Artificial Intelligence (AI): An Innovative Approach for Auditing Field Sampling in Waste Management Context

Igor Pinhal Luqueci Thomaz
Researcher, UFRJ, Rio de Janeiro, Brasil, pinhal@coc.ufrj.br

Claudio Fernando Mahler
Professor, COPPE/UFRJ, Rio de Janeiro, Brasil, mahler@coc.ufrj.br

ABSTRACT: Audit and control models are necessary in an open, transparent and accountable government. The use of Artificial Neural Networks (ANN) in audit models to predict the physical properties of Municipal Solid Waste (MSW), such as gravimetric composition, has been discussed, especially after the world experienced the COVID-19 pandemic. Traditional on-site sampling is expensive, slow and requires specialized professionals, who are exposed to physical and chemical risks, while predictions made by ANN can be carried out with little or no waste handling, and can be done retroactively, to fill gaps in information left during the mandatory quarantine period. ANNs depend on data sets provided by third parties, which need to be chosen looking for reliability, periodicity and availability. In this work we show the possibility of adopting ANN models fed by socioeconomic datasets relating to an important capital in Brazil to make predictions of the physical properties of MSW produced in this city destined for landfills.

KEYWORDS: Municipal Solid Waste; Artificial Neural Networks; gravimetric composition; landfills; audit

RESUMO: Modelos de auditoria e controle são necessários em um governo aberto, transparente e responsável. O uso de Redes Neurais Artificiais (RNA) em modelos de auditoria para previsão de propriedades físicas de Resíduos Sólidos Urbanos (RSU), como a composição gravimétrica, tem sido discutido, especialmente após o mundo ter vivenciado a pandemia COVID-19. A amostragem tradicional in loco é cara, lenta e requer profissionais especializados, que ficam expostos a riscos físicos e químicos, enquanto as previsões feitas pelas RNA podem ser realizadas com pouco ou nenhum manuseio de resíduos, e podem ser feitas retroativamente, para preencher lacunas de informações deixadas durante o período de quarentenas obrigatórias. As RNA dependem de conjuntos de dados fornecidos por terceiros, que precisam ser escolhidos buscando confiabilidade, periodicidade e disponibilidade. Neste trabalho mostramos a possibilidade de adotar modelos de RNA alimentados por conjuntos de dados socioeconômicos relativos a uma importante capital do Brasil para realizar previsões das propriedades físicas dos RSU produzidos nesta cidade destinados a aterros sanitários.

PALAVRAS-CHAVE: Resíduos Sólidos Urbanos; Redes Neurais Artificiais; composição gravimétrica; aterros sanitários; auditoria

1 INTRODUCTION

The Municipal Solid Waste (MSW) can be analyzed by its physical characteristics, such as gravimetric composition and specific weight. They are very relevant to planning suitable services of urban cleaning, and reliable models for estimating waste generation are crucial to allow better use of public funds (THOMAZ, 2016). The waste sector traditionally uses on-site sampling procedures, which demand the hiring of specialized personnel, consumes a significant amount of time and financial resources, and are dangerous for employees responsible for handling them (THOMAZ et al., 2023). But the COVID-19 pandemic has changed the world, causing transformations in the economy, employment and daily life (IKIZ et al., 2021). Concerns about the risks of solid waste contamination have increased, mainly due to the sudden increase in waste infected with virulent agents (PENTEADO and CASTRO, 2021).

At that juncture, when the limitations imposed by the COVID-19 pandemic had to be respected, it was simultaneously mandatory to maintain a minimum level of quality in waste management. By adopting healthy and safe methods to monitor the waste produced, predictive models utilizing Artificial Intelligence (AI), such as Artificial Neural Networks (ANN), gained visibility and importance them (ADELEKE et al., 2021; ANDEOBU et al., 2022).

