

# Prediction of Engine Emissions using Linear Regression Algorithm in Machine Learning

Kongara Venkatesh, Sivanesan Murugesan



**Abstract:** A large set of data are being generated in engine testing which is used for the evaluation of performance and prediction of emission characteristics. For any engine modifications or required improvements in the results, the whole testing procedure to be repeated again for further evaluation. To overcome this repetition, we need some data handling and analysis techniques such as machine learning and prediction models. The datasets which were collected by testing procedures help in building a prediction model by which the expected results of the test can be predicted without conducting repeated trials. This study mainly focusses on predicting the emissions of a diesel engine using a prediction model built by Linear Regression Algorithm in Machine Learning using Regression Learner Application in MATLAB. Linear Regression prediction model was built from the emission data collected from the single-cylinder diesel engine testing. The prediction model is validated and compared with the actual testing data obtained. Errors such as RMSE, MSE, MAE, R-squared errors are evaluated and found to be minimum. Using a validated prediction model, the emissions values can be predicted for any range of data set. This will reduce the time and cost involved by the repetition of testing procedures.

**Keywords:** Data Pre-processing, Engine Emissions, Linear Regression, Machine Learning.

## I. INTRODUCTION

Emissions are the pollutants which are emitting out of the vehicles whenever any technical inabilities occurred. CO, HC and NO<sub>x</sub> are the principle engine emissions from the SI and CO, HC, NO<sub>x</sub> and PM from the CI engines.

### A. Causes of emissions

The constituents of NO<sub>x</sub> are NO and NO<sub>2</sub>. When the fuel burns in the engine it reaches to high temperatures. Nitrogen and oxygen available in the air will combine at this high temperatures and forms nitrogen oxide. When coming out of the tail pipe it will mix with oxygen and forms nitrogen dioxide. When fuel is injected into the combustion chamber, all the fuel may not be burned. Some of the regions where the combustion flame can't reach are filled with fuel very rich mixtures.

These pockets of unburned fuel will cause (HC) hydrocarbon emissions. Fuel contains carbon compounds which need to be oxidized when burned with sufficient air. After burning Carbon dioxide will forms. But when there is no enough oxygen and due to incomplete combustion of carbon containing compounds, Carbon monoxide (CO) is released. Unburned fuel and lubricating oil will be having a large number of heavy hydrocarbons. These particles will condense or adsorb on to the carbon particles and these particles will lead to the formation of particulate matter (PM). This is reason for formation of black smoke in the exhaust of the vehicle.

### B. Parameters Affecting Emissions in Vehicles

The parameters which are going to affect the emission behavior in the engine are classified as design parameters and operating variables. Design parameters like compression ratio, surface to volume ratio, ignition timing, motion of air into the chamber, charge stratification, etc., and some of the parameters like air fuel ratio, charge dilution and exhaust gas recirculation, speed, load, coolant temperature and engine operating conditions like acceleration, deceleration and cruising are classified as operating variables will affect the emission behavior in engines.

The process of using the patterns hidden in the data sets collected, in order to predict the forthcoming or any behavior of the parameters which is crucial in any operation is called as Prediction analysis. For this process to be happening, the data sets play a key role in the whole prediction analysis. The raw data sets which are collected during any testing procedure, is going to use so that the patterns hidden in them will uncover and helps to predict the forthcoming. This whole pattern unveiling process is called as training. This prediction analysis is possible through the Machine Learning and its algorithms.

### C. Application of Machine Learning in Predictive Analysis

In the past two decades the Machine Learning became crucial in data analysis and prediction. In daily life it plays an important role in organizing the things needed such as email classification, internet browsing, weather forecasting, facial recognition and speech recognition, etc.,

Machine Learning helps in developing some intelligent systems which are going to work in autonomous way. These intelligent systems will work with the help of algorithms such as Linear Regression, Support Vector Regression and much more which will be selected appropriately according to the application and data types. These algorithms will learn with the help of previously obtained or historical data. This historical data will be gone through statistical analysis and pattern recognition.

Revised Manuscript Received on May 30, 2020.

\* Correspondence Author

Kongara Venkatesh\*, Mechanical Engineering department, Amrita School of Engineering, Coimbatore, Amrita Viswa Vidyapeetham, India. Email: kmani733@gmail.com

Sivanesan Murugesan, Mechanical Engineering department, Amrita School of Engineering, Coimbatore, Amrita Viswa Vidyapeetham, India. Email: m\_sivanesan@cb.amrita.edu

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

The completion of analysis will lead to the predicted results. The linear regression prediction model takes the predictor variables such as temperature and load on engine and checks for the relativity between the inputs and the outputs in the data. It will learn the pattern containing in the data imported and used in predicting the response variables such as NO<sub>x</sub> and smoke emissions of the engine.

