



Driving Mode Estimation Model Based in Machine Learning Through PID's Signals Analysis Obtained From OBD II

Juan José Molina Campoverde^(✉)

Automotive Mechanical Engineering: GIIT Transport Engineering
Research Group, Salesian Polytechnic University, Cuenca, Ecuador
jmolina4@est.ups.edu.ec

Abstract. In this paper a driving mode estimation model based in machine learning architecture is presented. With the statistic method, Random Forest, the highest inference of driving variables is determined through the best attributes for a training model based in OBD II data. Engine sensors variables are obtained with the aim of explaining the behavior of the PID signals in relation to the driving mode of a person, according to specific consumption and engine performance, characterizing the signals behavior in relation to the different driving modes. The investigation consists of 4 power tests in the dynamometer bank at 25%, 50%, 75% and 100% throttle valve opening to determine the relationship between engine performance and normal vehicle circulation, through the engine most influential variables like MAP, TPS, VSS, Ax and each the transmission ratio infer in the fuel consumption study and engine performance. In this study Random Forest is used achieving an accuracy rate of 0.98905.

Keywords: Random forest · OBD II · Fuel consumption · Eco drive · K-means · S-Golay

1 Introduction

Facing the increase in environmental pollution caused by greenhouse gas emissions produced by combustion engines [1], electronic injection systems were implemented. Through the on-board diagnostic (OBD-II), available since 1994 [2], different variables of sensors are obtained, like longitudinal acceleration (A_x), Engine Speed (RPM), Vehicle Speed Sensor (VSS), Manifold Air Flow (MAF), Manifold Absolute Pressure (MAP) [3]. A methodology used to determine sensors behavior is carried out by driving modes study (Sport, Normal and ECO) where Lv et al. [4], highlight behaviors and interactions between the driver and the vehicle, these modes are identified according to driving styles by means of maneuvers [5], to estimate fuel consumption and polluting emissions [6]. Due to the average driver can increase efficiency by 10% simply by changing the driving mode [7], Hooker et al., in case of Eco drive mode, covers low consumption driving behaviors due to the accelerator pedal differential use [8], which influences in fuel economy and vehicle emissions to reduce its environmental impact [9]. In addition, an efficient driving mode is the main solution to reduce greenhouse

gases emission produced by the vehicles [10]. According to Sivak et al., factors such as drivers decision respect to the vehicle maintenance or route selection influence in fuel economy [9], as well as making early gear changes avoid high speeds, reduce polluting emissions and fuel consumption [11, 12]. Lapuerta et al. [13], considers how it infers the altitude above sea level effect in internal combustion engines due to the air density that modifies its thermodynamic cycle and as consequence decreases the engine power, increases the polluting emissions and fuel consumption, considering Cuenca altitude, 2550 m.a.s.l. [14].

2 Materials and Methods

2.1 Experimental Overview

For the development of the present investigation, a review of previous researches was realized to propose an adequate data acquisition and analysis methodology. Zhou et al. [15], propose an optimization model of engine repair based on the classification of a Random Forest algorithm, which by adaptation and decision of more significant variables allows classifying engine real data, demonstrating that the more samples are trained the higher the classification accuracy. Considering the weight of applied variables can reach 79.1% of precision in the applied model. Damström et al. [16], propose a methodology that consists of finding the ideal number of groupings controlling similar characteristics between them by means of a cross correlation, but it has the disadvantage that it doesn't guarantee to find a global minimum which the approximation of cluster centroid changes its results. Based in the Corcoba et al. methodology [17], a diffuse logic system was used to simulate human behavior to estimate an efficient driver behavior by input and output variables, its output is rated with a classification between 0 to 1 depending how efficient the driving is, the proposal characterized the driving style by the data acquisition through telemetry. Pereira et al. [18], present a mobile application based on diffuse logic, through OBD-II data acquisition to characterize the driving efficiency by estimating the specific consumption with an accuracy of 85% in this model.

