2017 IEEE International Conference on Big Data (BIGDATA)

# Application of Machine Learning for Fuel Consumption Modelling of Trucks

Federico Perrotta[1], Tony Parry[2]
University of Nottingham: *Faculty of Engineering*
Nottingham Transportation Engineering Centre, NTEC
Nottingham, United Kingdom
[1]Federico.Perrotta@nottingham.ac.uk
[2]Tony.Parry@nottingham.ac.uk

Luis C. Neves
University of Nottingham: *Faculty of Engineering*
Resilience Engineering Research Group
Nottingham, United Kingdom
Luis.Neves@nottingham.ac.uk

***Abstract*— This paper presents the application of three Machine Learning techniques to fuel consumption modelling of articulated trucks for a large dataset. In particular, Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) models have been developed for the purpose and their performance compared. Fleet managers use telematic data to monitor the performance of their fleets and take decisions regarding maintenance of the vehicles and training of their drivers. The data, which include fuel consumption, are collected by standard sensors (SAE J1939) for modern vehicles. Data regarding the characteristics of the road come from the Highways Agency Pavement Management System (HAPMS) of Highways England, the manager of the strategic road network in the UK. Together, these data can be used to develop a new fuel consumption model, which may help fleet managers in reviewing the existing vehicle routing decisions, based on road geometry. The model would also be useful for road managers to better understand the fuel consumption of road vehicles and the influence of road geometry. Ten-fold cross-validation has been performed to train the SVM, RF, and ANN models. Results of the study shows the feasibility of using telematic data together with the information in HAPMS for the purpose of modelling fuel consumption. The study also shows that although all the three methods make it possible to develop models with good precision, the RF slightly outperforms SVM and ANN giving higher $R^2$, and lower error.**

***Keywords - fuel consumption; machine learning; neural networks, random forests, support vector machine, truck fleet management***

## I. Introduction

Nowadays, one of the biggest challenges to face is the reduction of greenhouse gas (GHG) emissions from the transport industry. In particular, the road transport sector accounts for about 80% of the whole energy demand required by transportation and, due to its reliance on fossil fuels, represents one of the most important sources of GHG emissions in the world [1].

Smart routing is used by fleet managers to direct their vehicles and minimise costs. Usually, the shortest path (e.g. [2]) or the least congested route (e.g. [3]) is chosen, however, some studies (e.g. [4]) showed that the road geometry and the condition of the road infrastructure can significantly affect fuel economy. A new fuel consumption model that takes into account these two factors would therefore, help fleet managers in reviewing their routing decisions. Furthermore, the model would be useful for pavement engineers and road managers to estimate the life-cycle costs of new and existing roads.

In the past, several fuel consumption and emissions models have been developed that include the impact of the road infrastructure (e.g. gradient, roughness and macrotexture) (e.g. [4], [5], [6], [7]), however, most of these models base their estimates on standard drive cycles, are calibrated only for specific vehicles, or offer just a simplified mechanistic model. Although these capture the physical processes controlling the fuel consumption, they may be representative of only a few specific cases and may not describe what happens in reality. This is because of the assumptions made in the models. These regard the driving mode, constant speed, acceleration, weather conditions, etc. and using the same equations or methodologies in more general conditions may be computationally expensive or inaccurate due to the highly nonlinear phenomena involved.

In the era of 'Big Data', large quantities of data are continuously collected by companies all over the world. For example, truck fleet managers use standard sensors [8] installed on the most recent vehicles to optimize the operational costs of the fleet, for example, understanding when a vehicle needs maintenance, or a driver needs training. These are communicated using wireless telemetry and are commonly referred to as telematic data.

However, analyzing these data using traditional methods can be time consuming and computationally expensive. Moreover, when the number of data can be so large, it is difficult to select the most significant variables to include in a regression model to avoid overfitting.

Machine learning techniques are widely applied to a number of topics and represent the most advanced methods for regression problems. Among others, Support Vector Machine (SVM), Random Forests (RF), and Artificial Neural Networks (ANN) have been demonstrated to be powerful tools for regression analysis thanks to their learning ability and fault tolerance. These are applied on a daily basis to estimate, for example, the prices in the stock market (e.g. [9], [10]), hydrology (e.g. [11], [12]), and health monitoring (e.g. [13], [14]). More recently, these techniques have also been