Extended Abstract

Analysis of the feasibility of fuel consumption prediction of transport trucks in road works through machine learning

Hugo Filipe da Silva Freitas Fernandes

hugo.fernandes@ist.utl.pt

Instituto Superior Técnico, Universidade de Lisboa, Portugal, 2021

Abstract

In our current technological and business setting, the main goal is to do more with less, i.e., to use fewer resources, through process optimization. Therefore, the use of the latest technology in areas such as artificial intelligence, as a way to predict and estimate costs, is crucial. Knowing that the cost of fuels, more specifically in heavy vehicles transporting materials for road paving works, represents a significant part of the total cost of the process, this dissertation's goal is to analyze the feasibility of predicting fuel consumption. Through the installation of sensors in a transport truck, the real numbers of fuel consumption were analyzed. This database allowed the forecast of fuel consumption through the implementation of Artificial Intelligence models, more specifically machine learning. This has enabled an evaluation, considering the following dimensions: route traveled, average speed, loads transported and recording of the actual consumption associated with each route, which corresponds to the comparison term on which the study is based. This premise was validated in a real-life context (road construction site) in collaboration with a Portuguese construction company. The results of this study support the feasibility of the forecast, providing relevant insights into the advantages associated with the combination of sensorization for data collection and artificial intelligence in a real construction scenario.

Keywords: Fuel consumption, Machine Learning, Pavements, Sensors, Trucks

1. Introduction

One of the major costs in linear transport infrastructure projects is related to the fuel consumption of heavy machinery. Among these, construction material transport trucks are typically associated with a very high variability regarding their fuel consumption, as it is a function of several factors typically difficult to estimate. The latter include not only aspects related to the specifications of the equipment, but also with the weight of the transported cargo and the trajectory through which the transportation of materials is carried out. Typically, project concept and management teams use average values for fuel consumption of this type of equipment to estimate their fuel and budget requirements throughout construction. However, the average values are not accurate enough to allow designers to properly estimate fuel-related costs throughout a construction project. Thus, this work describes the first results of a study aimed at combining remote monitoring technologies based on intelligent sensors with artificial intelligence (AI) prediction models (i.e., machine learning (ML)), to create a framework that aims to estimate the fuel usage by trucks in this type of construction projects, with greater precision. Ultimately, the project is a step towards current digitization trends advocated by concepts such as cyber-physical systems, as well as Industry 4.0 and 5.0.

In this context, this dissertation has as the main objective the use of databases of real works resulting from the installation of sensors in transport equipment to study the feasibility of predicting their fuel consumption. This prediction is based on the implementation of ML models, in order to analyze of the mentioned databases according to the characterization of the routes covered by the sensorised trucks (e.g., number of light and steep climbs, number of descents), the average speed achieved by the trucks, the loads transported and the record of the actual consumption associated with each route, which corresponds to the "ground truth" on which the study is based.

The potential of ML applications in Transport Infrastructure and Geotechnics has been the target of a considerable focus in the last decade (Gomes Correia et al., 2012). Indeed, following the increasing development of remote monitoring and data storage technologies, the successful ML applications in this field cover different areas, such as earthmoving productivity (Parente et al., 2014; Hola and Schabowicz, 2010), the safety of slats (Tinoco et al., 2017) and compressive strength of jet grouting (Tinoco et al., 2011), management, maintenance and monitoring of pavements (Ma et al., 2019; Zhao et al., 2019). These can often address specific processes, such as estimating the compaction work rate (Marques et al., 2008), or the excavator cycle time (Edwards and Griffiths, 2000; Tam et al., 2002), as well as understanding the central part of larger and more complex systems, such as fleet management and allocation systems (Parente et al., 2016), or systems for concept and managing pavements (Ma et al., 2019; Zhao et al., 2019).

In the latter area, there has also been an increasing number of applications regarding the evaluation and maintenance of pavement conditions (Souza et al.,2018; Nunes and Mota, 2019; Majidifard et al., 2020). A notable aspect of these systems is related to the fact that they leverage concepts such as sensorization and digital twins to collect data, which in turn provide the training and testing database. In other words, the predictive models in these pavement management systems are trained with data from different sensors placed in the field, either in inspection vehicles (Souza et al.,2018; Nunes and Mota, 2019) or on the pavement itself (Ma et al., 2019; Zhao et al., 2019; Majidifard et al., 2020).