Given the success of models that aim to fill information gaps in times of pandemic, such as the one presented by Thomaz et al. (2023), this work proposes an audit model, based on ANN, that can be used to audit gravimetric composition and specific weight datasets obtained from in-situ sampling of MSW. To achieve the objectives, an ANN model was proposed, fed by socioeconomic datasets related to a significant capital of Brazil. To properly focus on the theme of Auditing, it was decided not to use data related to the pandemic and post-pandemic periods, on which other dedicated works exist.

To select appropriate datasets, a literature review was conducted, from which it was concluded that the properties of MSW are influenced by indicators such as: income; GDP; educational degree; family size; average age of family member; access to medical care insurance; access to electricity, potable water supply and sanitation system; population (VAZQUEZ et al., 2020; GHANBARI et al., 2021; NOMAN et al., 2023; THOMAZ et al., 2023).

The model presented in this paper has been tested with a dataset consisting of population data, Gross Domestic Product (GDP) and total annual electricity consumption, made available for a 15-year interval, from 2004 to 2019. However, due to the pragmatic spectrum of the paper, it was decided to present only the forecasts made with the data for the year 2011, which represents the average year within the studied set, and for which there is a huge availability of detailed data, necessary for the correct understanding of the concepts explored.

2 MATERIALS AND METHODS

2.1 Study area and datasets

THOMAZ (2016) tested different scenarios to understand how the socioeconomic characteristics of the population of a given region determines the gravimetric composition of Municipal Solid Waste (MSW), allowing a reliable prediction of the gravimetric characteristics of the waste produced, knowing only its population profile. ANN models enable a clean approach, without the need for waste handling by workers.

ANN are capable of simulating the behavior of the animal brain, which is an organ able to learn without know the algorithm that produces the problem, able to generalize and robust. The use of ANN allows the understanding of complex phenomena as well as allows the estimation of dependent variables, making use of other more accessible variables.

Models, such as those proposed by Thomaz (2016) and Thomaz et al. (2023), consider that waste reflects the daily behavior of society, then, it is possible to infer that there is a relationship of predictability between the characteristics of society and the properties of the waste produced by it.

The model uses socioeconomic data, assuming that factors such as purchasing power, educational level, habits, customs and density population govern the production and characteristics of the waste generated. The use of input variables that are not relevant for determining the outputs causes the introduction of noise at the input and impairs the performance of the network, while the elimination of input variables that contain relevant information about the outputs results in the deterioration in the accuracy of the output results prediction.

The study area chosen to validate the model was the Municipality of Rio de Janeiro, capital of the State of Rio de Janeiro, located at southeast region of Brazil. Besides being one of the main economic and financial regions of the country, Rio de Janeiro is known internationally for its attractive culture and landscape. It is important to note that the data collection stage is one of the most time-consuming steps in carrying out the work, because a huge amount of raw data should be obtained from different sources, in different formats, with different reference parameters.

In the year 2010, the population resident in the Municipality was 6,320,446, equivalent to almost 40% of the population of the State of Rio de Janeiro (IBGE, 2010). Almost 10 years later, in the year 2019, the population in the municipality of Rio de Janeiro grew to 6,718,903 inhabitants (IBGE, 2019).

The collection of MSW in the Municipality of Rio de Janeiro is carried out by the Municipal Cleaning Company (COMLURB), which provides numerous data regarding the collection and disposal of MSW (PMRJ, 2021). The MSW gravimetric composition, divided into standardized fractions Paper-Cardboard, Plastic, Glass, Metal, Organic matter, and others, and the MSW specific weight were provided by COMLURB (2023).

The GDP of the municipality of Rio de Janeiro exhibited a growth trend for most of the years corresponding to the period between 2004 and 2020, with the period between the years 2010 and 2013 standing out, where the minimum annual growth exceeded 10%. However, starting from the year 2018, when the GDP reached its peak of R$ 363 billion, a decline began, with a drop of 2.33% recorded between the years 2018 and 2019, and an even larger drop, of 6.66%, recorded between the years 2019 and 2020 (IBGE, 2023).