Oñate et al. [19], propose an intelligent system of erratic driving diagnosis through OBD-II and GPS data acquisition, which proposes reduce accidents by extracting driving signals, mathematical signal analysis and driving fault detection through Neural Networks, evaluating the model they guaranteed its viability and reliability. Chen et al. [20], present a semi – supervised classification method through a Support Vector Machine (SVM) algorithm, by data acquisition of velocity and engine speed data by OBD II from gasoline and diesel passengers vehicles with the vehicle identification number (VIN), it can classify the vehicles through a specific classification rule, whose purpose is applying to vehicles with internet connection and to improve estimation of CO2 emissions. Corcoba et al. present an efficient driving model through OBD-II data acquisition by a smartphone which provides ecological driving suggestions to the driver for bad driving habits, with the aim of improve the driving style through the efficiency consumption by algorithms such as J48 [21], Naïve Bayes and Random Forest, where the last one allows a better control of random attributes for each node and

was selected because its execution time is less than J48, especially when the attributes number or the training set is bigger with a classification certainty of 83.17 and 71.15% respectively.

Given the background, the following methodology is proposed (Fig. 1):

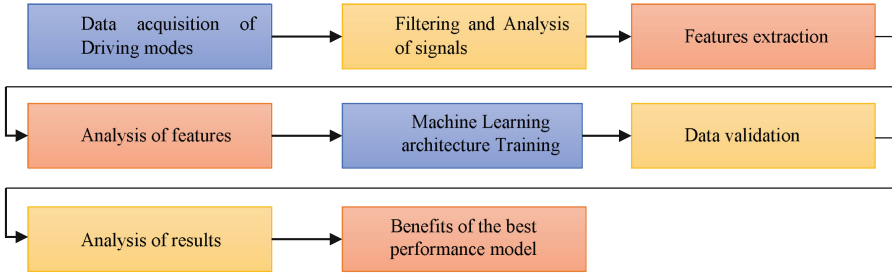


Fig. 1. Methodology

2.2 OBD-II Data Acquisition Through Data Logger Device

OBD-II data acquisition is realized by the data logger device, “Freematics ONE+”, which allows to extract all the Electronic Control Unit (ECU) sensors data and register the information in a “.CSV” archive, it’s saved in a micro SD for a later processing and research. Stored Variables are shown on Table 1.

Table 1. Stored variables

Variables	Nomenclature	Units
Longitudinal acceleration	Acceleration	m/s ²
Vehicle speed sensor	VSS	m/s
Engine coolant temperature	ECT	°C
Intake air temperature	IAT	°C
Manifold absolute pressure	MAP	KPa
Oxygen sensor	O2	V
Engine speed	RPM	rpm
Throttle position sensor	TPS	%
Long term fuel trim	LTFT	%
Short term fuel trim	STFT	%

For the experiment development, some test were realized on the dynamometer bank MAHA LP 3000 based in the ISO 17359 2018 “Condition monitoring and diagnostics of machines” [22] to obtain engine power and specific consumption data according to quality criterions.

To characterize method based in Meseguer et al. [1–3], Dörr et al. propose analyze driving incidence factors, such as, input, study, noise and controllable variables. These variables are shown in Table 2.

Table 2. Incidence factors

Variables	Factors	Symbol	Unit
Input	Longitudinal acceleration	A	[m/s ²]
	Vehicle speed	VSS	[m/s]
	Manifold absolute pressure	MAP	[kPa]
	Engine speed	η	[rpm]
	Throttle position sensor	TPS	[%]
Controllable	Coolant temperature	T Coolant	[°C]
	Rolling temperature	T Rolling	[°C]
	Lubricant temperature	T Lubricant	[°C]
Noise	Oxygen sensor	O2	[V]
	Air temperature	T Air	[%]
	Fuel temperature	T Fuel	[°C]
	Atmospheric pressure	P Atm	[Pa]
	Relative humidity of air	H Rel	[%]
Study	Power	P	[kW]
	Specific consumption	C esp	[g/kWh]

2.3 Signal Filtering

Motor signals when acquiring data present noise due to the existence of incidence factors in the vehicle, for this reason, these signals go through a filter stage; this consists in a comparison between real signals versus those obtained from sensors, for this purpose the S-Golay filter [24] was applied in order to obtain clean signals for a correct interpretation.