Nevertheless, there is less attention to the estimation of costs associated with construction processes. In particular, the estimation of fuel consumption using a predictive model of machine learning is an issue which, despite having had some developments in other areas, such as logistics and long-haul truck routes (Svard, 2014; Perrotta et al., 2017; Schoen et al., 2019), no equivalent effort was made in the field of construction. In large transport infrastructure projects, fuel is one of the main costs associated with the execution of construction processes and its data-based estimation is a very significant advantage for project managers and designers. Decisions regarding heavy machine allocation, task planning, or performance evaluation are typically made using fuel consumption estimates based on operators' professional experience or generic documents and guides, such as the CATERPILLAR Performance Handbook (2018). Thus, the availability of sensor-based information in real context allows decision-making teams to adequately manage their resources (both economic and physical, such as construction equipment) not only during the design and planning phases, but also during the construction phases themselves, where the allocation of resources can be adjusted due to unforeseeable constraints and occurrences (e.g. equipment malfunction or equipment malfunction by low productivity).

Most of the studies conducted on fuel consumption prediction are based on simulations through physical calculations and are generally slow to perform because they simulate the internal components of trucks (Sandberg, 2001; Nasser et al., 1998). An example of this is *the Scania Truck and Road Simulation* (STARS), which is a simulation system, that only predicts consumption after obtaining certain vehicle and driver configurations (Svard, 2014). One disadvantage of this type of prediction is that simulations require time and manual configuration. As more parameters are added, the longer the simulation turns out to be due to increased complexity.

There are many studies related to the prediction of consumption in the different types of transport through ML algorithms, serving as the main source of inspiration STARS. Due to the quality of the study, there are many projects derived from this. However, very few have been able to conclude whether there is the possibility of making the necessary prediction from the ML algorithms.

One study implemented a STARS-inspired model using random forest and gradient boosts (Viswanathan, 2013) and concluded that the former achieved better results as well as greater relevance of speed parameters, descents without consumption, distance with trailer, distance with cruise control and maximum speed. However, it only considered the parameters of the driver's behavior and did not include the characteristics of the road, the vehicle, or the weather conditions. It also did not address how to train the model for route choice or anomaly detection.

A study carried out by (Lindberg, 2014) tried to develop a predictive model using data about the position, amount of fuel and other characteristics inherent to the truck, to be sent with a given frequency. After several collections following a road between two different cities, when applying the different models, a definitive conclusion was not reached, and it was not clear which model would have the best results. However, the author was able to conclude that the weight of the vehicle and the slope of the road would be the most relevant variables for fuel consumption predictions. Due to the limitation of the data and its low sample rate, the published results show a low accuracy of 26%.

Another study (Svard, 2014) confirmed that it was possible to make this type of prediction and even outline that the weight of the vehicle load and the inclination would be the most determining factors in the associated calculations. Factors as driver behavior, slowing speed and other variables that influence consumption were not addressed. Different models were trained, such as decision trees, artificial neural networks, linear regression models and SVM. This last model turned out to be the most productive within this study, but the results may not be reliable, as different data was used to train the remaining models.

A similar project was carried out regarding the logistical issues (Almer, 2015). However, despite the very positive results with the use of ML models for the prediction of fuel consumption, with a wide variety of parameters, the objective would be a little different, both because of the nature of the process (long-distance goods transport trips) and the constraints of the process. In fact, road works include trips that are generally cyclical, on varied terrain (paved, unpaved, construction site paths), with varied weights and materials, and generally a much less controlled environment. Thus, this project also aims to add more knowledge to the state of the art, specifically for the area of construction and its equipment.

2. Methodology

The methodology adopted in the development of this dissertation began by defining the fundamental indicators to predict fuel consumption and how to affect it. Thus, the following sensors were chosen, IMU, to measure acceleration on the three axes and the slope, and a GNSS receiver, to measure the spatial position, to be part of a prototype that was installed in a transport vehicle (truck) to monitor and obtain information in the form of computer files, as described and schematized in Figure 1. The process continued with the gathering of automated data from sensors and other manual input collected by the operator (i.e., weight and real fuel consumption), followed by the consolidation process, the pre-processing and analysis of the data and the application of machine learning algorithms in the creation of the prediction models.

