PREDICTION MODELS FOR IMPROVING WASTE DECISION SUPPORT MANAGEMENT IN ROMANIA IN ASSOCIATION WITH V4 MEMBER COUNTRIES

Stefan-Mihai PETREA^{1, 2, 3}, Ira-Adeline SIMIONOV^{1, 2, 4}, Alina ANTACHE^{1, 2, 6}, Aurelia NICA², Cristina ANTOHI³, Dragos Sebastian CRISTEA³, Adrian ROȘU^{1, 5}, Valentina CALMUC^{1, 5}, Bogdan ROȘU⁴

 ¹"Dunarea de Jos" University of Galati, REXDAN Research Infrastructure, 98 George Cosbuc Street, Galati, Romania
 ²"Dunarea de Jos" University of Galati, Faculty of Food Science and Engineering, 11 Domneasca Street, Galati, Romania
 ³"Dunarea de Jos" University of Galati, Faculty of Economics and Business Administration, 59-61 Nicolae Balcescu Street, Galati, Romania
 ⁴"Dunarea de Jos" University of Galati, Faculty of Automation, Computers, Electrical Engineering and Electronics, 47 Domneasca Street, Galati, Romania
 ⁵"Dunarea de Jos" University of Galati Faculty of Sciences and Environment, Department of Chemistry, Physics and Environment, 47 Domneasca Street, Galati, Romania

Corresponding author email: ira.simionov@gmail.com

Abstract

The present study results are based on the application of XGBoost machine-learning algorithms and indicate that total waste, as a dependent parameter, can be accurately evaluated considering plastic wastes (feature importance-FI = 1.53, Rsq. = 0.75, RMSE = 0.47) in the case of V4 group, while for Romania, the dependent parameters identified as most reliable are chemical wastes (FI = 0.58) and industrial effluent sludges (FI = 0.04), with lower accuracy metrics (Rsq. = 0.46, RMSE = 0.75). In terms of waste treatment (WT), the portable batteries and accumulators' market (FI=0.45) presents high reliability to be used as the main predictor (Rsq. = 0.80, RMSE=0.42) for V4 support tool, while for Romania, the waste generation (FI = 1.57, Rsq.= 0.85, RMSE=0.36) highly explains the variability of WT. However, batteries and accumulators waste (FI = 0.77, Rsq.= 0.82, RMSE=0.39) can be used as a reliable predictor for WT variation in a more extended analytical framework, in the case of Romania. It can be concluded that waste decision support management can be supported based on ML models which are different in the case of Romania, compared to V4, emphasizing the regional importance when developing environmental modeling-based tools.

Key words: prediction models, waste treatment, waste decision support models, XGBoost, waste framework.

INTRODUCTION

The expansion of waste footprint is both environmentally and economically unsustainable. As a response, the European Union (EU) has set a Waste Framework Directive which regulates a series of concepts related to waste management, considering risks linked to human health, water – air – plants – animals pollution, as well as noise and odour pollution. In order to improve the capacity to respond to different social, economic and environmental challenges, regional interest groups have been created, such as the Visegrad Group (V4). A demonstrated increase in efficiency associated with already establish regional groups could encourage their extension and could optimize the environmental strategies, considering the regional specificity.