As previously stated, the year 2011 represents the midpoint of the 15-year interval studied from 2004 to 2019, in addition, there is a large amount and diversity of socioeconomic data available between the years 2004 and 2011, including geographically referenced data (SNIS, 2023). This scenario is one of the best possible, which is when a wide range of socioeconomic information is available to feed the forecasting algorithm, allowing it to be verified which socioeconomic characteristics are most relevant.

2.2 Computational Tool

The MATLAB software was selected to program the ANN, because it presented the best performance among the other tested programs, such as “Weka”, a freeware from the University of Waikato, and the paid Excel add-in called “NeuralTools” from Palisade company (THOMAZ et al., 2023).

Within the computational environment, population and MSW data were grouped into spreadsheets that were filtered, processed and transformed into a detailed information database, used to assemble sets “entries-targets” compatible with the software used during study.

The Toolbox incorporated into MATLAB allows real-time monitoring of network learning (figure 1), making it possible to cancel, reconfigure, change the hyperparameters and restart the process at any time. The preliminary evaluation of the ANN training performance can also be conducted through a plot of the Mean Squared Error (MSE) values against the number of training epochs. The MSE metric emphasizes larger errors by squaring each individual error prior to averaging these squared errors.

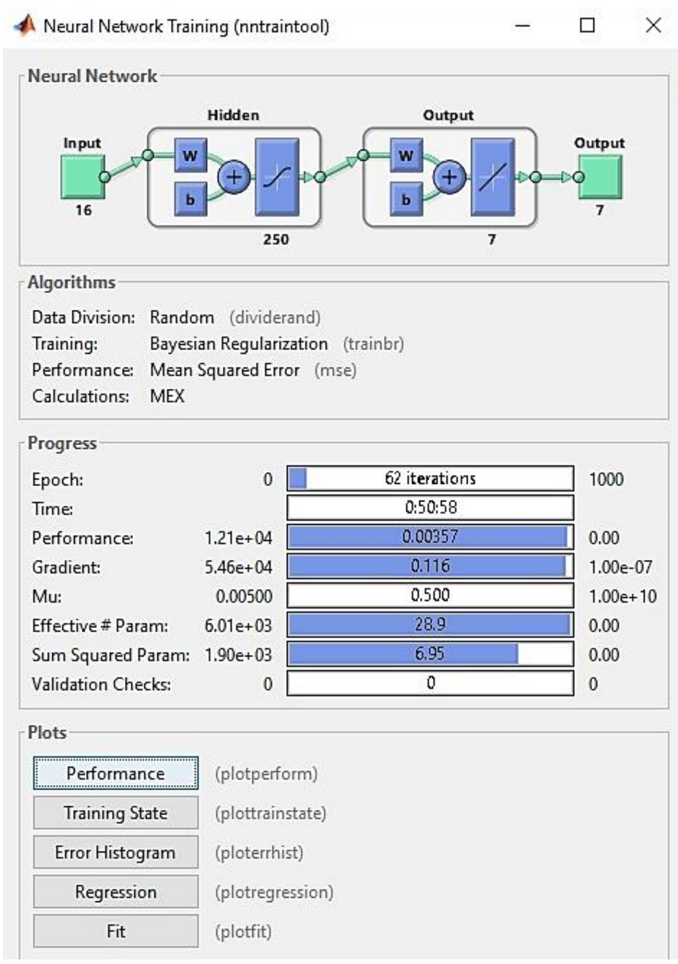

Figure 1. Real-time monitoring of network learning. Source: Author.

It was chosen networks fed by multiple layers of output nodes, or Multilayer Perceptrons (MLP) (figure 2). They have one or more hidden (intermediate) layers, that is, each neuron in a layer has direct connections to neurons in the adjacent layer. In this case, there is also a comparison between the results, and the error found is inserted back into the network in order to adjust the values, further decreasing the value of the error function.