The algorithm will use the previous learning experience whenever a prediction is required from a new data set. The optimization is required based upon the predictions or results in order to increase the accuracy of the prediction model.

## II. LITERATURE SURVEY

**Ayon Dey** [1] mentioned about various machine learning algorithms and their applications such as Decision Trees, Naïve Bayes, Support Vector Machines, K-Means, K-Nearest Neighbor, etc., and their applications in data mining, image processing, voice recognition, forecasting and prediction of events. The classification of Machine Learning and their sub-classification with an overview was presented. **Jiang Zheng, Aldo Dagnino** [2] mentioned about the present data analytics situations and their limitations in industrial growth. The forecasting of substations fault systems and power load with the help of R and Rapid miner tools and a variety of the machine learning algorithms like Neural Networks, Support Vector Machines, Naïve Bayes were used as prediction models. Linear regression with stochastic gradient descent algorithm was used as a prediction model in predicting the power loads. The subsystem faults were predicted with the 0.5994 accuracy and f-measure of 0.5963. **Gulden Kaya Uyanik<sup>i\*</sup>, Nese Guler<sup>ii</sup>** [3] mainly focused on the Linear Regression analysis and its type, Multi-Variate Linear Regression. Prediction analysis were carried out on the Education Faculty student's lessons score and their score using Multi-Variate Linear Regression technique. It was found that the score prediction from the input variables was accurate with the model's degree of prediction,  $R=0.932$  and the model's degree of explaining the variance in the response variable is  $R^2=0.87$ . **Yashavant S. Ingle<sup>1</sup>, Prof. Anil Mokhade<sup>2</sup>** [4] focused on using Linear Regression and its importance in prediction analysis. A prediction model was built using Linear Regression in order to find the ranking for the students based on the CGPA as an independent variable. It was mentioned that the other learning to rank methods will bring good results in prediction analysis and also exploring the other methods will brings out the predicted results nearer to target values. **Nelson Fumo, M.A.Rafe Biswas** [5] focused on a comparative study on available machine learning algorithms such as Linear Regression, K-means, Naïve Bayes, Support Vector Machines for the predictive analysis process. A model was built to predict the residential energy consumption in terms of outdoor temperature, solar radiation, hourly consumption. It was shown that the Linear Regression algorithm among the all mentioned machine learning algorithms showed better accuracy of  $R^2=0.89$ , in other words, the model's degree of explaining the variance is 89% in predicting the residential energy consumption. **Aulia Qisthi Mairizal<sup>a,b</sup>, Sary Awad<sup>a,\*</sup>, Cindy Rianti Priadi<sup>b,c</sup>, Djoko M. Hartono<sup>b</sup>, Setyo S. Moersidik<sup>b</sup>, Mohand Tazerout<sup>a</sup>, Yves Andres<sup>a</sup>** [6] focused on predicting the properties of Bio-Diesel such as a viscosity, density, etc., with the help of a prediction model built by using Multivariate

Linear Regression model. It was mentioned that these properties will affect the choice of feedstock for the standards of Bio-diesel to reach. It was found that the prediction model built showed the high accuracy for parameters like density and higher heating value with prediction errors  $< 5\%$ ,  $< 10\%$  for viscosity and  $< 15\%$  for flash point and oxidative stability. **Ki-Young Lee<sup>1</sup>, Kyu-Ho Kim<sup>1,\*</sup>, Jeong-Jin Kang<sup>2</sup>, Sung-Jai Choi<sup>3</sup>, Yong-Soon Im<sup>4</sup>, Young-Dae Lee<sup>5</sup>, Yun-Sik Lim<sup>6</sup>** [7] focused on comparison between Linear Regression and Artificial Neural Networks with the help of training the models with survey data about status of school firms in terms of type of establishment, class of school, etc., It was found after the training and analysis of survey data, Linear Regression was found to show the R-Square value to be 0.6111 i.e., variance is about 60%. This explains that the regression analysis of data sets is the best way in predicting the responses as the data sets deals with the numbers. It was found that Linear Regression is best to use and simple of all the regression It will best fit a line with the minimum errors and high efficiency. It is the least complex algorithm of all which is easier to modify to get better outputs. However, it is having some disadvantages like overfitting which can be avoided with preprocessing of the data.

