testing of specimens for generating the data bank were performed in controlled approvals of the government.

Range of various parameters

Number of cylinders in a car = 4 to 12

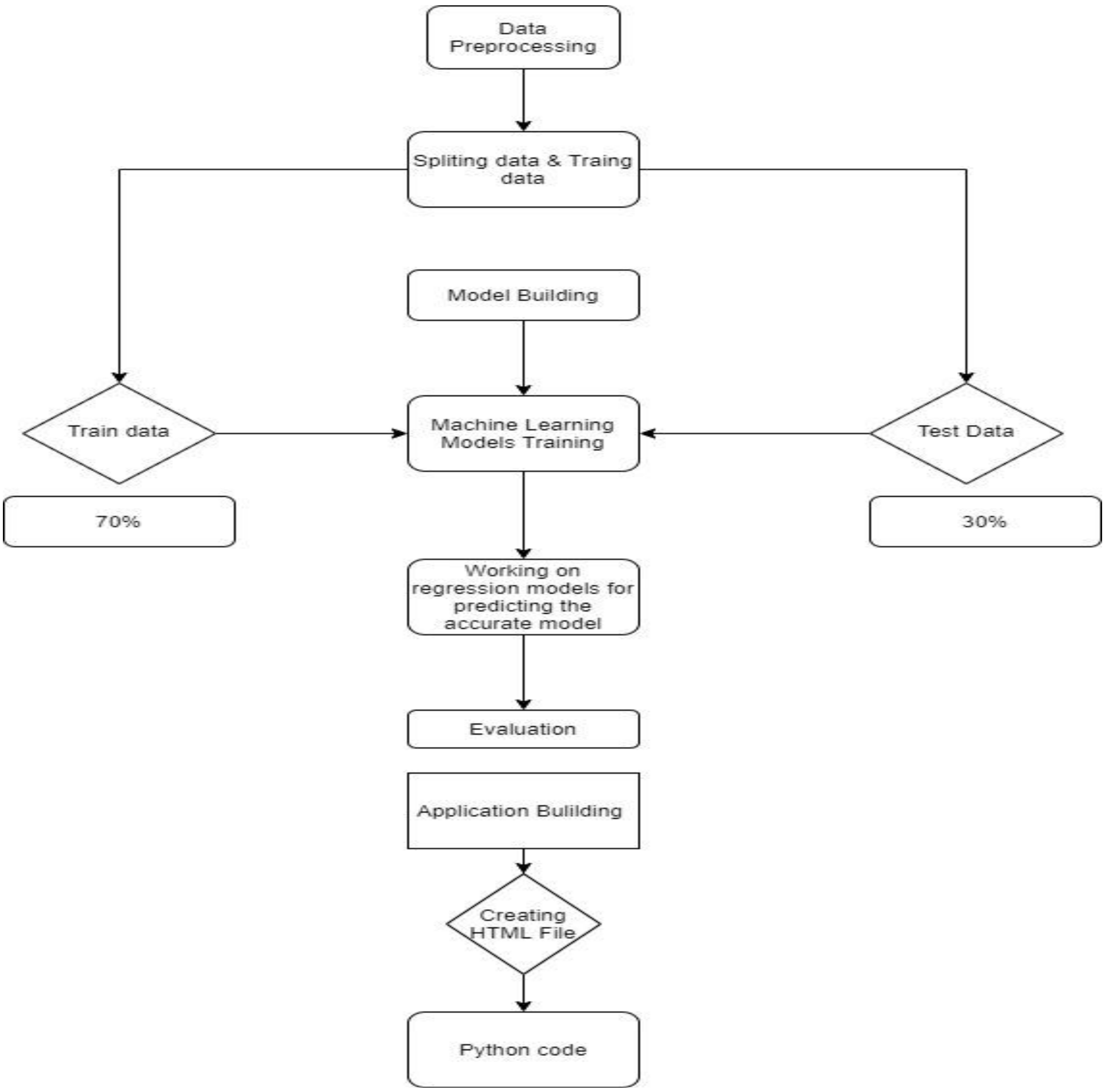
Engine size (in liters) = 2 to 7

Fuel combustion(gallons per mile) = 6 to 20

Note: Fuel combustion varies as per the path such as on highways, cities (urban) and sometimes combined paths.

These all parameters have impact on emission of CO_2 from cars.

5) Flowchart



6) Result

We have analysed the data of co_2 emissions of various cars and used Machine Learning to Predict the co_2 emissions. We have used Linear Regression and its variations, Lasso, Ridge and Random Forests to make predictions and compared their performance. Random Forest Regressor has highest accuracy and is a good choice for this problem. Random Forest Regressor trains randomly initialized trees with random subsets of data sampled from the training data, this will make our model more robust

7) Advantages and Disadvantages

Advantages:

Using Machine learning to predict the CO_2 emissions from cars will be time effective and more accurate. It is more trust worthy and cost effective .It also reduces the man power for doing the experiments to find the strength of the concrete in different unknown situations.

Disadvantages:

There is a 4 % chance that the outcome will not predict the approximate value and in that situation it can be troublesome.

8) Applications:

- Can predict the emission of CO_2 using the inputs provided.
- Implementable on the website

9) Conclusion

In this study, a prediction model of CO_2 emission is predicted using Random Forest Regressor. A total of 1060 sample data collected from the experimental test were used to develop the Random Forest Regression model for predicting compressive strength. The Random Forest model was first calibrated and then verified using the experimental data from samples. Conclusions can be drawn as follows:

- Compare to all other Machine Learning Models Random Forest was best suitable for this data.
- Random Forest Regressor gave the maximum accuracy when tested using r2_score and mean errors.
- Maximum accuracy received is 96.43 %.

10) Future Scope

This model can predict the outcome with many different inputs within seconds. The model will save time for the RTA to analyse a vehicle's CO_2 emission. Experiment cost is also reduced for car owners, who could find pollution being caused by their own cars.

11) Bibliography

Books

Hastie, Friedman, and Tibshirani, The Elements of Statistical Learning, 2001
Bishop, Pattern Recognition and Machine Learning, 2006
Ripley, Pattern Recognition and Neural Networks, 1996
Duda, Hart, and Stork, Pattern Classification, 2nd Ed., 2002
Tan, Steinbach, and Kumar, Introduction to Data Mining, Addison-Wesley, 2005.

Data repositories

Self prepared data using GreenVehicleGuide.com

Algorithms

Thesmartbridgeteachable.com

12) Appendix

Source Code :

Importing the libraries

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import pickle
```

Reading data

```
dataset = pd.read_csv('Emission_Co2.csv')
```

Taking Care Of Missing Values

```
dataset.isnull().any()
```

Removing the unnecessary Parameters

```
data1=dataset[['CYLINDERS','ENGINESIZE','FUELCONSUMPTION_CITY','FUELCO
NSUMPTION_HWY','FUELCONSUMPTION_COMB','CO2EMISSIONS']]

x = data1.iloc[:, :6]
y = dataset.iloc[:, -1::]
```

Splitting Training and Test Set

```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=10)
```

we will train our model with all available data

```
from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators=20, random_state=10,max_depth =10)
```

Fitting model with training data

```
regressor.fit(x_train, y_train)
y_predict = regressor.predict(x_test)
```

Evalution

```
from sklearn.metrics import r2_score
accuracy = r2_score(y_test,y_pred1)
accuracy
```

Saving model to disk

```
pickle.dump(regressor, open('model.pkl','wb'))
```

Loading model to compare the results

```
model = pickle.load(open('model.pkl','rb'))
print(model.predict([[4,2.4,11.2,7.7,9.6,29]]))
```