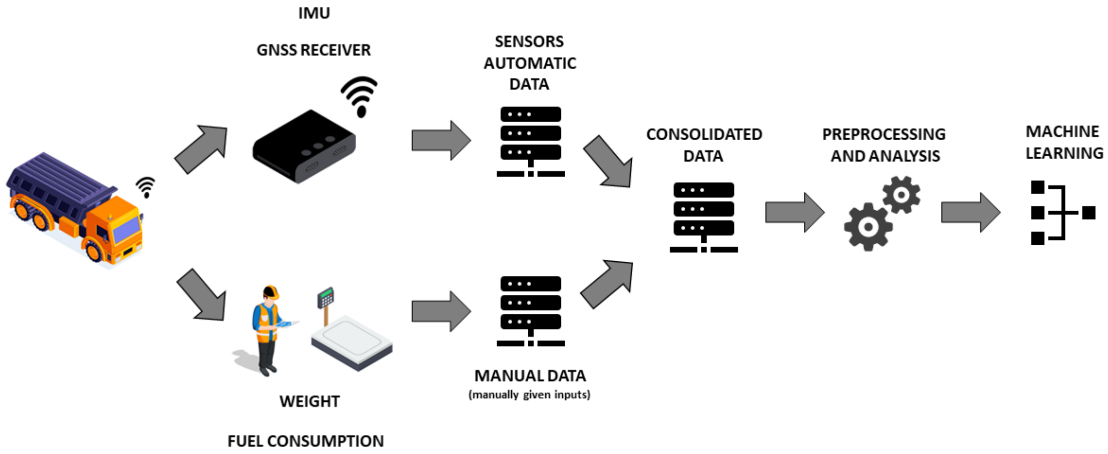

Figure 1 - Process of collection, consolidation and processing of field data and creation of prediction models

After information is provided, it would be essential to treat it and to withdraw results, which was carried out with the help of a programming software, R (R Development Core Team 2011). This software, consisting of packages that give it the ability to simulate ML methods, allowed to estimate the fuel consumption which was compared with the actual value observed, to validate and measure the robustness of the model and to draw conclusions regarding the cause-effect relationships of certain parameters with the consumption of the transport vehicle.

3. Results and Discussion

In this work, the implemented techniques focus on regression models. As such, their evaluation is primarily based on three main metrics, namely the mean absolute error, MAE (Equation 1), depicting the error associated with the degree of learning of a given model, the square root of the mean of squared errors, the RMSE (Equation 2), which penalizes higher error values, and the correlation coefficient, R^2 (Equation 3), which comprises the correlation between the observed values and the predicted values (Hastie et al., 2009).

$$MAE = \frac{\sum_{i=1}^{N}(y-\hat{y})}{N} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(y-\hat{y})^2}{N}} \tag{2}$$

$$R^2 = \left(\frac{\sum_{i=1}^{N}(y-\bar{y}) \times (\hat{y}-\bar{\hat{y}})}{\sqrt{\sum_{i=1}^{N}(y-\bar{y})^2 \times \sum_{i=1}^{N}(\hat{y}-\bar{\hat{y}})^2}}\right)^2 \tag{3}$$

Where:

y - is the computed network output vector,

$\hat{y}$ - is the target output vector, and

N - is the number of samples in the database.

Additionally, the Regression Error Characteristic Curve (REC) (Bi and Bennett, 2003) was also adopted as a measure of the cumulative error distribution function of different regression algorithms, allowing a comparative analysis between the latter. As previously mentioned, the data is the foundation of ML models. However, the fact that these predictive algorithms depend so heavily on the availability of data represents both their greatest potential and limitation. Indeed, the predictive capacity of an ML model depends both on the quantity and quality of the available data. In other words, to achieve a perfect prediction of the behavior of a target variable, the model must not only be fed with sufficient data, but this data should also cover all independent variables that may have some level of influence on that behavior. Moreover, data variability should be consistent enough to allow ML algorithms to gather knowledge and insight on how much each independent variable affects the target variable, as well as how the independent variables connect and affect each other.

In this context, one can easily find several sets of pre-processed data associated with different fields which can result in almost perfect models (i.e., with near 100% prediction accuracy in the case of classification models, or very low error margins in the case of regression models), depending on the ML technique adopted, as is the case of the well-known VGG-16 (Simonyan and Zisserman, 2015) or the ResNet50 databases (He et al., 2016) in the image classification field. However, this level of accuracy/error is not always attainable when it comes to real-world data, as it tends to be noisy and affected by higher levels of unreliability or missing data. While there is a number of actions that can be taken to deal with these issues during data cleaning and processing phase (which will be discussed in the ensuing subsection), typical real-world based models can only very rarely achieve perfect fits to the data, and subsequently perfect predictive capabilities.

Bearing these considerations in mind, the data used in this work result from a real scenario setting of a road construction site in Portugal. The raw data is a consequence of the sensorised activities of a vehicle transporting road paving materials, featuring around 21 trips through different routes and pavement surfaces (such as construction site paths, national road and motorway), while transporting different materials, among which the most common was bituminous mix from an asphalt plant where it is manufactured for a road construction site. Besides cargo and pavement surfaces, the trips also show some variation regarding the total distance, which varies between 20 and 60km. The main parameters measured include time, location/GPS data, altitude, speed and 3-axle inclination of the truck, as exemplified in Table 1.