Waste management is a critical issue in the EU, and the EU has implemented several environmental policies to manage waste in a sustainable and environmentally friendly way. These policies aim to promote a circular economy, reduce waste, and improve resource efficiency (*Single-Use Plastics Directive -European Bioplastics e.V.*, n.d.). Some of the key European directives related to waste management in the EU are, as follows: Waste Framework Directive (WFD) (Implementation of the Waste Framework Directive, n.d.), Packaging and Packaging Waste Directive (PPWD) (Packaging Waste, n.d.), Waste Electrical and Electronic Equipment Directive (WEEE) (EUR-Lex - 02012L0019-20180704 -EN - EUR-Lex, n.d.), Landfill Directive (LD) (Landfill Waste, n.d.)and Circular Economy Package (CEP) (Circular Economy Policy -Library, n.d.). The WFD establishes the basic concepts and definitions related to waste management and lays out a hierarchy of waste management options that prioritize prevention. reuse, and recycling over landfilling and incineration. The directive requires member states to take measures to reduce waste generation and to ensure that waste is managed without endangering human health or harming the environment. The PPWD aims to reduce the environmental impact of packaging and packaging waste by setting targets for the recovery and recycling of packaging waste. Member states are required to establish systems for the collection and recovery of packaging waste and to ensure that a certain percentage of the waste is recycled. The WEEE establishes rules for the collection, treatment, and disposal, which include electronic devices such as computers, televisions, and refrigerators. The directive requires member states to establish collection systems and to ensure that the waste is treated and disposed of in an environmentally sound manner. The LD aims to reduce the amount of biodegradable waste that is sent to landfills, which are a significant source of methane emissions. The directive establishes strict standards for the operation and closure of landfills and requires member states to reduce the amount of biodegradable waste that is landfilled. The CEP consists of several legislative proposals aimed at promoting a circular economy in the EU. The package includes new initiatives such as the Circular Economy Action Plan and the EU Strategy for Plastics in a Circular Economy. The package aims to increase the recycling and reuse of waste, reduce waste generation, and promote resource efficiency. Overall, the EU's waste management policies aim to reduce waste, promote recycling and reuse, and protect the environment and human health (Virsta, 2020). The directives discussed above are some of the

key tools the EU is using to achieve these goals.

In the Visegrad Group, waste management policies vary among the member states. Poland and Slovakia, for instance, have been criticized for their lack of progress in waste reduction and recycling compared to other EU member states (Mišík, 2019). Hungary has implemented measures to promote recycling and improve waste treatment facilities (Buczkó, 2018), while the Czech Republic has focused on waste reduction and reuse (Fialová, 2019).

The EU's waste management policies continue to evolve, with a focus on reducing waste generation and promoting a circular economy. Romania and the Visegrad Group are subject to these policies, and while progress has been made, some member states, particularly Poland and Slovakia, face criticism for their lack of progress in waste reduction and recycling.

Therefore, the aim of the present study is to develop a series of prediction models which target improving waste decision support management, considering the specificity of Romania, in association with V4 member countries.

MATERIALS AND METHODS

Bibliographic analysis

The analysis was performed using R code -Bibliometric and Scientometric Python Library, considering the WoS database and the following dimensions, within WoS database: waste management modeling. waste machine learning, management waste neural networks. management waste management deep learning, waste management Romania, waste management Visegrad, waste treatment modeling, households waste modeling and recycling modeling. Thus, during the 2008-2022 timespan, a number of 101 research documents (64 articles, 1 book and 36 proceedings papers) were considered and analysed, with an annual growth rate of 8.78%, counting 285 authors with an international coauthorship percentage of 9.90%.

The scientific papers publishing dynamics (Figure 1) reveal a considerable increase in the last 4 years of the analysed period, emphasizing the upward interest trend related to the analysed scientific areas.



Figure 1. Annual scientific production dynamics considering the keywords-based, already established, analytical framework

Database descriptiton

Data provided by Eurostat, related to a number of 14 indicators, between the years 2004-2020, were taken into consideration for the analytical framework: used oils wastes (tons) OUW; chemical wastes (tons) CW; industrial effluent sludges (tons) IES; health care and biological wastes (tons) HCBW; plastic wastes (tons) PW; textile wastes (tons) TW; batteries and accumulators wastes (tons) BAW; household and similar wastes (tons) HSW; total waste (tons) TW; waste treatment (kg per capita) WT; waste generated (kg per capita) WG; recycling rates of packaging waste RRPW; portable batteries and accumulators market (tons) PBAM; wastes collected - Portable batteries and accumulators (tons) WCPBA.

The dependent variables were considered as being TW and WT. However, for TW, the OUW, CW, IES, HCBW, PW, BAW and HSW are considered predictors, while for WT, two dimensions were considered as follows: D1 (including WG, RRPW, PBAM, WCPBA as predictors) and D2 (including OUW, CW, IES, HCBW, PW, TW, BAW, HSW as predictors).

