# PREDICTION OF VEHICLE EMISSION (CARBON MONOOXIDE) USING MACHINE LEARNING ALGORITHM

Mahi Agrawal
Student
M.E, VLSI & Embedded
G.H.R.I.E.T, Pune, India

***Abstract :*** The pollutants a vehicle emits, during the combustion process are known as vehicle emissions. Carbon monoxide (CO) is emitted due to incomplete combustion when there is not enough oxygen to burn fuel. This gas has no smell, color or flavor, so it is difficult to recognize. It causes fatal air poisoning and is primarily responsible for seizures, coma and deaths in humans. In this paper we will predict the amount of Carbon monoxide (CO) emitted from the vehicle using Machine Learning algorithms. Machine learning is the field, in which computers can learn without being explicitly programmed. In this project a training dataset is given as a input to the machine and learning algorithm produces an inferred function to make predictions about the output values ie the amount of CO for each hill of drive cycle of given vehicle.

***IndexTerms*** **-Algorithm; Emission; Hydrocarbons; Machine learning; Reinforcement**

## I. INTRODUCTION

The population is increasing rapidly, which leads to an increase in the number of vehicles without the length of the road increasing accordingly. The number of vehicles registered in Delhi has already exceeded 6 million, and significant vehicle traffic connects the Delhi road from neighboring countries. It was found that vehicles drive less than 20 km / h in approximately 30% of all cases, regardless of road class. It has been observed that only vehicles cause around 64% of pollution in Delhi, while other sources such as power plants, industry and households contribute 16%, 12% and 8%, respectively, to environmental pollution . The maximum contribution of air pollution is growing rapidly from the sources of vehicles

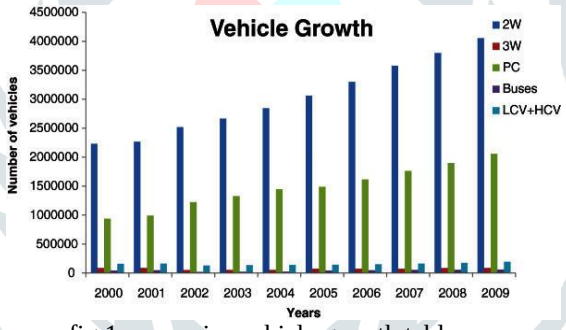



fig 1. year wise vehicle growth table

The pollutants a vehicle emits during the combustion process are known as vehicle emissions. The main four combustion components are: carbon monoxide (CO), carbon dioxide (CO2) and nitrogen oxide (NO2). Carbon monoxide (CO) is due to incomplete combustion, if there is no enough oxygen to burn fuel. This gas has no smell, color or flavor. It is difficult to recognize Causes fatal air intoxication and is primarily responsible for seizures, coma and deaths in humans.

The pollution of cars is one of the main causes of global warming. Cars and trucks emit carbon dioxide and other greenhouse gases, which make up a fifth of the pollution of total global warming in the United States. Greenhouse gases store heat in the atmosphere, increasing global temperatures. The effects of vehicle pollution are widespread and affect the quality of air, soil and water.

The hydrocarbons, carbon monoxide and other pollutants in the car are harmful to human health. Diesel engines emit a large amount of fine dust, which is a soot and metallic particle of air. This causes irritation and allergies to the skin and eyes, and very fine particles are deposited in the lungs and cause respiratory problems there.

Machine learning is an application for artificial intelligence that allows the system to automatically learn from experience and improve without explicit programming. Machine learning aims to develop computer programs that can access and learn the data itself.

In this project we will predict the amount of Carbon monoxide (CO) emitted from the vehicle using Machine Learning algorithms. Machine learning is a field of computer science that gives computers the ability to learn without being programmed explicitly. So, this project will automatically predict the value of the emission gases of a vehicle just with the input as a drive cycle

## II. ENVIRONMENTAL PROTECTION AGENCY (EPA)

The Environmental Protection Agency (EPA) was established in December 1970 under the direction of US President Richard Nixon. Its mission is to protect human health and the environment. DC-based EPA is accountable for developing standards and laws that promote human health and the environment. The purpose of the EPA is to protect and protect the natural environment and improve people's health by groping the impact of contaminants and setting limits on their use. EPA smears its findings through fines, fines and other records.