Table 1 - Values extracted from the raw database

Inclination _X	Inclination _Y	Inclination _Z	Latitude	Longitude	Altitude	Speed	Clock
0.778198	-0.29755	-1.31226	39.4447	-7.47812	447	0.043	2021-07-22T12:42:36.100Z
0.778198	-0.29755	-1.31226	39.4447	-7.47812	447	0.038	2021-07-22T12:42:36.400Z
0.778198	-0.29755	-1.31226	39.4447	-7.47812	447	0.038	2021-07-22T12:42:36.400Z
0.839233	-0.30518	-0.03052	39.4447	-7.47812	447	0.038	2021-07-22T12:42:36.700Z
0.839233	-0.30518	-0.03052	39.4447	-7.47812	447	0.013	2021-07-22T12:42:36.800Z
0.839233	-0.30518	-0.03052	39.4447	-7.47812	447	0.014	2021-07-22T12:42:37.100Z
0.839233	-0.30518	-0.03052	39.4447	-7.47812	447	0.044	2021-07-22T12:42:37.500Z
0.923157	-0.28992	-1.95313	39.4447	-7.47812	447	0.004	2021-07-22T12:42:37.900Z
0.923157	-0.28992	-1.95313	39.4447	-7.47812	447	0.036	2021-07-22T12:42:38.200Z

As can be easily inferred, the refresh rate associated with the IoT framework collects data several times per second, ultimately generating very large databases in the form of .csv files, each corresponding to a trip from the truck. Since these trips can be as long as nearly 70 kilometers, the associated data files also increase exponentially to around 12,500 entries in those cases. At this stage in the project, the database contains the afore mentioned 21 routes for this truck, carrying different types (and weights) of cargo.

Figure 2 depicts one of the most typical altimetric profiles through which the truck has travelled. These were extrapolated by integrating the speed data over time and validated by resorting to the inclinometer data. From this point on, a sliding window methodology was adopted to obtain sections every 10 meters to adjust the lines along the altimetric profile, allowing the determination of the slope of each 10-meter section. Through this methodology, each trip was translated into the accumulated distance which the truck travels in each type of slope, according to the considerations described in Table 2.

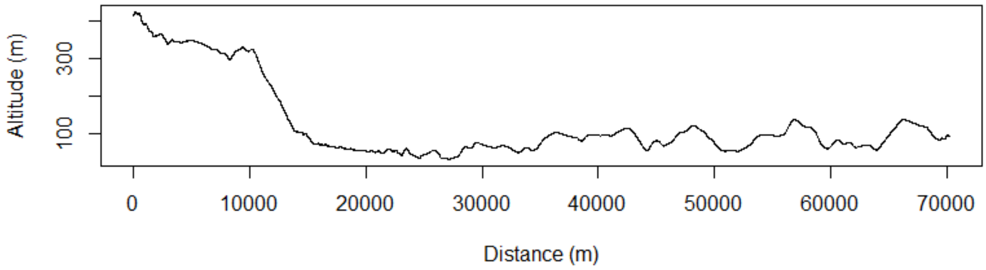

Figure 2 – Example of an altimetric profile of a trip as measured by the sensorized truck

Table 2 - Considered slope ranges and description

Slope description	**Range**	**Feature description**
Flat surface	-1% < Slope ≤+1%	AD_0.01n_0.01
Light upwards slope	+1% < Slope ≤+5%	AD_0.01_0.05
Moderate upwards slope	+5% < Slope ≤+10%	AD_0.05_0.1
Steep upwards slope	Slope ≥ +10%	AD_0.1
Light downwards slope	-5% < Slope ≤ -1%	AD_0.01_0.05n
Moderate downwards slope	Slope ≤ -5%	AD_0.05n

Ultimately, this conversion consisted of determining the percentage of the total route distance that the truck passes on each type of slope throughout each trip, resulting in a large part of the training and test database. The latter are complemented with data on the total distance of the trip (TDistance, meters), average speed (AvSp, meters per second), load weight (Load, tons) and fuel consumption (FConsumption, liters, target variable). The last two characteristics are manually entered into the database, as they originate from the manual records made by the truck driver between trips and during each refuelling action. Table 3 depicts the processed database that supports the final ML algorithms.