3 Results

3.1 Experimentation Results

With 4 power test, the vehicle performance was characterized as the volumetric fuel consumption that a low engine speed it is low, then when speed engine is increased by the increased throttle opening, the volumetric fuel consumption gets a gradually increase due to a bigger amount of air – fuel mixture causing more engine power, (see Fig. 2).

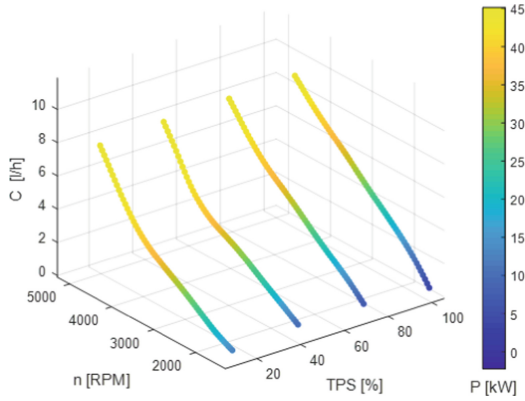


Fig. 2. Volumetric consumption

Relating obtained results of volumetric fuel consumption with engine power at 25%, 50%, 75% and 100% throttle valve opening, engine specific consumption is determined by (1):

$$\text{Equation: } C_{esp} = \frac{\rho \dot{v}}{P} \left[\frac{g}{kWh} \right] \tag{1}$$

Where:

\dot{v} = Fuel Volumetric Flow (l/h)

ρ = Fuel Density (g/l)

P = Engine Power (kW).

The engine specific consumption depends directly of engine speed and the input air – fuel mixture regulated by the throttle valve opening angle, (see Fig. 3).

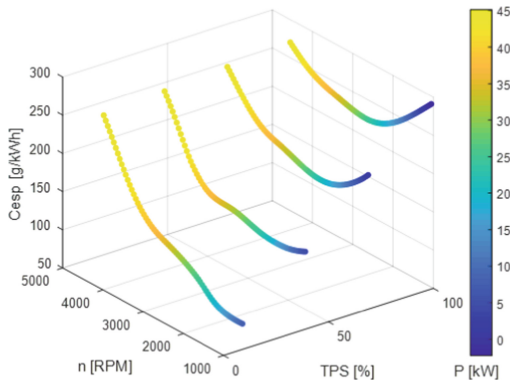


Fig. 3. Engine specific consumption.

3.2 Machine Learning Architectures Application

For obtained data application from dynamometer bank in real driving cycles a model that explain the relationship between engine performance and normal circulation vehicle is proposed, for later determine by means of an artificial neural network application the fuel consumption and finally to apply fuzzy logic to obtain the driving mode.

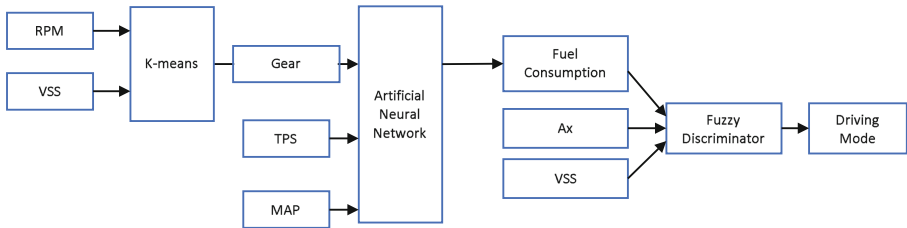


Fig. 4. Driving estimation model

Vehicle performance depends directly of transmissions ratios of each gear, so these values are obtained by the application of K-means method based in relation analysis between speed vehicle and speed engine (VSS/RPM) taken from a normal driving cycle, (see Fig. 4).