Machine learning algorithms

For creating an in-depth machine learningbased analytical framework, firstly, generalized additive models (GAM) algorithms were used. The generalized additive models (GAM) extend the linear model with nonlinear functions of each variable. Therefore, for GAM, the linear component $b_i X_i$ is replaced with a non-linear $f_j(X_j)$ function associated with feature j, according to eq. 1. The functions are calculated for each predictor having the contributions added to the result. Based on dispersion diagrams and by using cubic spline functions, the f_i functions are evaluated through interpolation (eq. 2).

$$g^{-1}(E[Y]) = a + f_1(X_1) + f_2(X_2) + \dots + f_n(X_n) + \varepsilon$$
(1)

where:

 ε - minimized error;

E[Y] - the arithmetic mean of the dependent variable *Y*;

a - point of origin of the trajectory;

 g^{-1} is the inverse of the function g, called the link function;

 $f_1(X_1),..., f_n(X_n)$ - spline functions associated with independent variables.

$$S(x) = a_i + b_i x + c_i x^2 + d_i x^3, for \ \forall x \in [x_{i-1}, x_i]$$
(2)

where:

$$S: [a, b] \to \mathbb{R};$$

$$f: [a, b] \to \mathbb{R};$$

$$(x_i) = S(x_i), \ i = \overline{0, n}.$$

According to underlying patterns in the data, generalized additive models can provide an understanding of the impact of the predictive variables through linear or non-linear smooth functions. Some of the best reasons for using GAM in predictive problems (Murphy et al., 2019) are. as follows: interpretability, regularization and flexibility/automation. Thus, GAM models provide a good balance between the interpretable linear model and the "black box" learning algorithms. For additive models, the marginal impact interpretation of a single variable is not related to the values of the other variables. As a consequence, the model can provide various insights into the effects of the predictive variables.

For a GAM, it is possible to control the smoothness of the predictor functions, avoiding too many inflexion points, by adjusting the level of smoothness (Khamma et al., 2020). Also, even if the dataset contains noisy relationships, a prior belief that predictive

relationships are inherently smooth in nature can be imposed.

The XGBoost algorithm was applied in order to create prediction models based on a strong classifier, using data from weak classifiers (accuracy under 20-30%). The applied data analysis framework is presented in Figure 2.

Sampled data is used as input, based on a subset of features randomly selected. The XGBoost - extreme gradient boosting uses boosting for merging and combining several models with low prediction accuracy in order to generate a core model with high accuracy. The algorithm target to set target outcomes for the next models for minimizing the errors. Among the unique features of XGBoost it can be mentioned the following ones: regularization, handling spare data, weighted quantile sketch, block structure for parallel learning, cache awareness and out-of-core computing. The application of the algorithm provides several advantages. as follows: high accuracy. scalability, efficiency, flexibility, regularization, interpretability and open source.



Figure 2. Data analysis framework

Also, since the database is structured, the XGBoost is considered to perform better, compared to the random forest tree-based algorithm (RF). Therefore, the XGBoost represent an important background which can be used in correlation with GAM in order to develop a predictable analytical framework which can be used, further on as bases for the development of decision support systems.

RESULTS AND DISCUSSIONS

Bibliographic analysis

The bibliometric analysis revealed that the top journals which had published articles framed into the search profile described in the material and methods section are Environmental Engineering and Management Journal (15 publications) and Sustainability - MDPI (10 publications). respectively. Most relevant keywords are wastes management (18)appearance), followed by the waste model (9 appearances), waste systems (8 appearances), followed by waste impact, municipal solidwaste and solid waste management (each with 6 appearances) (Figure 3). Waste management appears mostly linked to waste recycling and circular economy, within the first keywords cluster (blue cluster within Figure 3), while waste related to management, recovery, sustainability, environmental economics and households defines the 2nd resulted keywords cluster (the red cluster emphasized in Figure 3). However, considering the keywords' conceptual structure map presented in Figure 4, it can be observed that it explains over 65% of the data variance. However, most of the red dimension revealed by the conceptual structure map is widened and, together with the green dimension, can be used in order to create the necessary roadmap for developing multiples frameworks which can be used as a baseline for developing high accuracy decision support tools.



Figure 3. Keywords occurrences network



Figure 4. Keywords conceptual structure map

No search resmethodologieslink waste and waste management within Romania or V4 to artificial intelligence methodology were identified, a fact which confirms the novelty of the present research structure and desideratum. However, if analysing the total citations per country, it can be observed that Romania is in the 1st position (total citations - TC 815, average article citation AAC 9.2), followed by Italy (173 TC, 57.7 AAC) and Portugal (149 TC, 149 AAC). Reported to the V4, the geopolitical group is mostly represented by Poland (5th place) with 8 TC and 8 AAC, followed by Slovakia (6th place) with 5 TC and AAC 5 and the Czech Republic (7th place,) with 3 TC and 3 AAC.