## III. PROPOSED METHODOLOGY

### A. Experiment

In this study, prediction of emissions from a single cylinder diesel engine was performed. The test was performed on GL-400 diesel engine which is coupled with 50 KW Eddy current dynamometer. Five-gas analyzer is used to measure and collect the emission levels and a computer setup is connected to monitor the performance parameters such as load on the engine, temperature and emissions such as smoke and NO<sub>x</sub>.

### B. Engine Specification

The Greaves GL-400 is a single cylinder, diesel fueled engine. The technical specifications of the diesel engine used in the experiment are shown in below Table I.

**Table- I: Technical Specifications of the Engine**

S. No	Parameter	Specification
1	Make	Greaves Cotton Limited
2	Engine Configuration	Single Cylinder, Four Stroke
3	Type of Injection	Direct Injection
4	Type of Cooling	Air Cooled
5	Compression Ratio	18:1
6	Rated speed	3600 RPM
7	Maximum torque	20 Nm
8	Maximum power	8.5 Hp

### C. Dynamometer Specifications

The dynamometer used in performing the test on engine is Eddy Current Dynamometer and its technical specifications are given in the below Table II.

**Table- II: Technical Specifications of Dynamometer**

S. No	Parameter	Specification
1	Type	Eddy Current Dynamometer
2	Capacity	50 KW

#### D. Testing and Acquiring Data

The GL400 diesel engine was coupled with the Eddy current dynamometer to maintain the loads for the testing. The five-gas analyzer and smoke meter was placed at the exhaust manifold of the engine to measure the  $\text{NO}_x$  emissions and smoke respectively. A temperature sensor was mounted onto the engine which measures the engine temperature. The whole setup was integrated with the data acquisition system and a monitor to collect and store the data.

The tests were performed on Engine in different conditions of load, i.e., 0 Nm and 7 Nm with the speed maintained as constant. At each load  $\text{NO}_x$  (ppm), smoke and the temperature ( $^{\circ}\text{C}$ ) values were taken from the data acquisition system. Load should be slowly increased in order to avoid any failure of the setup. The experimental setup for engine emission testing is shown below in Fig. 1.

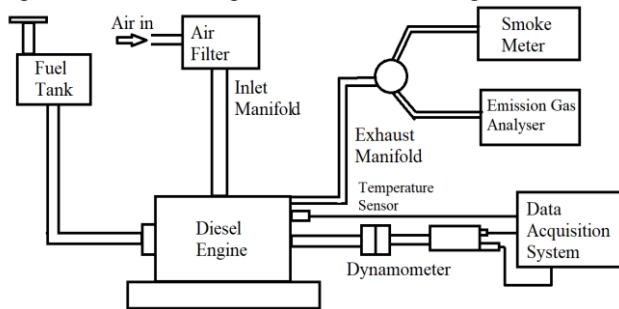


Fig. 1. Experimental Setup for Engine Emission Testing

### IV. DATA ANALYSIS AND TRAINING

#### A. Data Pre-processing Techniques

Data preprocessing or Data Mining is the process of extracting the useful information from raw data. It will eliminate the data rich information poor situation and will uncover the pattern between the datasets and gives us some useful information and knowledge. These preprocessing techniques are of two ways which are used in order to eliminate the errors and achieve better prediction results. The common technique in handling the missing data values is by either deleting the entire row or sometimes by the entire column only if 75% of column values are found to be missing or null. It is important to make sure that after deleting there should be no bias added. The other method is by replacing all the missing with median, mean or most frequent values. This method will give better results compared to deleting rows and columns.

#### B. Linear Regression Algorithm

Linear Regression is classified under supervised learning followed by the regression type and is the most used and less complex algorithm. It is a method of modelling a target value based on independent variables by fitting a straight line in the data with minimum errors. This algorithm consists of independent variables or predictor variables and response variables. This algorithm will fit a straight line into the data for the process of predictions using both the variables. The straight-line equation is shown in the below equation (1).

$$y = mx + c \quad (1)$$

As the algorithms are not fully accurate in nature, some amount of error will be involved in the predictions. Therefore, above equation can be rewritten in the following terms as shown in the equation (2).

$$y = mx + c + E \quad (2)$$

Error  $E$  should be minimized in order to make effective predictions and get the accurate results. The workflow of the Linear Regression algorithm is represented in the below Fig. 2.