Table 3 – Values extracted from the training and testing database

AD_0.01n_ 0.01	AD_0.01_0.05	AD_0.05_0.1	AD_0.1	AD_0.01_0.05n	AD_0.05n	TDistance	AvSp	Cargo	FConsumption
0.32	0.25	0.01	0	0.36	0.06	66341.91	18.48	29.48	30
0.40	0.33	0.05	0	0.21	0.02	65704.09	20.52	0	23
0.20	0.33	0.13	0.01	0.23	0.11	52992.28	12.39	0	20.5
0.20	0.25	0.08	0.01	0.28	0.17	35887.31	9.15	33	25.5
0.19	0.25	0.10	0.01	0.28	0.17	36563.17	8.52	32.68	25
0.37	0.23	0.01	0	0.35	0.05	62786.77	17.40	33.76	26.5
0.43	0.30	0.05	0	0.21	0.01	62282.68	20.79	0	23.5
0.38	0.22	0.01	0	0.33	0.05	62786.77	18.91	32.46	26
0.41	0.31	0.05	0	0.22	0.01	62606.6	19.16	0	23

The results were obtained using the rminer package (Cortez, 2010) for R. With the use of this tool, several models in the processed database were trained and tested, namely RF, ANN and SVM. Since generalization capacity is a key concern for future implementation, as well as for the assessment of models, a 10-run under cross-validation approach was adopted. The relatively low amount of data justifies the k-fold value of 3. This means that the data is evaluated across the entire training set by dividing it into 3 folds. The model is then trained 3 times while reserving a different fold as a testing dataset each time, thus using the data available to its full potential (Hastie et al.,2009). Figure 3 shows a comparison between the REC curves of the three models. The analysis of this figure suggest that RF seems to have a slightly higher performance than the other two models, which is consistent with the results obtained in the literature concerning the prediction of heavy vehicles fuel consumption (Perrotta et al., 2017).

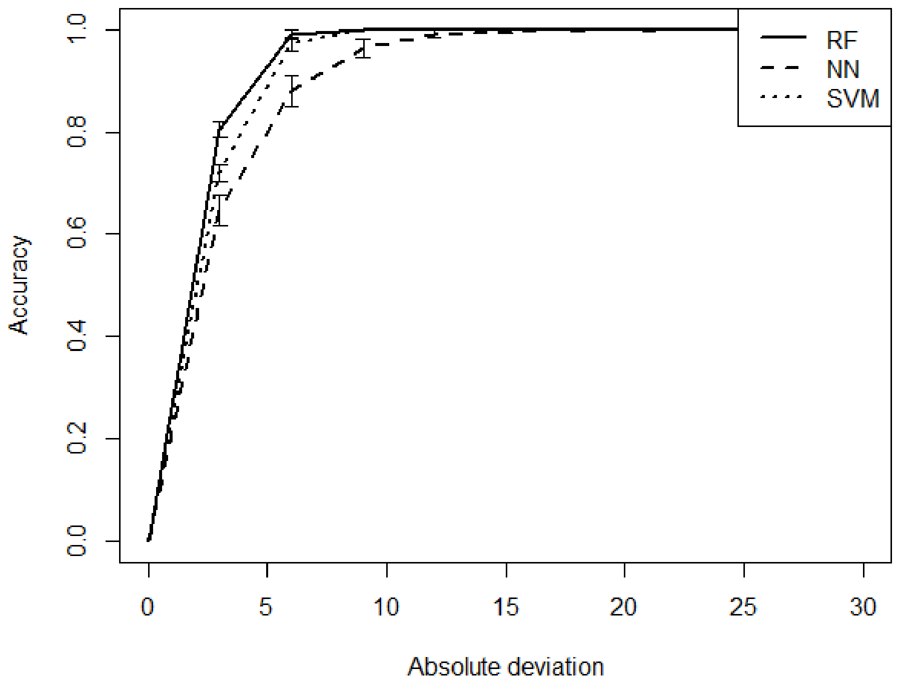

Figure 3 - Graphical representation of the result of the various models

In fact, the RF model exhibits a slightly lower error, with an MAE of 1.723915 (corresponding to a percentage error of about 7.6%), a RMSE of 2.127431 and a slightly better R^2 for the data, with a value of 61.4%. Although this value is expected to improve as more data becomes available, it is still considered a reasonable value considering that the model is based entirely on real-world data. This hypothesis is supported by the value of the Pearson's correlation coefficient, R, which is at 78.4% for the RF model, denoting a very good interdependence between the independent variables and the target variable.

