International Journal for Research in Engineering Application & Management (IJREAM)
ISSN : 2454-9150 Vol-07, Issue-01, APR 2021

Fuel Prediction Based on Driving Behavior using Machine Learning Techniques

[1]Malarvizhi K, [2]Abinaya Subamyl M, [3]Harshithaa P V, [4]Nethra R

[1]Associate professor, [2,3,4]Student, Coimbatore Institute of Technology, Coimbatore, India,

[1]malarvizhi@cit.edu.in, [2]1807001it@cit.edu.in, [3]1807021it@cit.edu.in, [4]1807037it@cit.edu.in

Abstract : **The fuel consumption prediction based on driving behavior is very important in automobile industry. As there is a hike in fuel price on daily basis, it is important to know the amount of fuel consumed by the vehicle. Each vehicle running on road have a default mileage rate that is determined by the manufacturer with the help of initial vehicle conditions but it varies with the driving styles exhibited by the driver. In the proposed method the pre processed dataset is passed to the two ML (machine learning)algorithms, LR(Linear Regression) and RF(Random Forest). This method lays a foundation for fine management of transportation fuel consumption. The trained model of LR is integrated with Streamlit framework and a web application is designed for better user experience. Experimental results on real-world values are carried and compared with the recent existing works which shows a better results in terms of accuracy. Driving behavior has a large impact on vehicle fuel consumption. Driving behavior can be measured using various types of sensors connected to a control area network. The measured multi-dimensional time series data are called driving behavior data. Modeling and predicting fuel consumption based on driving behavior is a vital in enhancing fuel economy of vehicles. Transportation is an important factor that affects energy consumption, and driving behavior is one of the main factors affecting vehicle fuel consumption. The fuel consumption prediction models are built using LR and RF. All two models could predict fuel consumption accurately. The random forest model is proved to have the highest accuracy and runs faster, making it suitable for wide application. This model lays a foundation for fine management of transportation fuel consumption.**

Keywords —Ml-Machine Learning, LR-Linear Regression, RF-Random Forest, Driving Behavior, Fuel Consumption, StreamLit.

I. Introduction

Over the past few years, automotive manufactures have been concerned about reducing emissions and the overall utilization of fuel resources that is associated with the transportation industry. This evolving problem has urged government agencies and decision-makers to set regulations and standards on efficiency and low emissions. Moreover, the high costs of oil, together with the rising worries about environmental and atmospheric pollution, has forced automotive manufacturers to the development and marketing of energy efficient vehicles, by adopting strategies such as (i) designing more efficient small displacement engines, (ii) reducing weight and coefficient of drag of the vehicle, (iii) usage of low profile tires to minimize rolling resistance, (iv) adding an electric power train along with the conventional fuel engine, etc. Worldwide, governments are imploring for more efficient vehicles; therefore, there have been outstanding advancements in the use of alternative and low emission fuels such as hydrogen combustion cells. For the past decade, the Japanese government has been urging Japan's automotive manufacturers to increase the development work spent on battery-powered electric vehicles (EVs) and hybrid electric vehicles (HEVs). Fuel cell electric vehicles (FCVs) such as hydrogen cells is one more types that is either used to generate power using hydrogen combustion engine which moves the vehicle or indirectly generating electricity to power up an electric motor [3].

Earlier, non-spark-ignition engines (diesel) were known for their weakness in terms of emissions and reliability. However, only very recently, modern technologies have significantly improved such engines. In general, diesel engines get better fuel mileage when compared with gasoline engines. Despite that, this work studies gasoline powered vehicles because they produce less harmful emissions and because the overall trend nowadays is moving towards gasoline and hybrid/electric vehicles. This paper brings to discussion fuel consumption in real-time using instantaneous vehicle parameters and tries to estimate such consumption using an SVM.

 DOI : 10.35291/2454-9150.2021.0173

II. TOOLS AND TECHNIQUES

A. *Technology*

Machine learning Ml is a field in artificial intelligence that concerns construction of systems which can learn from examples in different ways. The concept can be described with the following definition by Tom M. Mitchell . A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E. — Tom M. Mitchell In the context of this study the experience E is the information about fuel consumptions and the other data collected from the different data sources. The task T is to estimate fuel consumption and the performance measure P is related to the size of the error in the prediction.