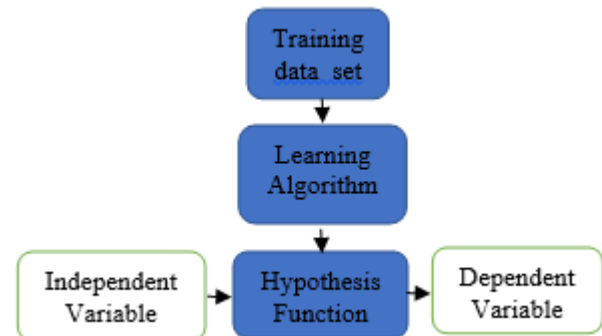


Fig. 2. Flowchart of Working of Linear Regression Algorithm

The hypothesis function will map the inputs variables to the output which will gives the predicted values. It is represented as follows in the equation (3).

$$H_{\theta}(x) = \theta_0 + \theta_1(x) \quad (3)$$

The accuracy of the algorithm which was built lies in choosing the  $\theta_0, \theta_1$  values and it should be in such a way that they should reduce the error in the algorithm. For that to be possible the values of  $\theta_0, \theta_1$  should be as minimum as possible. They will decide the straight line going to fit in the data for the prediction.

#### C. Cost Function or Squared Function

The minimization function which will reduce the values of  $\theta_0, \theta_1$  is known as Cost function or Squared Error function. It is denoted as  $J(\theta_0, \theta_1)$ . This cost function is formulated as follows in equation (4).

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^i) - y^i)^2 \quad (4)$$

Here  
 $J(\theta_0, \theta_1)$  = cost function  
 $\theta_0, \theta_1$  = parametres  
 $m$  = number of training examples  
 $i$  = count of the current training example  
 $h_{\theta}(x^i)$  = hypothesis function

#### D. Gradient Descent Algorithm

The algorithm which is used for minimizing the cost function  $J$  is Gradient Descent algorithm. It will be doing the iteration for a number of times until we get the optimized values for  $\theta_0, \theta_1$ . The objective of this algorithm is to reduce the  $\theta_0, \theta_1$  values. The initial assumptions for the values  $\theta_0, \theta_1$  will be taken as 0, 0 respectively. Then the algorithm will perform the iterations and simultaneously updates the values. After replacing them with new values it will check the compatibility and again replaces with new values until it reaches the minimization point of  $\theta_0, \theta_1$ . This is formulated as below.



Repeat until convergence {

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$$

}

(for  $j = 0$  and  $j = 1$ )

Here

$\alpha$  = learning rate of algorithm

$$\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = \text{slope}$$

The simultaneous updating of values of  $\theta$  after finding each value will be as follows given below.

$$\text{temp } 0 = \theta_0 - \alpha \frac{\partial}{\partial \theta_0} J(\theta_0, \theta_1)$$

$$\text{temp } 1 = \theta_1 - \alpha \frac{\partial}{\partial \theta_1} J(\theta_0, \theta_1)$$

$$\theta_0 = \text{temp } 0$$

$$\theta_1 = \text{temp } 1$$

The term  $\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1)$  in the mentioned formula above

will determine the slope of the tangent line to the point  $\theta$ . This term will decide how the  $\theta$  value should update in order to reach the minimal point. This is shown in the below Fig. 3.

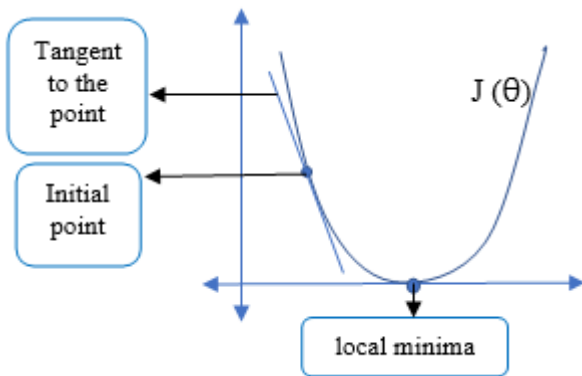


Fig. 3. Local Optima Point

The slope will determine the point  $\theta$  whether to move right or left on the curve to minimize and to obtain local optimal point. Some of conditions of slope are given below.

- 1) If the slope is positive,  $\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \geq 0$ , the point  $\theta$  will move towards left to reach local optimal point.
- 2) If the slope is negative,  $\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) \leq 0$ , the point  $\theta$  will move towards right to reach local optimal point.
- 3) If slope is,  $\frac{\partial}{\partial \theta_j} J(\theta_0, \theta_1) = 0$ , this point is called local optimal point.

As we approach the local optimum point, the gradient descent algorithm itself will take smaller steps in such a way to reach the minimum point exactly. So, there is no need to reduce the learning rate  $\alpha$ .

## E. Data Collected

The data sets collected from the engine testing which was maintained at load 0 and 7 N and temperature limits obtained were ranging between 154-345°C are shown below. From the graph, we can say that the  $\text{NO}_x$  and smoke emissions which were collected, increased with the increase in load and temperature. The graphs are shown in below Fig.4 and Fig.5.