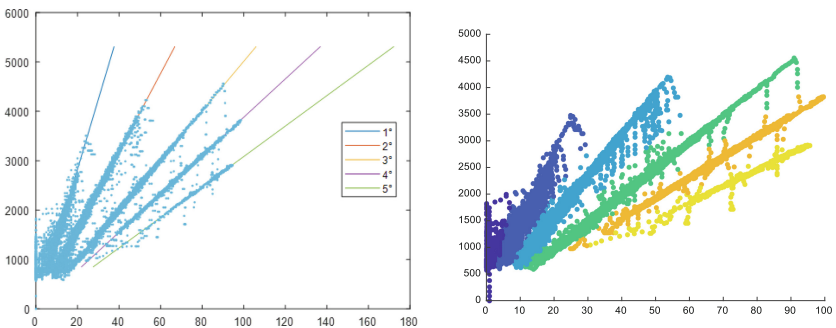


Fig. 5. K-means classification.

Where the following transmission ratios values are obtained (Table 3):

Table 3. Transmission ratios

Transmission ratios					
Rn	R1	R2	R3	R4	R5
0	0.0088	0.0137	0.0212	0.0299	0.0425

Fuel Volumetric Consumption depends directly of engine speed and the vehicle gear selected, as shown in (2) (see Fig. 5).

$$C_n \left[\frac{l}{h} \right] = R_n * RPM \tag{2}$$

Where:

- C_n = Fuel Volumetric Consumption for each gear (l/h)
- R_n = Transmission Ration for each gear
- RPM = Engine Speed (rpm)

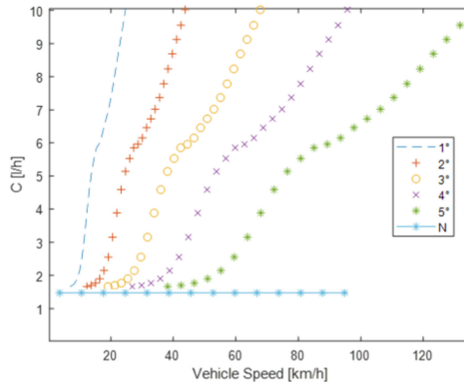


Fig. 6. Fuel volumetric consumption

Relating fuel consumption with engine power delivered, the specific consumption for each gear is found, in which can be noted minimum values for each one of the selected gear, that are ideal values in which the vehicle should circulate (see Fig. 6).

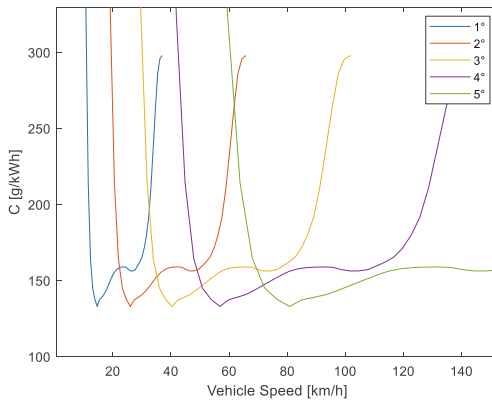


Fig. 7. Specific consumption for each gear.

Given the performance, maximum engine acceleration for each selected gear is determined, as shown in (3)

$$A_x = \frac{M_{nor} * \eta_i * \frac{R_c R_{Cn}}{r_e}}{m_{veh} * \gamma_m} \quad (3)$$

Where:

A_x = Longitudinal Acceleration

η_i = Gear efficiency

M_{nor} = Normal Momentum

$\frac{R_c R_{Cn}}{r_e} = R_n$ = Total Transmission Ratio

m_{veh} = Vehicle Mass

γ_m = Rotative Mass Coficient

n = Selected gear

$\gamma_m = 1,05 + 0,0025R_n^2$

Vehicle maximum acceleration for each gear depends on engine speed and due to transmission ratio of the vehicle velocity [25], as shown in Fig. 8.