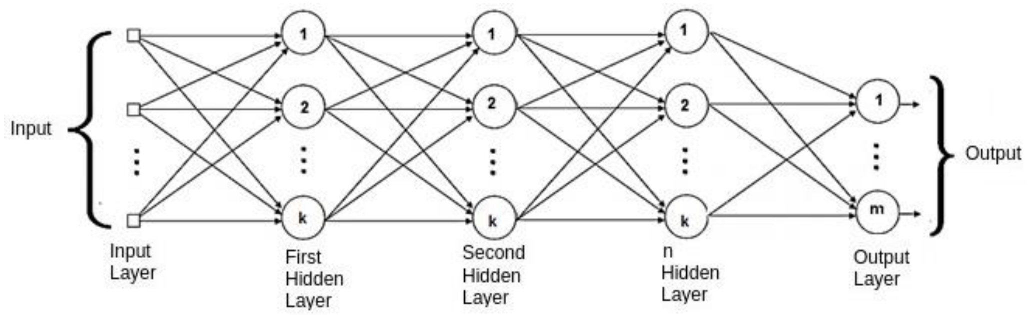

Figure 2. Schematic diagram of a Multilayer Perceptrons (MLP) neural network.

The number of layers hidden at the script was chosen according to the empirical method proposed by BEALE (2022), avoiding excessive layers, which could cause overfitting problem, damaging the process of generalization. Tests with few hidden layers were started, and the number of layers was increased to hundreds of them. The simulations carried out with more than two hundred layers required a very large computational effort, and the results were not satisfactory. For this case, it was observed that fifty layers are sufficient.

3 RESULTS AND DISCUSSION

For the Municipality of Rio de Janeiro, the combination of population data, GDP and total annual electricity consumption, proved to be adequate for the 2004-2011 period, therefore, they were used to build the MATLAB input database (Table 1) and choose targets (Table 2). The results for the chosen paradigm year can be seen in Table 3, and the compliance check is shown in Table 4.

Table 1. Input database to feed the algorithm developed in MATLAB.

INPUTS		2004	2005	2006	2007	2008	2009	2010	2011
Population		6,051,399	6,094,183	6,136,652	6,093,472	6,161,047	6,186,710	6,320,446	6,355,949
GDP Present Value (millions R$)		112,675	117,772	128,026	140,095	158,757	170,517	190,018	209,366
Classes of electricity consumption (MWh)	Residential	4,840,630	5,264,761	5,286,605	5,394,924	5,382,944	5,759,607	5,985,329	6,018,867
	Industrial	2,495,426	2,251,276	1,852,207	3,875,665	3,891,549	2,583,374	2,860,884	2,562,750
	Commercial	4,343,455	4,685,725	5,066,104	4,963,312	5,055,236	5,280,937	5,535,628	5,684,810
	Rural	1,779	1,894	1,888	2,109	2,130	2,197	2,245	2,360
	Public powers	936,013	1,021,809	1,224,161	1,154,018	1,149,348	1,229,854	1,245,820	1,286,482
	Street lighting	423,173	504,554	166,819	461,413	447,082	435,053	442,492	448,021
	Public service	588,144	570,904	1,104,240	603,604	609,149	783,308	796,484	824,004
	Own consumption	42,669	40,997	66,768	65,446	59,100	55,815	66,511	74,411

Table 2. Targets selected for the year 2011 gravimetric composition and specific weight prediction.

TARGETS		2004	2005	2006	2007	2008	2009	2010	2011
Gravimetric composition	Paper - Cardboard (%)	12.5	13.5	14.8	14.6	16.1	16.1	16.5	Target ⋔
	Plastic (%)	15.4	15.3	14.7	17.2	19.2	20.3	19.1	Target ⋔
	Glass (%)	3.2	3.2	2.7	2.9	2.9	2.8	3.0	Target ⋔
	Metal (%)	1.7	1.7	1.6	1.6	1.6	1.7	1.4	Target ⋔
	Others (%)	7.4	5.5	4.8	5.6	4.7	5.4	5.0	Target ⋔
	Organic matter (%)	59.7	60.7	61.4	58.1	55.7	53.6	55.0	Target ⋔
Specific Weight (Kg/m³)		153.6	148.4	144.9	144.5	31.2	124.0	111.1	Target ⋔

Table 3. Prediction results for the year 2011 gravimetric composition and specific weight.