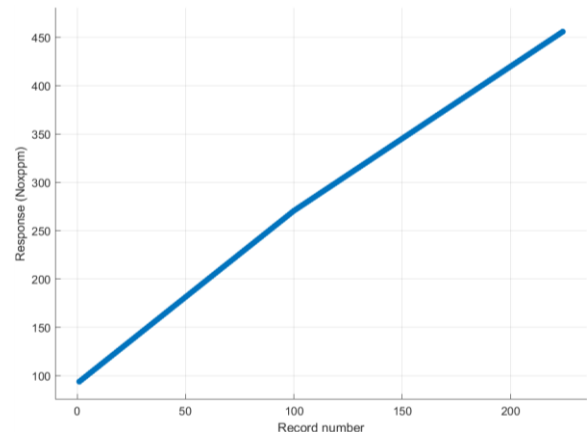


Fig. 4.  $\text{NO}_x$  Emission Data Collected

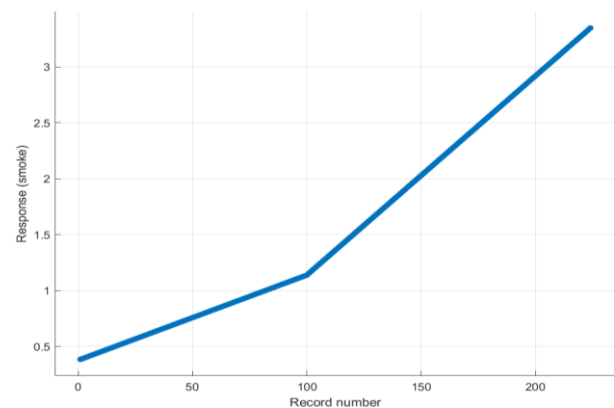


Fig. 5. Smoke Data Collected

## F. Training

The prediction analysis was performed in MATLAB, Regression Learners Application which is a part of Statistics and Machine Learning Toolbox. The data sets collected were imported into the Regression Learners Application and the features such as predictor and response variables were selected. It allows the user to select the mode of prediction analysis like validation and without validation processes which depends upon the results which are going to predict. This toolbox provides all types of regression algorithms available in the machine learning and gives the freedom to select among those.

## G. Types of Errors

The prediction results are evaluated on the basis of error between the predicted values and the target values. The regression learner application provides four different types of errors with the errors are calculated and the results are validated. Root Mean Square Error is the standard deviation of residuals (predicted errors). Hence, this error will show how far the data points are from the regression line. R-Squared Error is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. Mean Squared Error tells the deviation of the residuals. Mean Absolute Error measures the average vertical distance between the continuous points and identity line. All these errors are formulated as below and their limits are shown in the Table III.

$$\text{Residual} = \text{actual point} - \text{predicted point}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (actual_i - predicted_i)^2}{N}}$$

Here  $N$  = number of data points

$$Mean\ Squared\ Error = \frac{1}{N} \sum (actual - predicted)^2$$

$$Absolute\ Error = |actual - predicted|$$

$$Mean\ Absolute\ Error = \frac{1}{N} \sum |actual - predicted|$$

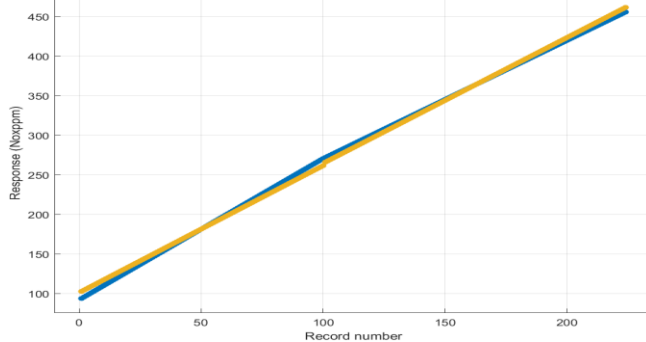
**Table- III: Types of Errors**

S. No	Error	Limits
1	Root Mean Square error	Smaller Values
2	R-Squared Error	Close to 1
3	Mean Squared Error	Smaller Values
4	Mean Absolute Error	Smaller Values