## III. DRIVE CYCLE

A driving cycle is a series of data points that show the vehicle speed over time. Different countries and organizations are creating driving cycles to assess vehicle performance in different ways, such as fuel consumption and pollutant emissions. Fuel consumption and emissions tests are performed on the chassis dynamometer.. The exhaust gases are composed and measured to determine the performance of the vehicle.

There are two types of driving cycles. Modal cycle as European standard or mode 10-15 in Japan, and transition cycle as FTP-75 cycle or Artemis. The main variance is that the modal cycle is a combination of direct acceleration and a constant speed period, which does not reflect the driver's actual comportment, whereas the transition cycle has many speed changes typical of road traffic conditions. It is to include In this article we have used the EPA 75 driving cycle.

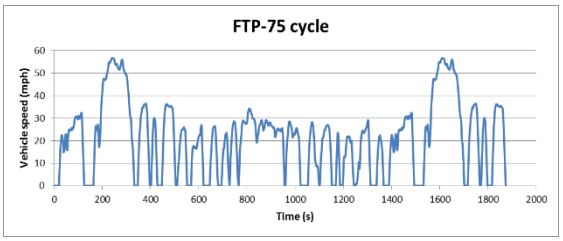
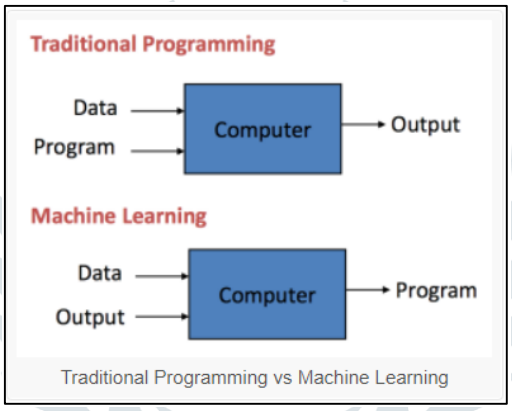



Fig 2. FTP-75 Cycle

## IV. MACHINE LEARNING

A machine learning algorithm is an algorithm that can learn from the information. The power of machine learning is that you can determine how to differentiate using models, rather than using human judgment. It is said that an algorithm / computer program learns from the measure of performance P and experience E with a class of tasks T as its performance improves in tasks in T, measured by P, with experience E.

- **Traditional programming:** Data and programs run on the computer to produce the output.
- **Machine Learning:** Data and results are run on the computer to create a program. This program can be used in conventional programming




## V. TYPES OF MACHINE LEARNING

1. Supervised learning: Supervised learning as a name indicates the presence of a administrator as a instructor. Fundamentally, supervised learning is the learning in which we teach or train the machine using well-labeled data, that is, some data has already been provided with the correct response. Thereafter, the machine receives a new set of examples (data) so that the supervised learning algorithm analyzes the training data (set of training examples) and generates a correct result of the tagged data.
2. Unsupervised learning: Unsupervised teaching is about training a machine with information that is not classified or labeled. This allows the algorithm to respond to this information without guidance. The mission of the machine here is to group the information not classified by similarity, pattern and difference, without previous exercise data. Unlike supervised learning, no teacher is provided. In other words, the machine is not trained. As a result, the machine is limited to finding the hidden structure of unlabeled data.
3. Reinforcement Learning - Reinforcement Learning is a type of Machine Learning in which the machine is required to determine the ideal behavior within a specific context, in order to maximize its rewards. It works on the rewards and punishment principle which means that for any decision which a machine takes, it will be either be rewarded or punished due to which it will understand whether or not the decision was correct. This is how the machine will learn to take the correct decisions to maximize the reward in the long run

As here we have the input dataset of different vehicle and on the basis of database machine will predict the outcome, so we will use the Supervised Machine Learning Algorithms.