B. *Performance Metrics for Regression Methods*

To evaluate how well an ml method for regression describes the underlying relationship several metrics may be applied to the trained model. The metrics I intend to use are described below.

C. *Mean Squared Error*

The mean squared error (mse) of a model is the average of the squares of the prediction errors. The error in this case is defined as the difference between the estimate and the true value. The mse incorporates both the variance of the estimator and its bias and can be expressed by (3.1) [16]. MSE(Variance) = Variance Estimate + Bias(Estimate, T rueV alue) 2 (3.1) Thus the mse assesses the quality of an estimator in terms of its variation and degree of bias. The root mean squared error (rmse) is simply the square root of the mse. Using the rmse as a measure will give the same results as using the mse, but the rmse can be considered a more meaningful representation of the error. In this study rmse will be used to evaluate the different models.

III. PROPOSED SYSTEM

A. *Data Collection*

This dataset was taken from the StatLib library which is maintained at Carnegie Mellon University. The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 3 multivalued discrete and 5 continuous attributes.

B. *Attribute Information*

1. mpg : continuous
2. Cylinders : multi-valued discrete
3. Displacement : continuous
4. Horsepower : continuous
5. Weight : continuous
6. Acceleration: continuous
7. Model year : multi-valued discrete
8. Origin : multi-valued discrete
9. Car name : string (unique for each instance)

C. *Feature Extraction*

By matching the data collected from the auto-mpg and the daily driving behavior of each driver and the corresponding fuel consumption could be obtained.

There are many driving behavior factors that affect the fuel consumption of vehicles. The features that most affect the fuel consumption is selected from the data set by feature extraction. The car specifications which had relationship with the driver behavior is correlated.

The high correlation of multiple indicators indicates that the method of using these data to predict fuel consumption is feasible.

D. *Model Construction*

The training set and test set were selected randomly, and the fuel consumption prediction models were built using Linear Regression and Random Forest.The prediction model uses predictor values to predict the targeted value. The predicted value is shown and further analysis is made on it's performance.

E. *Effect Evaluation*

Building the fuel consumption prediction models several times and cocomparing the accuracy and efficiency of the two prediction models ususing didifferent methods, the best method to predict vehicle fuel coconsumption based on auto-mpg is proposed.

IV. PROJECT FLOW DIAGRAM

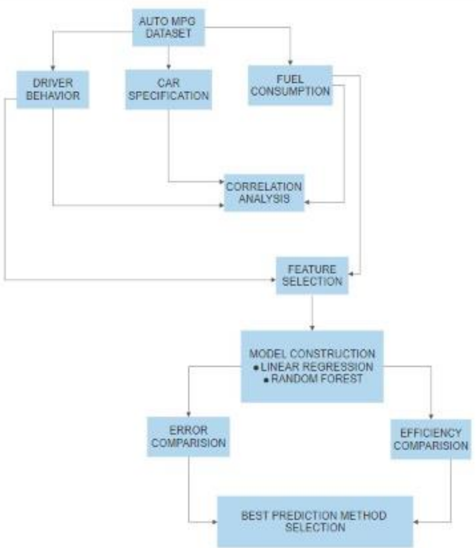

Fig.1.1 Project Architecture Diagram

 DOI : 10.35291/2454-9150.2021.0173

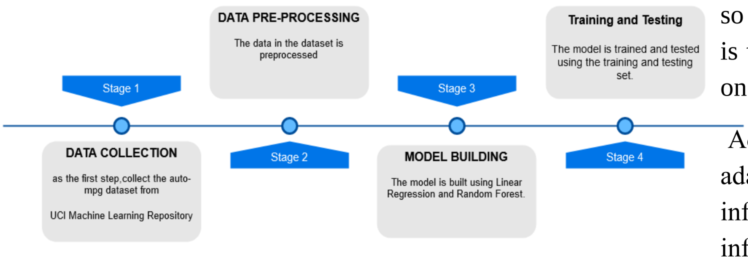

Fig.1.2 Project flow Diagram

V. DATA SET INFORMATION

A. Auto mpg:

This dataset is a slightly modified version of the dataset provided in the StatLib library. In line with the use by Ross Quinlan (1993) in predicting the attribute "mpg", 8 of the original instances were removed because they had unknown values for the "mpg" attribute. The original dataset is available in the file "auto-mpg.data-original"."The data concerns city-cycle fuel consumption in miles per gallon, to be predicted in terms of 3 multivalued discrete and 5 continuous attributes." (Quinlan, 1993).