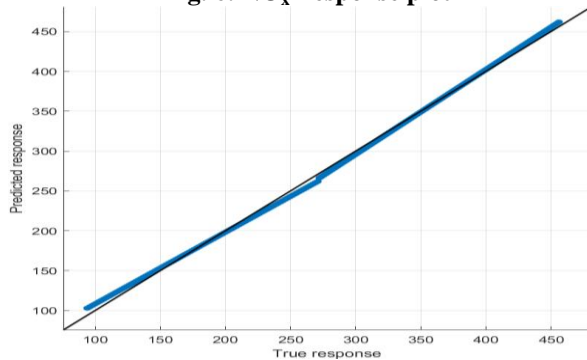
## V. RESULTS AND COMPARISION

### A. NO<sub>x</sub> Predictions

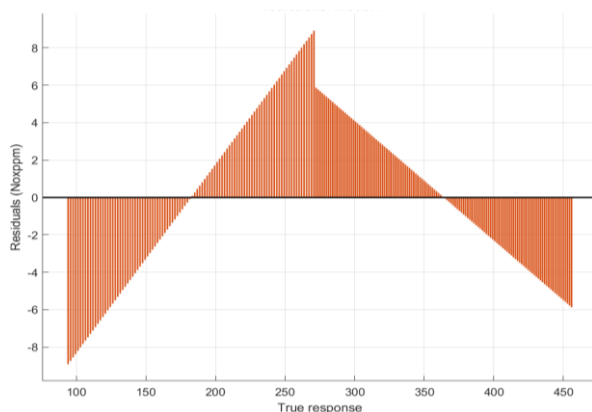
The observations are drawn between NO<sub>x</sub> on vertical axis and load, exhaust temperature on horizontal axis. The results are shown in following figures Fig.6, Fig.7, Fig.8.



**Fig. 6. NO<sub>x</sub> Response plot**



**Fig. 7. NO<sub>x</sub> Predicted vs Actual Plot**



**Fig. 8. NO<sub>x</sub> Residuals Plot**

### B. NO<sub>x</sub> Statistical Results

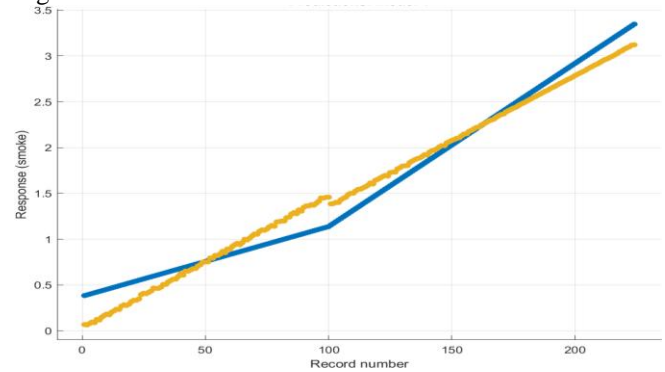
The statistical results for NO<sub>x</sub> predictions are shown in the below Table IV.

**Table- IV: NO<sub>x</sub> Statistical Results**

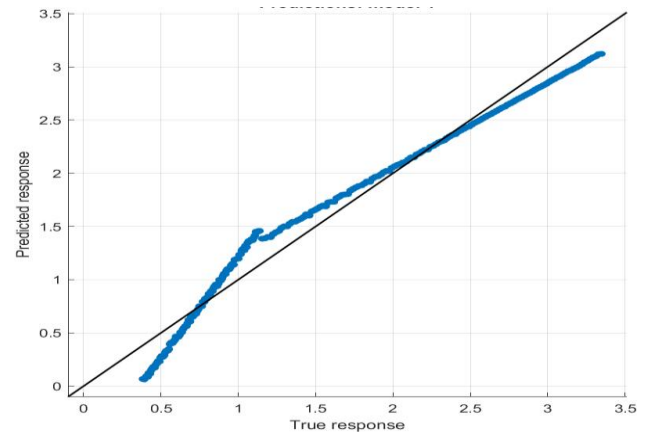
S. No	Error	Value
1	Root Mean Square error	4.304
2	R-Squared Error	0.98
3	Mean Squared Error	18.542
4	Mean Absolute Error	3.6477
5	Training Time	7.4389 sec

### C. Smoke Predictions

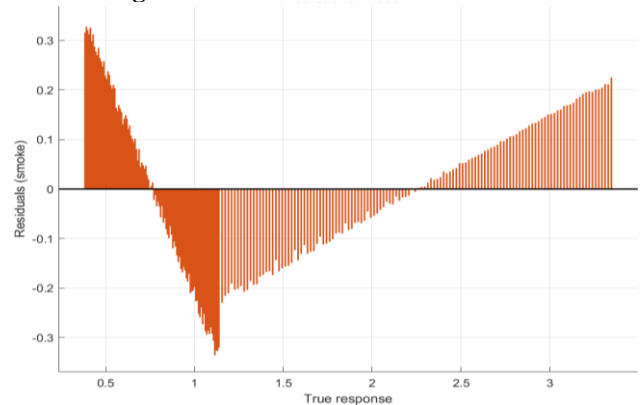
The observations are drawn between smoke values on vertical axis and load, exhaust temperature on horizontal axis. The results are shown in following figures Fig.9, Fig.10, Fig.11.