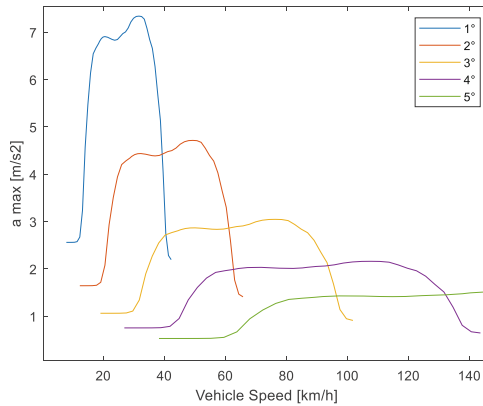


Fig. 8. Vehicle maximum acceleration for each gear

For an efficient driving gear changes should be done when specific consumption is relatively low according with Fig. 7. Table 4 shows values where fuel consumption is minimized.

Table 4. Specific consumption eco drive

Eco drive					
Gear change	Speed (km/h)	Rpm max	Previous RPM	Ax (m/s ²)	Previous Ax (m/s ²)
1–2	26,03	3400	1900	6,908	4,205
2–3	42,47	3100	2000	4,501	2,767
3–4	61,48	2900	2100	2,860	1,990
4–5	86,71	2900	2000	2,033	1,380

Random Forest algorithm is presented, which its main purpose is to find relationship between input and output study variables, to determine more important variables in fuel consumption according with driving characteristics. Through study attributes, Random Forest allows to classify real engines data, demonstrating that while more samples being trained the greater would be the classification accuracy [15], also it allows a better control of the random attributes number for each node with an execution time really efficient [21], proving that fuel consumption depends directly of MAP, TPS, VSS, Ax and the selected gear, but not from engine speed, because this factor is in close relationship with VSS and Ax, with a determination index of 0.98905, (see Fig. 9).

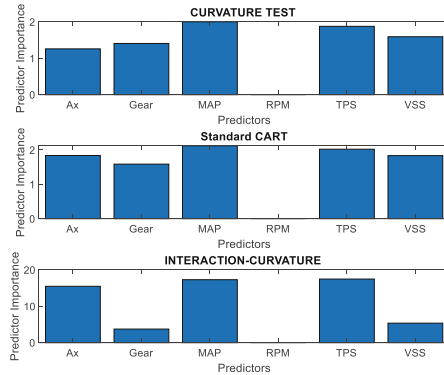


Fig. 9. Random Forrest.

From this, an artificial neural network is trained to estimate the volumetric fuel consumption based on determined variables with 16661 samples to estimate the driving mode. The prediction is efficient due to the high reliability of RNA with a global index of 0.98905 (Table 5).

Table 5. ANN training

	Samples	Samples (%)	R
Training	11663	70	0.98886
Validation	2499	15	0.98878
Testing	2499	15	0.99026

The found settings (see Fig. 10):

With this model an efficient vehicle speed is estimated by VSS and acceleration is determined to evaluate change gear according to engine speed. As the engine speed increases, a gear change is required, with the aim of increase vehicle speed and avoid excessive fuel consumption so that when the gear change is made the engine

revolutions fall until get the new gear, considering that if the change gear is not realized, fuel consumption gets increased.

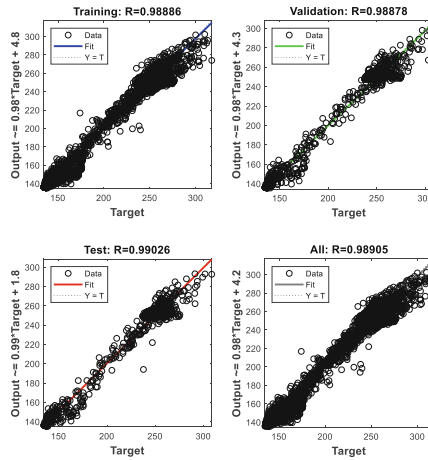


Fig. 10. Relationship between expected values and estimated values by ANN

By Predictor importance the strength of the relationship of the predictors and the PID data result is quantified by the import of the samples, the viability of the established model is determined according to its importance and the training given.

In this work a diffuse classifier such as Manami is proposed with a total of 43 rules, which has many input variables: speed, acceleration, volumetric fuel consumption of the vehicle and determines the driving mode between Sport, Normal and Eco. The classifier is applied to obtained data in real driving (see Fig. 11) during start-up the Sport driving mode is used meanwhile eco mode is used in decelerations and small displacements and for the rest of scenarios normal mode is frequently used.