OUTPUTS		2011
Gravimetric composition	Paper - Cardboard (%)	16.6486
	Plastic (%)	18.7446
	Glass (%)	2.8387
	Metal (%)	1.4758
	Others (%)	4.7447
	Organic matter (%)	55.5476
Specific Weight (Kg/m³)		104.2891

Table 4. Compliance check for the year 2011 gravimetric composition and specific weight prediction results.

COMPLIANCE CHECK		Reference 2011	Estimated 2011	Error Absolut	Error Relative
Gravimetric composition	Paper - Cardboard	16.84%	16.65%	0.1869	1.11%
	Plastic	19.29%	18.74%	0.5431	2.82%
	Glass	3.19%	2.84%	0.3479	10.92%
	Metal	1.68%	1.48%	0.2056	12.23%
	Others	6.33%	4.74%	1.5887	25.08%
	Organic matter	52.68%	55.55%	-2.8722	-5.45%
Specific Weight (Kg/m³)		109.09	104.29	4.7961	4.40%

The relative errors of the predictions (Table 4) were less than 11%, except for the fraction "metal" (relative error 12.23%) and for the fraction "other residues" (relative error 25.08%). Forecasting errors are shown visually in the following graphs (figure.3 and figure 4).

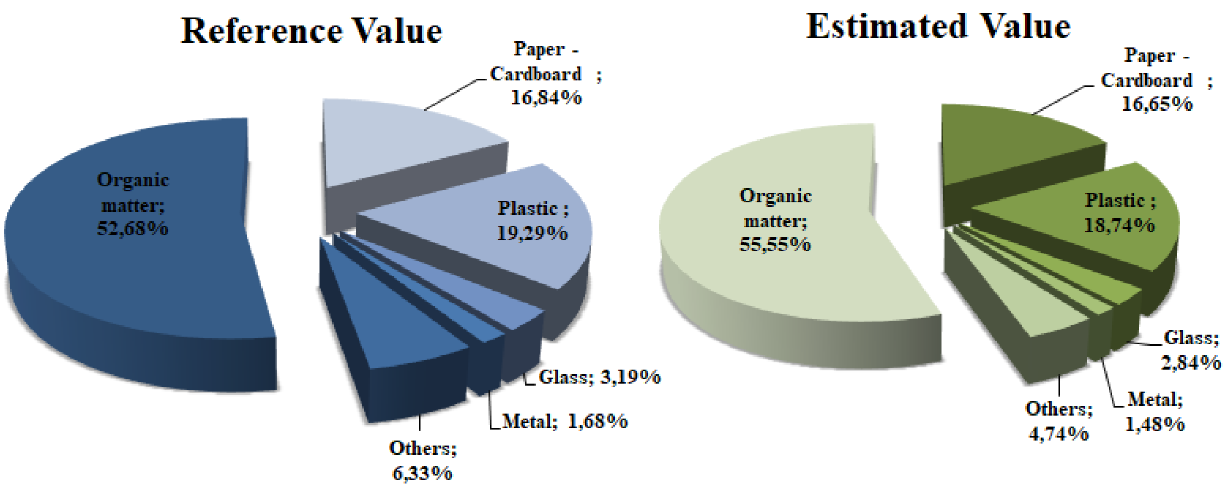

Figure 3. Graphical representations of the reference fractions and estimated Gravimetric Composition of the year 2011 in the Municipality of Rio de Janeiro.