VI. RELATED WORK

Wang et al. [4] examined the influence of driving patterns on fuel consumption using a portable emissions measurement system on ten passenger cars. They concluded that vehicle fuel consumption is optimal at speeds between 50 and 70 km/h and that fuel consumption increases significantly during acceleration. These results indicate that both the speed limit of the road and driver behavior have large impact on fuel consumption. Based on the mobile phone terminals and on-board diagnostic system (OBD) installed in taxis, driving behavior data and fuel consumption data are extracted, respectively. By matching the driving behavior data collected by a mobile phone with the fuel consumption data collected by OBD, the correlation between driving behavior and fuel consumption is explored, so that vehicle fuel consumption. The purpose of this paper is to improve fuel consumption monitoring databases based on mobile phone data.

Advanced driver assistance systems like cooperative adaptive cruise control (CACC) are designed to exploit information provided by vehicle-to-vehicle (V2V) and/or infrastructure-to-vehicle (I2V) communication systems to achieve desired objectives such as safety, traffic fluidity or fuel economy.

This information is used to train a set of nonlinear, autoregressive (NARX) models. Two scenarios are investigated, one of them assumes a V2V communication with the predecessor, the other uses only data acquired by on-board vehicle sensors. Depending on the applied approach and the moving space of the controlled vehicle, the thus obtained (imperfect) prediction allows fuel benefits in a range of 5% to 25% in the case of moderate, non-congested traffic.

To address the problems on road, driving behavior analysis and prediction models need to be developed.In this paper we have discussed some of existing driver behavior models. These models are classified as two types: Driver behavior analysis and driver behavior prediction models. In this paper we have carried out a detailed survey about the driver behavior analysis models and the driver behavior prediction models.

VII. RESULTS

The proposed system to predict the Fuel consumption rate based on driving behavior produced results which were accurate. In order to identify better algorithm to give more accurate results, Linear Regression and Random Forest Algorithm are compared using Performance Metrics: Mean Absolute error,Mean Squared error,Root Mean squared error and R2 scores.