Most cited paper in the area of the study presented in the present article is Synthesis and characterization of new zeolite materials obtained from fly ash for heavy metals removal in advanced wastewater treatment with 200 citations. followed by *Packaging* waste recycling in Europe: Is the industry paying for it? with 149 citations, Introduction of the circular economy within developing regions: A comparative analysis of advantages and opportunities for waste valorisation with 145 citations and Forecasting municipal solid waste generation using prognostic tools and regression analysis with 115 citations. In terms of countries' collaboration for the elaboration of research articles in the area of interest for the present paper, Romania mostly collaborates with Italy (Figure 5).



Figure 5. Countries' collaboration in the area of study

XGBoost prediction models

The results are based on the application of XGBoost machine-learning algorithms and indicate that total waste, as a dependent parameter, can be accurately evaluated considering plastic wastes (feature importance-FI = 1.53, Rsq. = 0.75, RMSE = 0.47) in the case of V4 group, while for Romania, the dependent parameters identified as most reliable are chemical wastes (FI = 0.58) and industrial effluent sludges (FI = 0.04), with

lower accuracy metrics (Rsq. = 0.46, RMSE = 0.75).

In terms of waste treatment (WT), the portable batteries and accumulators' market (FI = 0.45) presents high reliability to be used as the main predictor (Rsq. = 0.80, RMSE = 0.42) for V4 support tool, while for Romania, the waste generation (FI = 1.57, Rsq. = 0.85, RMSE = 0.36) highly explains the variability of WT. However, batteries and accumulators waste (FI = 0.77, Rsq. = 0.82, RMSE = 0.39) can be used as a reliable predictor for WT variation in a more extended analytical framework, in the case of Romania.

GAM modelling framework

The GAM prediction models are made in order to assist the decision-support process in selecting the correct mechanisms for limiting both the total wastes (TW), and also maximizing and encouraging waste treatment (WT).

The present study analysis both indicators in various analytical framework scenarios, based on supporting databases from a multitude of relevant predictors, in order to reveal the controlling mechanisms for both Romania and V4 geopolitical group and to identify to which degree the two analysed entities can merge in a future globalization taskforce for improvement of the resilience against the wastes in the Eastern part of European Union (EU). Therefore, if TW is targeted to be predicted, and UOW, CW, HCBW, PW and BAW are considered independent parameters, it can be stated that used oil wastes, chemical wastes and plastic wastes all significantly influence the total wastes production and can be successfully used as tools for minimizing the TW (Figures 6A, 5B, 5D). Thus, targeting to decrease the quantity of the nexus used oils-chemical-plastic wastes could be the solution for controlling the TW in the case of Romania. However, it seems that healthcare and biological wastes have a reverse impact on TW since their increase creates premises for the decrease of TW in Romania (Figure 6C).

The batteries and accumulator's wastes increase until 38000 tonnes generate an increase in TW, while a further increase, over the previously mentioned value generates a reverse effect on TW, decreasing the value of the predicted indicator (Figure 6E). This can be because increasing BAW could represent the beginning of a long-term strategy of increasing the sustainability of several industries by shifting to alternative sources which are more sustainable (i.e., adoption of electric vehicles powered by batteries and decreasing the number of conventional engine vehicles). Thus, the results of this strategy, in relation to TW, will be visible after a long-term period, therefore, explaining the prediction dynamics presented in figure 6E.



Figure 6. Romania GAM prediction models for the TW dependent variable, based on several predictors (A - OUW; B - CW; C - HCBW; D - PW; E - BAW)

The waste treatment indicator, in the case of Romania, is predicted to increase as both the waste generation and portable batteries and accumulators market increase (Figure 7A, 7C). Therefore, it can be observed that the Romanian mechanism of waste treatment is highly responsive to any increase in waste generation due to the increase in the waste generation products market, enhancing the opinion according to which Romania has the capacity to handle difficult scenarios which could appear in the future, in terms of waste management. Also, both recycling rates of packaging wastes and wastes collected quantity associated with portable batteries and

accumulators emphasize a Gaussian shape prediction in relation to TW in Romania (Figure 7B, 7D). This completes the finding previously mentioned for WT prediction, according to which the shifting to more sustainable tools for accomplishing the desideratum targeted by the EU Green Deal (GD) had already started in Romania and this will not imply immediate effects on WT, instead, will induce positive long-term effects.