**Fig. 9. Smoke Response Plot**



**Fig. 10. Smoke Predicted vs Actual Plot**



**Fig. 11. Smoke Residuals Plot**

## D. Smoke Statistical Results

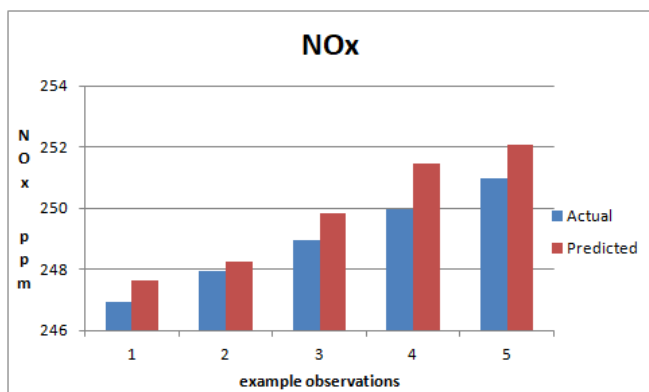
The statistical results for NO<sub>x</sub> predictions are shown in the below Table V.

**Table- V: Smoke Statistical Results**

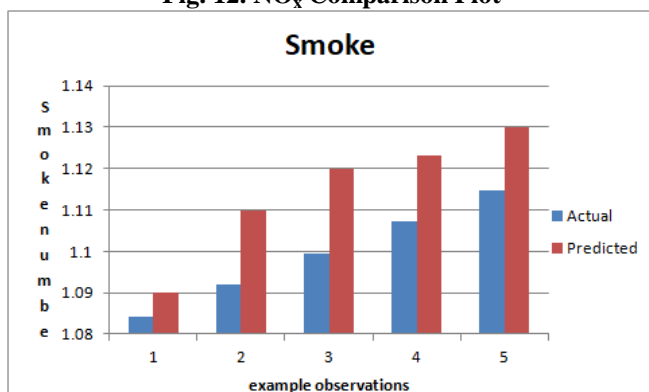
S. No	Error	Value
1	Root Mean Square error	0.1638
2	R-Squared Error	0.92
3	Mean Squared Error	0.026831
4	Mean Absolute Error	0.13901
5	Training Time	5.2489 sec

## E. Comparison plots

The predicted data are compared with the actual values for further more accuracy of the prediction model. The comparison plots of NO<sub>x</sub> and Smoke are shown in following figures Fig.12, Fig.13.



**Fig. 12. NO<sub>x</sub> Comparison Plot**



**Fig. 13. Smoke Comparison Plot**

## F. Validation

The results in performing the prediction of the emissions in an engine are appropriate with an accuracy of 98 % and 92 % in NO<sub>x</sub> and smoke respectively. Therefore, this model is good in performing prediction analysis of emissions of NO<sub>x</sub> and Smoke from the analysis of graphs and the error results obtained. It is explained as below.

- 1) Response Plot: The predicted data points should be as much as near to the actual values. We can see that the predicted points in the above-mentioned graph are pretty near to the observations or real values.
- 2) Predicted vs Actual Plot: In this plot, the predicted points should be as near as to the fitted regression line. This will say

that the predicted points are with less error as a measure in the vertical distance.

3) Residual Plot: This plot tells us about how appropriate the model is for the data. If there is no pattern seen in the graph but a random distribution of points it says that linear regression is good for the data. Otherwise non-linear regression should be used. So, it is observed that linear regression model is good for the data.

The model prediction accuracy could also be said by the statistical analysis presented above in the tabular forms.

1) RMSE error should be as low as possible for more accuracy and goodness of fit. It is observed that very less RMSE error in the smoke prediction and little bit more in NO<sub>x</sub> predictions.

2) R-Squared error is allowed to be in the limits between 0 and 1 and it should be close to 1. Therefore, good results in both smoke and NO<sub>x</sub> predictions are obtained.

3) MSE and MAE should also be as low as possible in order to increase the accuracy of the model. In both the predictions, low values for MSE and MAE are observed.