## VI. PROCEDURE

The basic stages that lead to machine learning and it's fundamentals, are described below image [1]

1. Data gathering
2. Data Preparation
3. Feature Selection
4. Model Creation
5. Evaluation
6. Validation.

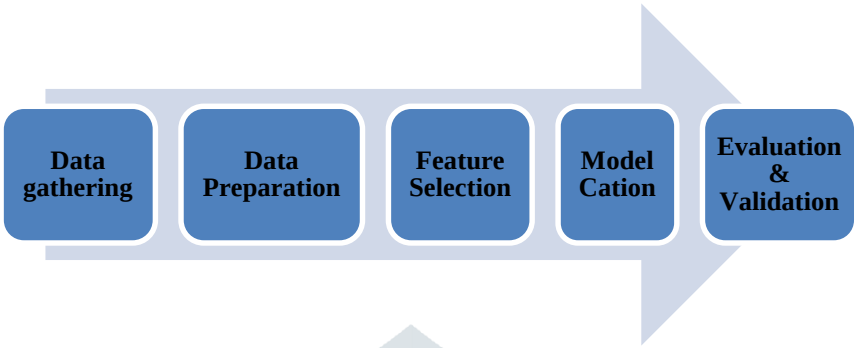



fig 3. phases in machine learning

1. Data Gathering: In this phase we need to get thxe maximum data which may be related to the target data. The number features (columns) can be very high in actual datasets. Mostly more is the data, the better it is. In our case the data is a drive cycles of the vehicle. This step is very imperative because the class and amount of data composed directly determines how upright the predictive model will be. The data composed is then charted and called as Training Data.

2. Data Preparation: Once you have collected training data, proceed to the next level of machine learning. Prepare the data Here, the data is loaded in the right place and ready for use in machine learning. Here the data is combined first, then the random order is combined. This is because the order of the data does not affect what is being investigated
This is also a good enough time to do any visualizations of the data, as that will help you see if there are any relevant associations between the diverse variables, how you can take their pro's and as well as show you if there are any data disparities present. Also, the data now has to be split into two parts. The first part that is used in exercise our model, will be the majority of the dataset and the second will be used for the evaluation of the skilled model's presentation. The other forms of adjusting and handling like normalization, error correction, and more take place at this step.[1]

3. Feature Selection: Not all features (columns) maybe helpful in predicting data. Instead of helping they may reduce the accuracy of final predicted target value. To prevent this using different techniques useful feature (columns) is taken for final model creation

4. Model Creation: Depending on the type of data (for example, classification / regression *), different algorithms are used to create models that predict the data. Most machine learning is known as training, where the data is used to progressively improve the predictability of the model. The training process involves initializing some random values for, for example, A and B of our model, predicting the result with those values, comparing it with the prediction of the model and adjusting the values to match the predictions, repetitions and update cycles previously referred to as a training step.

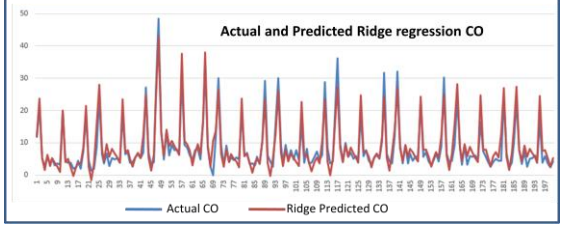
5. Evaluation: Once the exercise is complete, check this step to see if it is good enough. This is where the data set comes in, which you have held in reserve before. The evaluation allows the model to be tested using data that has never been seen and used for exercise, and is intended to be demonstrative of the performance of the model in practice.

6. Validation: The model is then validated with test data and using different analysis the model's strength is checked and if it performs up to the mark. First we need to divide our data set into train and test data. This is to check if our model **is** works properly on different dataset. Mostly the training set is taken as 70% of data whereas testing set is taken as 30% of data. If the difference between the testing and training data is not more then we can consider the model for prediction.