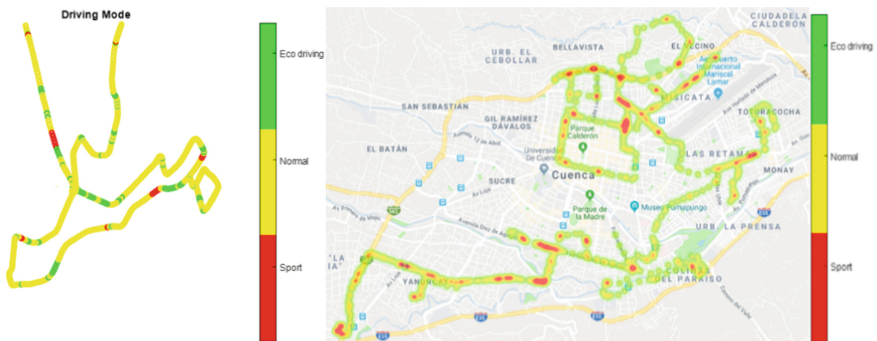


Fig. 11. Characterization of the driving mode [23]

Through the characterization of the driving modes it is determined that the Eco-driving mode is the most efficient due to the differentiated use of the acceleration pedal which establishes the specific consumption is low compared to the other modes by optimizing fuel consumption in a random route of conduction.

4 Conclusions

In this paper it was obtained a model that explains PID's signals behavior according to a person driving mode through thresholds values based on Machine Learning architecture by statistic method Random Forest that identifies the major importance variables on the estimation of fuel consumption in order to classify the driving modes in a random route, representing by sections the driving modes Eco drive, Normal and Sport according to driving behavior. Random Forest allows estimating sensors variables values that are unknown, with great precision through data set adjustments using characteristics of engine sensors variables, with the advantage that the model presents a minimum error for driving mode data classification of each vehicle with a determination index of 0.98905. PID's signals like MAP, TPS, VSS, Ax and the selected gear, directly influence in the person driving mode due to the big input mixture air – fuel, where highlights specific consumption, polluting emissions and vehicle performance considering that engine speed doesn't have a relationship with VSS and Ax, it doesn't infer in the model behavior. The application of K-means method based on the analysis of the relationship between vehicle speed and engine speed (VSS/RPM) though obtained data from the dynamometer bank establishes a model based on the specific consumption and engine performance for driving modes characterization under normal driving conditions.

This paper aims to facilitate the study of future projects for the determination of the estimation of the driving mode through Machine Learning architectures applied to the study of the analysis of polluting emissions.

Acknowledgment. To Mr. Nestor Diego Rivera Campoverde, for his direction and unconditional collaboration in the realization of the paper with his contributions and his knowledge throughout the entire process, in addition to the GIIT Transportation Engineering Research Group for his support for the completion of the paper.

References

1. Meseguer, J., Toh, C.K., Calafate, C.T., Cano, J., Manzoni, P.: Assessing the impact of driving behavior on instantaneous fuel consumption. In: 12th Annual IEEE Consumer Communications and Networking Conference (CCNC) (2015)
2. International Organization for Standardization, Road vehicles, Diagnostic systems, Keyword Protocol 2000 (1999)
3. Meseguer, J., Toh, C.K., Calafate, C., Cano, J.Y., Manzoni, P.: DrivingStyles: a mobile platform for driving styles and fuel consumption characterization. *J. Commun. Netw.* **19**(2), 162–168 (2017)