## VI. CONCLUSION

Linear Regression algorithm is appropriate and better in performing the predictions of engine emissions. The data transformation and optimization techniques will help more the prediction model to be accurate. The errors shown are acceptable with less training time and with R-Squared error as 0.98 and 0.92 for NO<sub>x</sub> and Smoke predictions respectively. Linear regression, fitting straight line may not be efficient in all the types of data like data sets in which the observations trend may not be linear, but by fitting nonlinear curve in the data may reduce the consistency with other data sets which use the same prediction models in future. This will show less bias but more variance with other data sets. Linear Regression is more bias and less variance with the other data sets. It will maintain consistency in performing the predictions. This consistency indeed needed by any good prediction model for achieving better results.

## REFERENCES

1. Jiang Zheng and Aldo Dagnino, "An Initial Study of Predictive Machine Learning Analytics on Large Volumes of Historical Data for Power System Applications" IEEE International Conference on Big Data, 2014.
2. Gulden Kaya Uyanik and Nese Guler, "A study on multiple linear regression analysis" Procedia - Social and Behavioral Sciences 106 (2013) 234 – 240.
3. Erdi Tosun, Kadir Aydin, and Mehmet Bilgili, "Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures" Alexandria Engineering Journal (2016) 55, 3081–3089 Journal, 2016.
4. Y. C. ay, I. Korkmaz, A. C. F. Kara, "Prediction of engine performance and exhaust emissions for gasoline and methanol using artificial neural network", Energy 50, 2013.
5. Y. Cay, "Prediction of a gasoline engine performance with artificial neural network", Fuel 111, 2013.
6. M. Canakci, A.N. Ozsezen, E. Arcaklioglu, A. Erdil, "Prediction of performance and exhaust emissions of a diesel engine fueled with biodiesel produced from waste frying palm oil", Expert Syst. Appl. 2009.
7. G. Najafi, B. Ghobadian, T.F. Yusaf, H. Rahimi, "Combustion analysis of a CI engine performance using waste cooking biodiesel fuel with an artificial neural network aid", Am. J. Appl. Sci. 2007.
8. Soteris A. Kalogirou, "Applications of artificial neural-networks for energy systems" Applied Energy 67 (2000) 17-35.

9. Aulia Qisthi Mairizal a, b, Sary Awad a, Cindy Rianti Priadi b, c, Djoko M. Hartono b, Setyo S. Moersidik b, Mohand Tazerout a, Yves Andres, "Experimental study on the effects of feedstock on the properties of biodiesel using multiple linear regressions" Renewable Energy 145 (2020) 375-381.
10. Alex Smola and S.V.N. Vishwanathan, "Introduction to Machine Learning" by Vishy, Revision: 252, (2010).
11. Subhradeep Biswas & Sudipa Biswas, "Inverse Linear Regression in Machine Learning", Volume 17 Issue 3 Version 1.0 Year 2017.
12. Kevin P. Murphy, "Machine Learning: A Probabilistic Perspective", by Massachusetts Institute of Technology (2012).
13. Yashavant S. Ingle, Prof. Anil Mokhad, "Use of Linear Regression in Machine Learning for Ranking", -International Journal for Scientific Research & Development| Vol. 1, Issue 5, (2013).
14. Henrik Almér, "Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles", Degree project in computer science, second level stockholm, Sweden (2015).
15. Brett Lantz, "Machine Learning with R- Second Edition", (2015).  
Jaiwei Han and Micheline Kamber, "Data Mining: Concepts and Techniques Second Edition".
16. Srihari S., Dr. Thirumalini S., and Prashanth, K., "An experimental study on the performance and emission characteristics of PCCI-DI engine fuelled with diethyl ether-biodiesel-diesel blends", Renewable Energy, 2017, vol. 107, pp. 440 – 447.
17. M N V R S S Sumanth and Sivanesan Murugesan, "Experimental Investigation of Wall Wetting Effect on Hydrocarbon Emission in Internal Combustion Engine", IOP Conf. Series: Materials Science and Engineering 577 (2019) 012029
18. Sivanesan Murugesan. Lakshmikanthan Chinnasamy, Abhijeet Patil., "Developing & Simulating the Duty Cycle on Engine Dynamometer based on Engine RWUP", "SAE Technical Paper", 2015, 2015-26-0024

## AUTHORS PROFILE



**Kongara Venkatesh**, Mechanical Engineering department, Amrita School of Engineering, Coimbatore, Amrita Viswa Vidyapeetham, India.  
Email: kmani733@gmail.com



**Sivanesan Murugesan, M. Tech**, Assistant Professor, Department of Mechanical Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore.  
Email: m\_sivanesan@cb.amrita.edu