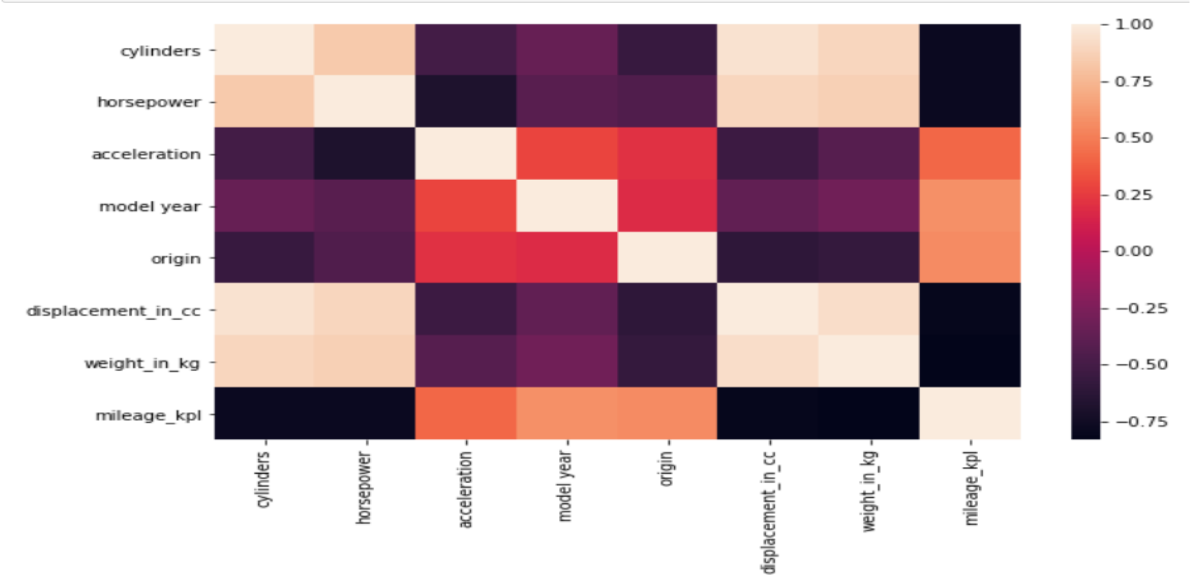

Fig.1.3 Corelation matrix

 DOI : 10.35291/2454-9150.2021.0173

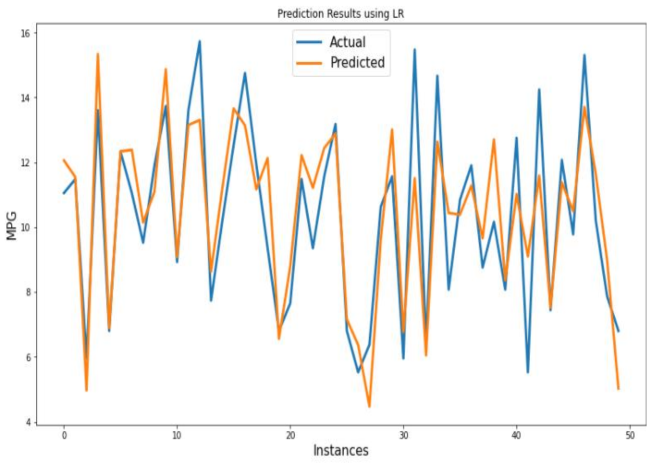

Fig.1.4 Linear Regression graph between actual values and predicted values.

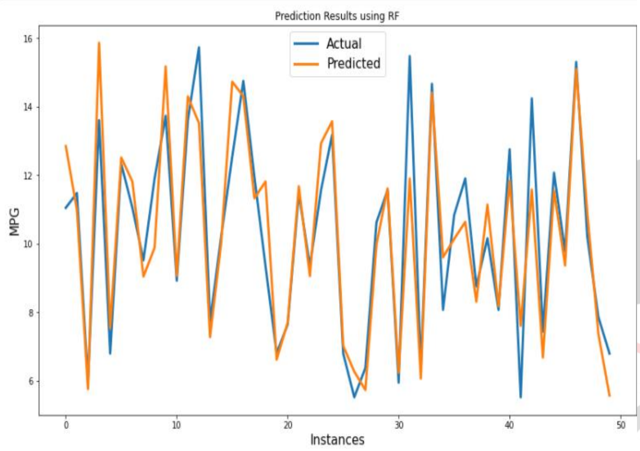

Fig.1.5 Random Forest graph between actual values and predicted values.

On observing the two graphs , we could find that in Linear regression, the predicted value and the actual lines do not go hand in hand. There is much difference in the predicted value. So we could infer that the linear Regression Algorithm does not predict accurate values. But if we consider the Random Forest Regressor graph, it is seen that the actual and predicted values are much similar with minimal variation. Thus we could conclude that the Random Forest predicts much accurately than Linear Regression in our project.

Fig.1.6 Web Application

When the inputs are given for number of cylinders , Horsepower, Time it takes to reach 0-60mph in seconds(Acceleration),Model Year, Origin of the Car, Engine Displacement in cc, weight in kg the accurate results for Mileage is predicted.

VIII. Conclusion

Thus, our model is shown to predict the values accurately given the right inputs. This project can also be used by car manufacturers in order to manufacturer fuel efficient vehicles. Users can also learn to drive efficiently by consuming less amount of fuel . As we can infer, the driving behaviour can have immense effect directly or indirectly on the fuel consumption rate. In our model, we used some parameters which are indirectly affected by the driving behaviour.The collected data is significantly more reliable, so our model will be able to perform better with a different, more accurate dataset. As a future implementation, we can improvise the project and make it to predict fuel consumption by integrating it with hardware module and bring iot together with machine Learning.The model can further be trained by having more number of features and use the method of combining them to provide more parameterized output. We can also increase the data set and train them in the cloud and deploy web services.

