Prediction performance of separate collection of packaging waste yields using Support Vector Machines

V. Sousa[1], I. Meireles[2], V. Oliveira[3,4], C. Dias-Ferreira[3,4]

[1]Department of Civil Engineering, Architecture and GeoResources, IST – University of Lisbon, Av. Rovisco Pais 1, 1049-001, Lisbon, Portugal

[2]Department of Civil Engineering, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

[3]Research Centre for Natural Resources, Environment and Society (CERNAS), College of Agriculture, Polytechnic Institute of Coimbra, Bencanta 3045-601 Coimbra, Portugal

[4]Materials and Ceramic Engineering Department, CICECO, University of Aveiro, Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

Keywords: household packaging, multiple linear regression, separate collection, support vector machines, waste collection.

Presenting author email: vitor.sousa@tecnico.ulisboa.pt

Understanding the drivers underlying waste segregation of glass, paper/cardboard, plastic/metal and bio-waste is a paramount to attain the targets set by the European Commission and transposed to the legislation of the member states. Within this context, the new Circular Economy Package set new waste management targets for 2030: i) reduction of landfilling to a maximum of 10% of the total municipal waste; ii) increase re-use and recycling of municipal waste to 65%, iii) increase recycling of packaging waste to 75% and iv) ban landfilling of separately collected waste (European Parliamentary Research Service, 2016). A better understanding of factors influencing the performance of source separated waste collection will aid waste management utilities in adjusting and optimizing the service to meet these targets.

Oliveira et al. (2017) modelled the separate collection of packaging waste yields in the coastal area of the "Centro" region of Portugal using ordinary least squares (OLS) multiple linear (OLS-L regression) and non-linear regression (OLS-NL regression) approaches. The models were built starting from a set of 14 possible explanatory variables, which were reduced to 5 statistically significant (i) inhabitants per bring-bank; ii) relative accessibility to bring-banks; iii) degree of urbanisation; iv) area and; v) number of school years attended). The OLS-NL regression model had the best performance, being capable to explain about 73% of variability on source separated collection yields data.

The results obtained reveal the existence of a non-linear relation between the explanatory variables and the separate waste collection yields as well as the existence of interaction between the explanatory variables. However, exploring possible interaction between the explanatory variables, even for a reduced set, results in large number of possible combinations. In these cases, tools from the artificial intelligence have revealed useful by enabling this task to be conducted using numerical algorithms.

In the present communication, an artificial intelligence tool such as support vector machines (SVM) were used to develop models to predict waste segregation yields. Then, the performance of this tool is compared with that of OLS-NL regression on the case study of the coastal area of the "Centro" region of Portugal, which comprise 42 municipalities (over 1 million of inhabitants). A SVM constructs a hyperplane or set of hyperplanes in a high- or infinite- dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data points of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier. Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space would be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load reasonable, the mapping used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function selected to suit the problem. Thus, two SVM models were built: i) with the same variables selected for OLS-NL regression (SVM-Reg) and; ii) using the genetic algorithm to select the best set of variables (SVM-GA).

The performance of SVM models for predict separate collection of packaging waste is shown in Fig. 1. Both SVM models reveal a superior performance when compared with the OLS-NL regression model develop in the past effort (Oliveira et al. 2017). Using the same explanatory variables of the OLS-NL regression model (SVM-Reg) or the set of variables selected using a genetic algorithm optimization (SVM-GA) yield the same performance of the SVM models (R^2=0.983).

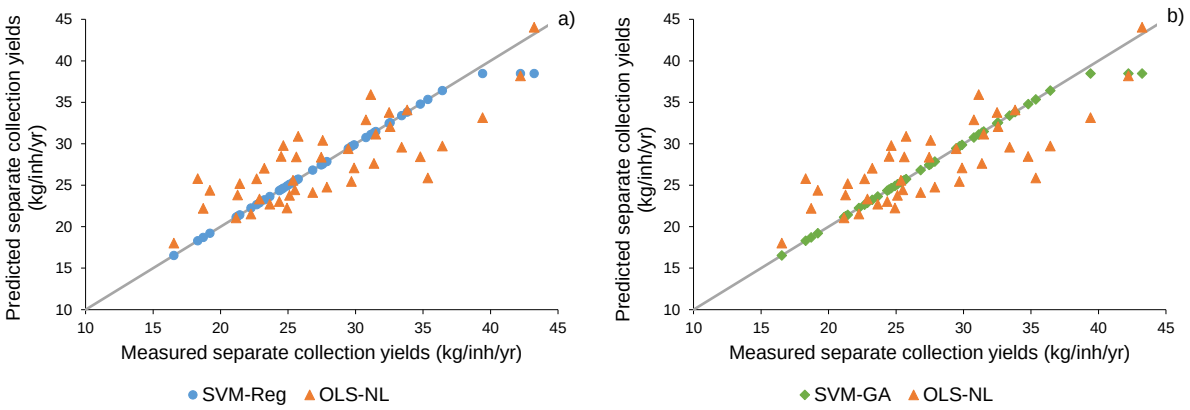

Figure 1. Comparison of predicted separate collection yields between: a) OLS-NL regression and SVM-Reg models and ii) OLS-NL regression and SVM-GA models

The performance of SVM models were almost 26% higher than the best non-linear regression model which indicates that this artificial intelligence tool could be a viable alternative for prediction of separate collection of packaging waste aiding waste management organisations in the definition of strategies and planning.

Acknowledgments

C. Dias-Ferreira and V. Oliveira gratefully acknowledge FCT – Fundação para a Ciência e para a Tecnologia (SFRH/BPD/100717/2014; SFRH/BD/115312/2016) for financial support, and project Life PAYT – "Tool to Reduce Waste in South Europe" (LIFE15 ENV/PT/000609).

References

Oliveira, V., Sousa, V., Vaz, J.M., Dias-Ferreira, C., 2017 (*in press*). Model for the separate collection of packaging waste in Portuguese low-performing recycling regions. Journal of Environmental Management. https://doi.org/10.1016/j.jenvman.2017.04.065

European Parliamentary Research Service, 2016. Closing the loop - new circular economy package: Briefing.