Figure 7. Romania GAM prediction models for the WT dependent variable, based on several predictors (A - WG; B - RRPW; C - PBAM; D - WCPBA)

If targeting to develop a wider analytical framework in order to predict the WT in Romania, based mainly on the waste sources, indicators such as OUW, CW, HCBW, BAW, TW, IES and PW are to be considered. Thus, it seems that after the application of GAM algorithms, two groups were identified, as follows: 1st group includes OUW, CW, TW, PW and IES, while the 2nd group includes HCBW and BAW (Figure 8). Therefore, waste treatment is highly responsive both to the variation of total waste quantitative parameters, but also to oil, chemical, plastics and industrialbased wastes (Figures 8A, 8B, 8E, 7F and 8G), an increase of these predictors generating a relatively similar response in the evolution of WT in Romania. However, the healthcare and biological wastes and the batteries and accumulators' wastes confirm their reverse impact on WT (Figure 8C, 8D), similar to what they had done in relation to TW (Figure 6C, 6E).



Figure 8. Romania GAM prediction models for the WT dependent variable, based on several predictors (A - UOW; B - CW; C - HCBW; D - BAW; E - TW; F - PW; G - IES)

In the case of V4, the prediction relation between TW, as a predicted variable, and UOW, CW and HCBW as predictors, is similar to that described in the case of Romania (Figure 9A, 9B, 9C). However, if PW, IES and HSV are considered predictors for predicting the TW dependent variable, it can be stated that their impact is the opposite, compared to the Romania case (Figure 9D, 9E, 9G). Thus, an increase in plastic waste and industrial effluent sludges generates a decrease in TW, revealing that in the case of V4 the industry sector is more environmentally sustainable, compared to Romania and the plastics wastes management is more efficient, so the control of TW can be done in a more generous way which can also mean the increase of plastic wastes as a low impact substitute of other high environmental impact types of wastes which mostly compose TW value within the analysed geopolitical group. However, household wastes can be used as a tool for detecting the TW as their increase also generates an increase in predicted values of this dependent parameter (Figure 9G). The batteries and accumulators waste increases are

164

related to TW decrease (Figure 9F), a fact which can be associated with the green technologies' implementation status in V4 - thus, if V4 had overcome the 1st state of transition towards the extended use and implementation of green technologies, this could indicate that the initial investments for shifting to EU GD desideratum were efficient in achieving their goal.



Figure 9. The V4 GAM prediction models for the TW dependent variable, based on several predictors (A - OUW; B - CW; C - HCBW; D - PW; E - IES; F - BAW; G - HSV)

The WT prediction by considering the WG, RRPW, PBAM and WCPBA as predictors reveal some similarities between V4 and Romania. Thus, in terms of using WG and PBAM as predictors for predicting WT, the situation between Romania and V4 is relatively similar since an increase of both predictors can relate to an upward trend of WT (Figures 10A, 10C, 7A, 7C). However, if considering RRPW and WCPBA, the situations are different, both indicating a Gaussian variation of WT predicted values, revealing that V4 has not yet reached maturity when it comes to recycling and packaging wastes management, as well as portable batteries and accumulators waste collection management (Figure 10B, 10D). Predicting the WT in relation to a wide range of predictors as revealed in Figure 11, concluded some similarities between V4 and Romania if considering UOW, CW, HCBW, IES and BAW (Figures 11A, 11B, 11C, 11G, 11H, 8A, 8B, 8C, 8D, 8G).



Figure 10. The V4 GAM prediction models for the WT dependent variable, based on several predictors (A - WG; B - RRPW; C - PBAM; D - WCPBA)



Figure 11. The V4 GAM prediction models for the WT dependent variable, based on several predictors (A - UOW; B - CW; C - HCBW; D - HSV; E - TW; F - PW; G - IES; H - BAW)

However, some indicators have a reverse impact, such as PW (Figures 11F, 8F) on WT, most probably due to the more advanced level of V4 in terms of implementing sustainable measurements in some waste-generating areas. In accordance with other studies, also this study is subject to several limitations. The main limitation is related to the dataset structure and time frame.