Figure 4 shows how the predicted values fit the observed values in the testing dataset. As can be easily inferred, the closer the points are to the diagonal line, the better the fit, and consequently the better the value of R^2. Value analysis shows that the model seems to be able to reproduce the behavior of the target variable especially well in the values of the average range (i.e. the fuel consumption values of 20 to 27 litres), though the values closest to the lower (i.e. around 17 litres), and higher ends (i.e. around 30 litres) are being slightly overestimated and underestimated, respectively. As expected, this is due the fact that there is a much lower amount of records in the database that fall within these extremity ranges, which is also expected to improve with the database increment over time.

Another noteworthy aspect for analysis is the relative importance of the variables for the RF model, depicted in Figure 5. The figure conveys the increase of the mean square error, %IncMSE, as a result of the corresponding variables being permuted. In other words, the higher the value of %IncMSE, the more important the corresponding variable is for the predictive capability of the model. In this context, it is interesting to observe that the weight of the cargo transported by the truck is considered by the model as its most important variable, followed by the predominance of slight and moderate upwards slopes, as well as moderate downward slopes. Intuitively, it is easy to infer how valid these aspects are, since the higher the weight and the upward slopes, the higher the fuel consumption, while the downward slopes imply a nearly null value of fuel consumption. On the other hand, the variable corresponding to steep climbs is considered the least relevant for the model, being easily explained by the low occurrence of sections with this type of inclination along the vehicle route, and, as such, the model identifies the associated variable as having a very low importance. However, the removal of this variable from the training data set has a negative impact on the predictive capacity of the final model, even if to a low scale, and as such, was maintained in the training dataset.

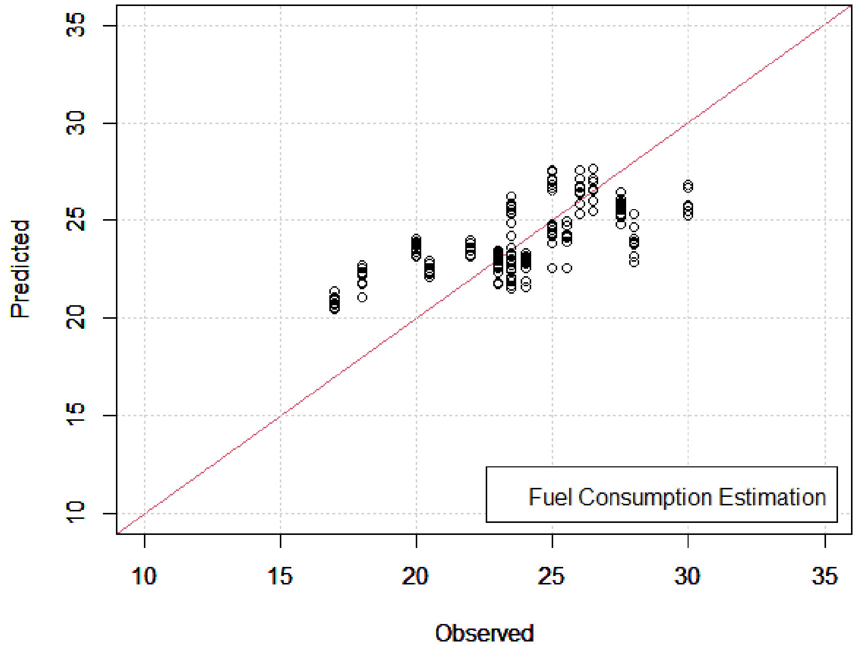

Figure 4 - Graphical representation of the result of the predicted data vs. observed