Acknowledgment

We express our sincere thanks to our Secretary **Dr.R.Prabhakar** and our Principal **Dr.V.Selladurai** for providing us a greater opportunity to carry out our work. The following words are rather very meager to express our gratitude to them. This work is the outcome of their inspiration and product of plethora of their knowledge and rich experience.

We record the deep sense of gratefulness to **Dr.N.K.Karthikeyan**, Head of the department of Information Technology, for his encouragement during this tenure.

We equally tender my sincere thankfulness to **Ms.K.Malarvizhi**, our guide, Department of Information Technology, for her valuable suggestions and guidance during this course.

During the entire period of study, the entire staff members of the Information Technology have offered ungrudging help. It is also a great pleasure to acknowledge the unfailing help we have received from our friends.

It is a matter of great pleasure to thank our parents and family members for their constant support and cooperation in the pursuit of this endeavor. We express our sincere thanks to God Almighty for His cherished blessings.

 DOI : 10.35291/2454-9150.2021.0173

REFERENCES

[1].J. E. Meseguer, C. T. Calafate, J. C. Cano, and P. Manzoni, ‘‘Assessing the impact of driving behavior on instantaneous fuel consumption,’’ inproc. 12th annu. IEEE consum. Commun. Netw. Conf. (ccnc), Jan. 2015,

[2]. H. Liimatainen, ‘‘Utilization of fuel consumption data in an ecodrivingin centive system for heavy-duty vehicle drivers,’’ IEEE trans. Intell. Transp. Syst., vol. 12, no. 4, pp. 1087–1095, Dec. 2011.

[3]. S. Kanarachos, j. Mathew, and m. E. Fitzpatrick, ‘‘Instantaneous vehiclefuel consumption estimation using smartphones and recurrent neural net-works,’’ expert syst. Appl., vol. 120, pp. 436–447, Apr. 2019.

[4]. P. Ping, W. Qin,Y. Xu, C. Miyajima, and T. Kazuya, ‘‘Spectral Clusteringbased approach for evaluating the effect of driving behavior on fuel economy,’’ in proc. IEEE int. Instrum. Meas. Technol. Conf. (i2mtc), Houston,tx, USA, May 2018,

[5]. A. Likas, N. Vlassis, and J. J. Verbeek, ‘‘The global k-means clustering algorithm,’’ pattern recognit., vol. 36, no. 2, pp. 451–461, Feb. 2003.

[6].H. Berndt, S. Wender, and K. Dietmayer, “Driver Braking behaviorduring intersection approaches and implications for warning strategiesfor driver assistant systems,” in proc. IEEE intell. Vehicles symp.,Jun. 2007, pp. 245–251.

[7]. J. C. Mccall and M. M. Trivedi, “Driver behavior and situation awarebrake assistance for intelligent vehicles,” proc. IEEE, vol. 95, no. 2,pp. 374–387, Feb. 2007.

[8].W. Khaisongkram, P. Raksincharoensak, M. Shimosaka, T. Mori, T. Sato,and M. Nagai, “Automobile driving behavior recognition using boostingsequential labeling method for adaptive driver assistance systems,” inki 2008: advances in artificial intelligence. Berlin, Germany: springer,2008,

[9].Support vector machines for binary classification. Accessed:Jun. 19, 2019. [online]. Available: https://ww2.Mathworks.Cn/help/stats/support-vector-machines-for-binary-classification.Html

[10].R. J. Kate and R. J. Mooney, ‘‘Semi-supervised learning for semanticparsing using Support Vector machines,’’ in proc. NACCL-short hum.Lang. Technol. Conf. North amer. Chapter assoc. Comput. Linguistics,2007, pp. 8184.

 DOI : 10.35291/2454-9150.2021.0173