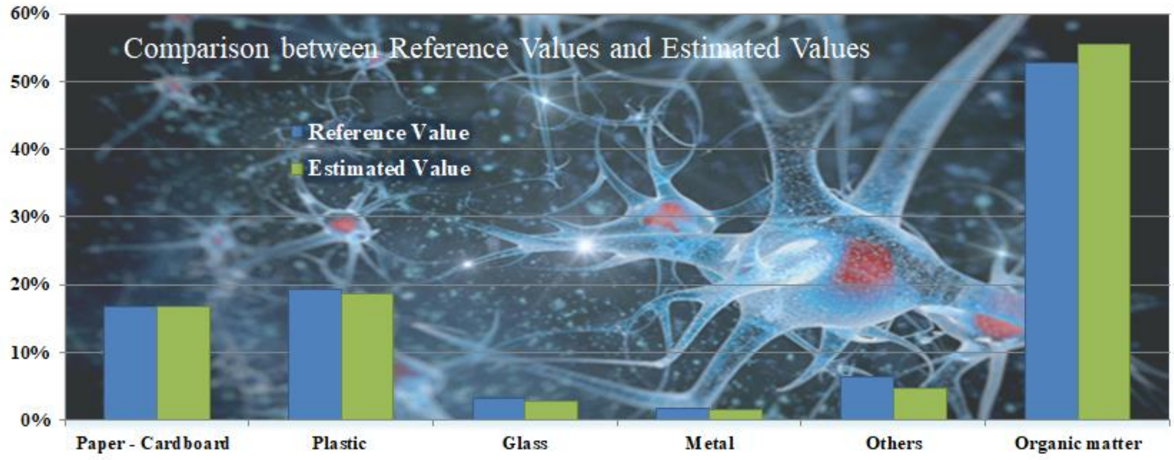

Figure 4. Comparison between Reference Values and Estimated Values of Gravimetric Composition of the year 2011 in the Municipality of Rio de Janeiro.

Understanding predictive models, in terms of interpreting and identifying actionable insights, is a challenging task, but it is crucial to make an analysis of the relative error of more than 25% in module in the forecast of the fraction “other wastes”, and the relative error of 12.23% in module in the prediction of the “metal” fraction.

It is possible to verify that the fractions “plastic”, “glass”, “metal” and “other wastes” were underestimated, while the fraction of “organic matter” was overestimated in the forecast (Tables 3 and 4). When analyzing the absolute errors, it could be verified that the gravimetric fractions of the underestimated plots compose the overestimated plot, which is intuitive. In this sense, it is probable that the simulation has classified “generically” part of the referred underestimated fractions as being “organic matter”.

Analyzing the history of the organic matter fraction, it can be observed that there was a growth trend between 2004 and 2006, followed by a downward trend between 2006 and 2009, with a return to the growth trend between 2009 and 2010 (Table 1). Looking at this perspective, there is an indication that the neural network considered that the growth trend would be maintained for at least one more year. However, the projection was conservative, with a slight positive difference of 0.14% between the forecasts for 2010 and 2011. The tendency is for errors to become smaller and smaller as simulations are made and databases grow.

4 CONCLUSIONS

The technological advancement has democratized the access to complex computational tools, previously limited to specific sectors, for widespread societal benefit. Opting for ANN models for predicting the gravimetric composition and specific weight of MSW proves to be more efficient than traditional on-site sampling. The latter necessitates the engagement of specialized professionals, is time-intensive, and dangerous to the workers responsible for waste handling, particularly in times of a pandemic, such as the COVID-19 outbreak.

Due to its qualities, this tool is highly recommended to support policymakers in their resource management strategies, including auditing, since it allows analyzing the consistency of gravity data and specific weight provided by third parties, making it easier for the auditor to decide where to concentrate efforts in the search for suspicious situations.

Despite errors exceeding 10% for two gravimetric fractions, the immediate implementation of the technique would be beneficial in audit tasks. The results are not used to fill information gaps, but as a benchmarking tool for auditors in the audit scope definition stage. For this purpose, the model can be useful in determining the relevance of each planned point for the audit.

A significant challenge for the future would be the implementation of audit tools automatically powered by waste management datasets from the main waste-generating municipalities. These tools could inherently contain ANN models capable of identifying anomalies that need to be urgently investigated, enabling audit work to be conducted while the problems are still small.

REFERENCES

Adeleke, O., Akinlabi, SA, Jen, T., Dunmade I. (2021) Application of artificial neural networks for predicting the physical composition of municipal solid waste: An assessment of the impact of seasonal variation. Waste Management & Research: The Journal for a Sustainable Circular Economy, vol. 39. http://dx.doi.org/10.1177/0734242x21991642 .