Widening the database by including other relevant synthetic indicators, corresponding to subgroups attached to the main waste management indicators, can offer new mechanisms for improving the management of waste. Also, a time-series data frame could offer the possibility of generating several forecasting analytical frameworks, using algorithms such as recurrent neural network LSTM.

CONCLUSIONS

The analytical framework of both Romania and V4, in terms of predicting the total waste production and total waste treatment, reveals several similarities, as follows: the total wastes can be predicted, with similar outputs, by using oil, chemical, healthcare and biological, as well as portable batteries and accumulators market size as main predictors; predictors as total wastes, oil, plastics and industry wastes, as well as healthcare and biological wastes can be used for predicting the waste treatment indicator, with similar results between Romania and V4. However, there are several peculiarities which can be associated with each of both analytical frameworks, and which reveal that V4 is more advanced in certain areas involving the shift to EU Green Deal desideratum, compared to Romania, in terms of waste management. Future studies should target widening the number of analytical framework dimensions to assure a more holistic approach.

ACKNOWLEDGEMENTS

The present research was supported by the project: "An Integrated System for the Complex Environmental Research and Monitoring in the Danube River Area", REXDAN, SMIS code 127065, co-financed by the European Regional Development Fund through the

Competitiveness Operational Programme 2014-2020, contract no. 309/10.07.2021.

The present research was supported by the project "Integrated research and sustainable solutions to protect and restore Lower Danube Basin and coastal Black Sea ecosystems, (ResPonSE)", 760010/30.12.2022, component C9. Private Sector Support, Research, Development and Innovation, Investment "I5. Establishment and operationalization of Competence Centers".

REFERENCES

- Buczkó, K. (2018). Municipal solid waste management in Hungary:A review. Waste management & research, 36(3), 195-204.
- Circular Economy Policy Library. (n.d.). Retrieved July 12, 2023, from https://circabc.europa.eu/ui/group/6e9b7f79-da96-4a53-956f-e8f62c9d7fed/library/37e8e207-6222-4212-ad7c-e809e64df72c; accessed on 12.03.2023.
- EUR-Lex 02012L0019-20180704 EN EUR-Lex. (n.d.). Retrieved July 12, 2023, https://eurlex.europa.eu/legalcontent/EN/TXT/?uri=CELEX:02012L0019-20180704: accessed on 17.03.2023.
- Fialová, J. (2019). Waste management in the Czech Republic and its comparison with selected European Union countries. *Polish Journal of Environmental Studies*, 28(6), 3759–3769.

- Implementation of the Waste Framework Directive. (n.d.). Retrieved July 12, 2023, from https://environment.ec.europa.eu/topics/waste-andrecycling/implementation-waste-frameworkdirective en; accessed on 23.04.2023.
- Khamma, T. R., Zhang, Y., Guerrier, S., & Boubekri, M. (2020). Generalized additive models: An efficient method for short-term energy prediction in office buildings. *Energy*, 213, 118834.
- Landfill waste. (n.d.). Retrieved July 12, 2023, from https://environment.ec.europa.eu/topics/waste-andrecycling/landfill-waste en; accessed on 02.05.2023.
- Mišík, M. (2019). Waste management policy in Slovakia and its European context. *E3S Web of Conferences*, 92, 02001.
- Murphy, R. R., Perry, E., Harcum, J., & Keisman, J. (2019). A Generalized Additive Model approach to evaluating water quality: Chesapeake Bay case study. *Environmental Modelling & Software*, 118, 1–13.
- Packaging waste. (n.d.). Retrieved July 12, 2023, from https://environment.ec.europa.eu/topics/waste-andrecycling/packaging-waste_en; accessed on 05.03.2023.
- Single-Use Plastics Directive European Bioplastics e.V. (n.d.). Retrieved July 12, 2023, from https://www.european-bioplastics.org/policy/singleuse-plastics-directive/; accessed on 25.03.2023.
- Virsta, A., Sandu, M.A., Daraban, A.E. (2020). Dealing with the transition from in line economy to circular economy - public awareness investigation in Bucharest. AgroLife Scientific Journal, 9(1), 355-361.