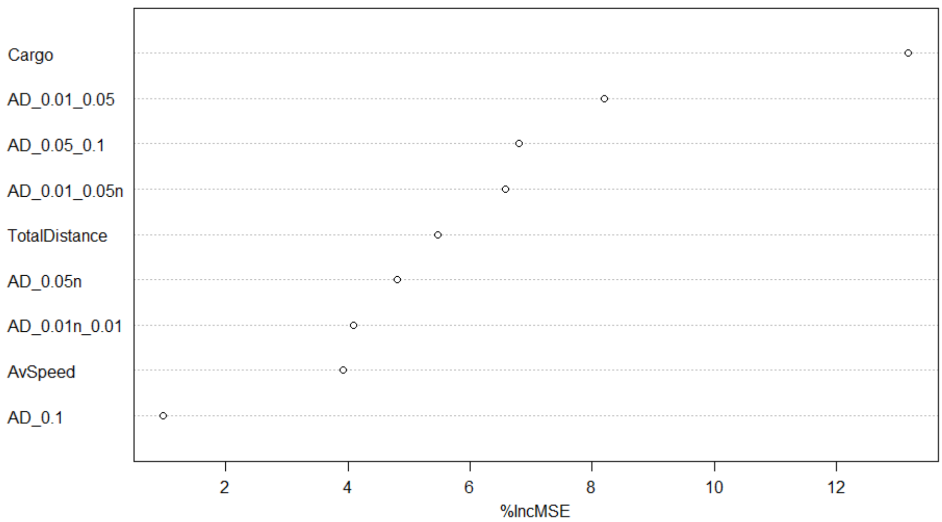

Figure 5 - Graphic representation of the relevance of the variables to the model

4. Conclusions

This dissertation presented a model for predicting fuel consumption applied to transport trucks, based on a real scenario, with the objective of being used as a tool for project planning and management and budget analysis. The innovation of this work was the integration of a IoT framework for the collection and transmission of information in the database, which comprises the training and test data for the predictive models. The results showed that fuel consumption can have a strong correlation with cargo, longitudinal inclination of the routes and total distance, thus proving to be key input parameters to achieve more accurate and reliable fuel consumption predictions. These results are particularly interesting to engineers with management, planning and decision-making functions, as this information is easily accessible to them through existing GIS systems (e.g. route inclination and total distance), and the construction project plans or BIM models themselves (e.g., cargo).

Although it is notorious that the amount of data needed at this stage was scarce to ML models to be considered generalizable and thus be implemented in practice, it is considered that the result of this dissertation has potential, and the results were promising. Above all, the methodology used was a relevant contribution to the state of knowledge, in the sense that it provides an initial step towards the real-world implementation of a digital twin, as well as a self-learning ML system in an IoT framework, thus following current trends in automation, digitalization and Industry/Construction 4.0.

One of the limitations that can be pointed to the developed model is that is required to the user to estimate the average speed over the route, which can comprise a significant obstacle. However, this issue can potentially be mitigated by the introduction of data stemming from accelerometers. As a matter of fact, leveraging on the vertical axis of the accelerometers to infer a rough classification of each type of surface through which the truck circulates (e.g., unpaved road, national road, highway) can provide insight on the behaviour of the truck in different environments (e.g., average speed, average number of full stops, traffic conditions, among others). Subsequently, this type of information may even be valuable enough to the model for it to eventually replace the need to estimate speed by the user, who instead may only have to estimate the percentage of each surface type in relation to the trip's total distance, similarly to the road slope features already present in the model. On the other hand, the reduced amount of data, due to the time window limit to collect them, as well as their quality, affected by failures in the collection of sensor data or human failures filling the values observed, limited the size and quality of the database that served as training and testing to the ML algorithm.

In addition to what has been mentioned above, future work should naturally include the expansion of the study to encompass a higher amount of vehicles, routes and carried cargos, in order to produce a more robust and generalizable prediction model. In addition, the results achieved motivate the development of a real-time sensor acquisition system capable of dealing with current sampling frequency bottlenecks, thus supporting a continuous and automatic process of training and testing of forecast models, improving their accuracy and reliability by increasing the amount of information obtained from sensors.

In addition, a number of other resources could have been explored, such as oversampling and undersampling, for the introduction of a bias, in order to select more samples from one class than from another, in order to compensate for an imbalance that may be present or likely to develop in the data of a purely random sample, as well as the use of other ML algorithms that obtained interesting results in the bibliography such as Gradient Boost Machines*,* among others. It will also be interesting to explore, with a more extensive database, the manual selection of training samples as opposed to what was done, automatically by the algorithm, which in this case would become irrelevant due to the small number of existing samples.

References

Almér, H. (2015), "Machine learning and statistical analysis in fuel consumption prediction for heavy vehicles", Master's thesis, KTH, School of Computer Science and Communications (CSC), Stockholm.

Bi, J., & Bennett, K.P. (2003). Regression error characteristic curves. In Proceedings of the Twentieth International Conference on Machine Learning, Washington, DC, USA, 21–24 August 2003; pp. 43–50.

Caterpillar (2018). Caterpillar Performance Handbook. Peoria, Illinois, U.S.A.

Cortez, P. (2010). "Data Mining with Neural Networks and Support Vector Machines using the R/miner Tool." Advances in Data Mining - Applications and Theoretical Aspects 10th Industrial Conference on Data Mining (ICDM 2010), P. Perner, ed., Springer, Berlin, Germany, 571–583.

Edwards, D. J., & Griffiths, I. J. (2000). "Artificial intelligence approach to calculation of hydraulic excavator cycle time and output," *Min. Technol.* 109 pp. 23–29.

Gomes Correia, A., Cortez, P., Tinoco J., & Marques R. (2012), "Artificial Intelligence Applications in Transportation Geotechnics," *Geotech. Geol. Eng.* 31, 861–879

Hastie, T., Tibshirani, R., e Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York.