Andeobu, L., Wibowo, S., Grandhi, S. (2022) Artificial intelligence applications for sustainable solid waste management practices in Australia: A systematic review. Science of The Total Environment, vol. 834. https://doi.org/10.1016/j.scitotenv.2022.155389.

Beale, M.H., Hagan, M.T., Demuth, H.B. (2022) Deep Learning Toolbox™ User's Guide, 14th Ed, Massachusetts, MathWorks.

COMLURB. Table 1494. (2023) Main characteristics of household waste - percentage gravimetric composition (…)(In Portuguese). [accessed 2023 Sep 11]. www.arcgis.com/sharing/rest/content/items/ccdc3c0946ff430db6ef479befe8a5a5/data

Ghanbari, F., Kamalan, H., Sarraf, A. (2021) An evolutionary machine learning approach for municipal solid waste generation estimation utilizing socioeconomic components. Arab J Geosci. 14(2). http://dx.doi.org/10.1007/s12517-020-06348-w.

Ikiz, E., Maclaren, V.W., Alfred, E., Sivanesan, S. (2021) Impact of COVID-19 on household waste flows, diversion and reuse: The case of multi-residential buildings in Toronto, Canada. Waste management, vol 164. https://doi.org/10.1016/j.resconrec.2020.105111

Instituto Brasileiro de Geografia e Estatística (IBGE). (2010) Censo Demográfico 2010 (In Portuguese), Rio de Janeiro.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2019) Estimativas da população residente para os municípios e para as unidades da federação brasileiros com data de referência em 1º de julho de 2019 (In Portuguese), Brasília.

Instituto Brasileiro de Geografia e Estatística (IBGE). (2023) Table 1517. Produto Interno Bruto (PIB) das capitais (...) Brasil - 2002-2012 (in portuguese). [accessed 2023 Sep 31]. https://www.arcgis.com/sharing/rest/content/items/0fee074296c9473aa9752813c453d1c3/data.

Noman, A.A., Rafizul, I.M., Moniruzzaman, S.M., Kraft, E., Berner, S. (2023) Assessment of municipal solid waste from households in Khulna city of Bangladesh. Heliyon. http://dx.doi.org/10.1016/j.heliyon.2023.e22446.

Penteado, C.S.G.; Castro, M.A.S. (2021) Covid-19 effects on municipal solid waste management: What can effectively be done in the Brazilian scenario? Waste management, vol 164. https://doi.org/10.1016/j.resconrec.2020.105152

PMRJ. (2021) Plano Municipal de Gestão Integrada de Resíduos Sólidos (PMGIRS) da Cidade do Rio de Janeiro (in portuguese). http://www.rio.rj.gov.br/dlstatic/10112/13305794/4334422/PMGIRSVERSAO12_08_21.pdf

SNIS. Table 1477. (2023) População total, população atendida, quantidade de ligações, quantidade de economias e extensão de rede de abastecimento de água (...) (in portuguese). [accessed 2023 Jul 7]. https://www.arcgis.com/sharing/rest/content/items/0e0c2674837242b18d78f9a66f638433/data.

Thomaz, I.P.L. (2016) Utilização de Redes Neurais Artificiais (RNA) para previsão da composição gravimétrica e peso específico de Resíduos Sólidos Urbanos (RSU) (In Portuguese), master's thesis, COPPE/UFRJ, Rio de Janeiro.

Thomaz, I.P.L., Mahler, C.F., Calôba, L.P. (2023) Artificial Intelligence (AI) applied to waste management: A contingency measure to fill out the lack of information resulting from restrictions on field sampling. Waste Management Bulletin. https://doi.org/10.1016/j.wmb.2023.06.002

Vazquez, Y.V., Barragán, F., Castillo, L.A., Barbosa, S.E. (2020) Analysis of the relationship between the amount and type of MSW and population socioeconomic level: Bahía Blanca case study, Argentina. Heliyon. http://dx.doi.org/10.1016/j.heliyon.2020.e04343