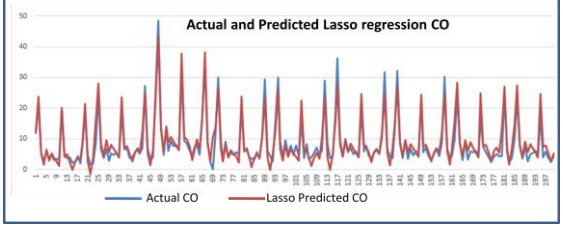
## VII. RESULTS

The predicted values of CO are compared with the actual values and results are shown in the following graphs for different algorithms:

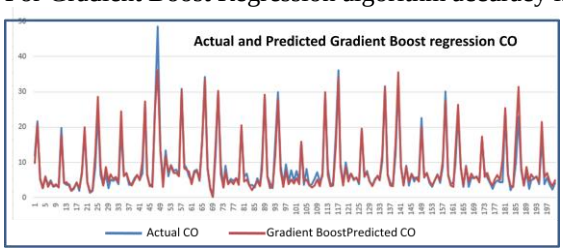
1. For Ridge Regression algorithm accuracy is 82%



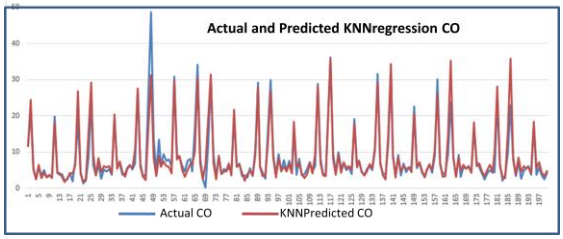

2. For Lasso Regression algorithm accuracy is 82%



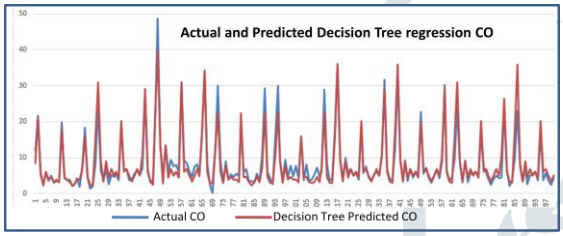
3. For Gradient Boost Regression algorithm accuracy is 88%



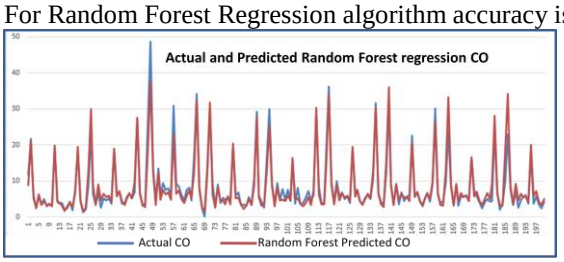

4. For KNN (K Nearest Neighbour) Regression algorithm accuracy is 87%



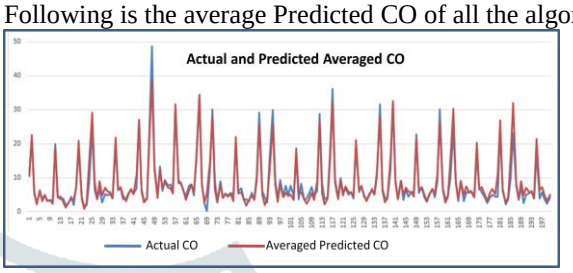

5. For Decision Tree Regression algorithm accuracy is 89%




6. For Random Forest Regression algorithm accuracy is 90%




Out of all the algorithms, Random Forest has the highest accuracy in this case.
Following is the average Predicted CO of all the algorithms.




## VIII. Conclusion

Vehicular emissions (e.g., CO) are one of the major sources of environmental pollution in urban areas, where road networks, intersections, and toll plaza areas are present. Vehicular emission prediction models are used to assess the impacts of vehicular emissions from different types of vehicles on human health and the environment. In this study different machine learning algorithms are applied on the vehicle drive cycle and predicted the amount of output Carbon Mono0xide. Also the average is calculated for the different algorithms. The predicted values of CO are compared with the actual values and results are shown.

## IX. References

1) https://www.techleer.com/articles/379-the-seven-steps-of-machine-learning/
2) https://machinelearningmastery.com/classification-and-regression-trees-for-machine-learning/
3) https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d
4) https://medium.com/@adi.bronshtein/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7
5) http://www.saedsayad.com/k_nearest_neighbors_reg.htm
6) http://dataaspirant.com/2017/05/22/random-forest-algorithm-machine-learing/
7) https://www.analyticsvidhya.com/blog/2016/01/complete-tutorial-ridge-lasso-regression-python/
8) http://www.statisticshowto.com/ridge-regression/