He, K., Zhang, X., Ren, S., & Sun J. (2016). Deep residual learning for image recognition, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2016-December 770–778. https://doi.org/10.1109/CVPR.2016.90.

Hola, B., & Schabowicz, K. (2010). "Estimation of earthworks execution time cost by means of artificial neural networks," *Autom. Constr.* 19 570–579.

Lindberg, J. (2014). Fuel consumption prediction for heavy vehicles using machine learning on log data. Master's thesis, KTH, School of Computer Science and Communications (CSC), Stockholm.

Ma, X., Dong, Z., Chen, F., Xiang, H., Cao C., & Sun J. (2019). "Airport asphalt pavement health monitoring system for mechanical model updating and distress evaluation under realistic random aircraft loads", Constr. Build. Mater. 226 227–237. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2019.07.174

Majidifard, H., Adu-Gyamfi Y., & Buttlar W.G. (2020). Deep machine learning approach to develop a new asphalt pavement condition index, Constr. Build. Mater. 247 118513. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2020.118513.

Nasser, S. H., WeiBermel, V., & Wiek, J. (1998). Computer simulation of vehicle's performance and fuel consumption under steady and dynamic driving conditions.

Nunes, D.E., & Mota, V.F.S. (2019). A participatory sensing framework to classify road surface quality, J. Internet Serv. Appl. 10 13. https://doi.org/10.1186/s13174-019-0111-1.

Parente, M., Gomes Correia, A., & Cortez, P. (2014). "Use of Data Mining Techniques in Earthwork Management: A Case Study", Earthwork Project Management, Slope Stability Analysis, and Wave-Based Testing Techniques: pp. 1-8. American Society of Civil Engineers, Geotechnical Special Publication (GSP) 252 (2014). http://dx.doi.org/10.1061/9780784478523.001.

Parente, M., Gomes Correia, A., & Cortez, P. (2016). "Metaheuristics, data mining and geographic information systems for earthworks equipment allocation", Advances in Transportation Geotechnics III 143 506-513. 10.1016/j.proeng.2016.06.064.

Perrotta, F., Parry T., & Neves, L. C. (2017). "Application of machine learning for fuel consumption modelling of trucks," *2017 IEEE International Conference on Big Data (Big Data)* 3810-3815. 10.1109/BigData.2017.8258382.

Sandberg, T. (2001). Heavy truck modeling for fuel consumption simulations and measurements. Master's thesis, Linköping University.

Schoen, A., Byerly, A., Hendrix, B., Bagwe, R. M., Santos E. C. d., & Miled, Z. B., "A Machine Learning Model for Average Fuel Consumption in Heavy Vehicles," in IEEE Transactions on Vehicular Technology, vol. 68 (2019) 6343-6351. 10.1109/TVT.2019.2916299.

Simonyan, K. & Zisserman A. (2015). Very deep convolutional networks for large-scale image recognition, 3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc. 1–14.

Souza, V.M.A., Giusti, R., & Batista, A.J.L. (2018). Asfault: A low-cost system to evaluate pavement conditions in real-time using smartphones and machine learning, Pervasive Mob. Comput. 51 121–137. https://doi.org/https://doi.org/10.1016/j.pmcj.2018.10.008.

Svärd, C. (2014). Predictive modelling of fuel consumption using machine learning techniques. Technical report, Scania CV AB.

Tam, C. M., Tong, T., & Tse, S. (2002). "Artificial neural networks model for predicting excavator productivity," *J. Eng. Constr. Archit. Manag.* 9 446–452

Tinoco J., Gomes Correia, A. & Cortez, P. (2011). "Application of data mining techniques in the estimation of the uniaxial compressive strength of jet grouting columns over time". Constr Build Mater 25 1257–1262. 10.1016/j.conbuildmat.2010.09.027

Tinoco, J., Gomes Correia, A., Cortez, P., & Toll, D. (2017). "Stability condition identification of rock and soil cutting slopes based on soft computing", Journal of computing in civil engineering 32 04017088.

Viswanathan, A. (2013). Data driven analysis of usage and driving parameters that affect fuel consumption of heavy vehicles. Master's thesis, Linköping University, Statistics, The Institute of Technology

Zhao, H., Wu, D., Zeng, M., Tian, Y., & Ling J. (2019). "Assessment of concrete pavement support conditions using distributed optical vibration sensing fiber and a neural network", Constr. Build. Mater. 216 214–226. https://doi.org/https://doi.org/10.1016/j.conbuildmat.2019.04.195