4. Lv, C., et al.: Cyber-physical system based optimization framework for intelligent powertrain control. *SAE Int. J. Commer. Veh.* **10**(1), 254–264 (2017)
5. Mohd, T.A.T., Hassan, M., Aris, I., Azura, C., Ibrahim, B.S.K.K.: Application of fuzzy logic in multi-mode driving for a battery electric vehicle energy management. *Int. J. Adv. Sci. Eng. Inf. Technol.* **7**, 284–290 (2017)
6. Dia, H., Panwai, S.: Impact of driving behavior on emissions and road network performance. In: *IEEE International Conference on Data Science and Data Intensive Systems* (2015)
7. Barkenbus, J.N.: Eco-driving: an overlooked climate change initiative. *Energy Policy* **38**(2), 762–769 (2010)
8. Hooker, J.N.: Optimal driving for single-vehicle fuel economy. *Transp. Res. A Gen.* **22**(3), 183–201 (1988)
9. Sivak, M., Schoettle, B.: Eco-driving: strategic, tactical, and operational decisions of the driver that influence vehicle fuel economy. *Transp. Policy* **22**, 96–99 (2012)
10. Andrieu, C., Saint Pierre, G.: Comparing effects of eco-driving training and simple advices on driving behavior. *Procedia – Soc. Behav. Sci.* **54**, 211–220 (2012)
11. Beckx, C., Panis, L.L., De Vlieger, I., Wets, G.: Influence of gear changing behavior on fuel-use and vehicular exhaust emissions. *Highw. Urban Environ.* **12**, 45–51 (2007)
12. McIlroy, R., Stanton, N., Godwin, L., Wood, A.: Encouraging eco-driving with visual, auditory, and vibrotactile stimuli. *IEEE Trans. Hum.-Mach. Syst.* **47**(5), 661–672 (2017)
13. Lapuerta, M., Armas, O., Agudelo, J., Sánchez, C.: Study of the altitude effect on internal combustion engine operation. Part 1: performance. *Technol. Inf.* **17**(5), 21–30 (2006)
14. Rivera, N., Chica, J., Zambrano, I., García, C.: Estudio del comportamiento de un motor ciclo otto de inyección electrónica respecto de la estequiometría de la mezcla y del adelanto al encendido para la ciudad de cuenca. *Revista Politécnica* **40**(1), 59–67 (2017)
15. Zhou, Y., Guo, J., Fu, L., Liang, T.: Research on aero-engine maintenance level decision based on improved artificial fish-swarm optimization random forest algorithm. In: *2018 International Conference on Sensing, Diagnostics, Prognostics, and Control* (2018)
16. Damström, J., Gerlitz, C.: Classification of Power Consumption Patterns for Swedish Households Using K-means. *K4. Power Consumption Analysis* (2016)
17. Corcoba, V., Muñoz, M.: Eco-driving: energy saving based on driver behavior. In: *XVI Jornadas de ARCA sobre Sistemas Cualitativos y sus Aplicaciones en Diagnosis, Robótica e Inteligencia Ambiental (JARCA)* (2015)
18. Pereira, A., Alves, M., Macedo, H.: Vehicle driving analysis in regard to fuel consumption using fuzzy logic and OBD-II devices. In: *2016 8th Euro American Conference on Telematics and Information Systems (EATIS)* (2016)
19. Oñate, J.A., Christian M. Quintero, G., Pérez, J.M.: Intelligent erratic driving diagnosis based on artificial neural networks. In: *IEEE ANDESCON* (2010)
20. Chen, S., Lin, R., Liu, W., Tsai, J.: The semi-supervised classification of petrol and diesel passenger cars based on OBD and support vector machine algorithm. In: *2017 International Conference on Orange Technologies (ICOT)* (2017)
21. Corcoba, V., Muñoz, M.: Artemisa: using an android device as an eco-driving assistant. *Cyber J.: Multidiscip. J. Sci. Technol. J. Sel. Areas Mechatron. (JMTC)*, June Edition, 3–7 (2011)
22. ISO 17359: Condition monitoring and diagnostics of machines (2018)
23. Google Maps. <https://www.google.com/maps>. Accessed 01 July 2019
24. Pang, C.K., et al.: Intelligent energy audit and machine management for energy-efficient manufacturing. In: *IEEE 5th International Conference on Cybernetics and Intelligent Systems (CIS)* (2011)
25. Aparicio, F., Vera, C., Díaz, V.: *Teoría de los vehículos automóviles*. Universidad Politécnica de Madrid, Madrid, pp. 279–285 (1995)