Waste Management 140 (2022) 31–39

Contents lists available at ScienceDirect

Waste Management

journal homepage: www.elsevier.com/locate/wasman

Forecasting municipal solid waste in Lithuania by incorporating socioeconomic and geographical factors

Agne Paulauskaite-Taraseviciene [a], Vidas Raudonis [b], Kristina Sutiene [c,*]

[a] Department of Applied Informatics, Kaunas University of Technology, Studentu 50, Kaunas 51368, Lithuania
[b] Department of Automation, Kaunas University of Technology, Studentu 48, Kaunas 51367, Lithuania
[c] Department of Mathematical Modelling, Kaunas University of Technology, Studentu 50, Kaunas 51368, Lithuania

ARTICLE INFO

Keywords:
Municipal solid waste
Forecasting
Composition
Machine learning
K-nearest neighbours

ABSTRACT

Forecasting municipal solid waste (MSW) generation and composition plays an essential role in effective waste management, policy decision-making and the MSW treatment process. An intelligent forecasting system could be used for short-term and long-term waste handling, ensuring a circular economy and a sustainable use of resources. This study contributes to the field by proposing a hybrid k-nearest neighbours (H-kNN) approach to forecasting municipal solid waste and its composition in the regions that experience data incompleteness and inaccessibility, as is the case for Lithuania and many other countries. For this purpose, the average MSW generation of neighbouring municipalities, as a geographical factor, was used to impute missing values, and socioeconomic factors together with demographic indicator affecting waste collected in municipalities were identified and quantified using correlation analysis. Among them, the most influential factors, such as population density, GDP per capita, private property, foreign investment per capita, and tourism, were then incorporated in the hierarchical setting of the H-kNN approach. The results showed that, in forecasting MSW generation, H-kNN achieved MAPE of 11.05%, on average, including all Lithuanian municipalities, which is by 7.17 percentage points lower than obtained using kNN. This implies that by finding relevant factors at the municipal level, we can compensate for the data incompleteness and enhance the forecasting results of MSW generation and composition.

1. Introduction

Waste disposal and waste handling are some of the world's most pressing problems. Proper municipal solid waste (MSW) management is a priority for achieving the EU sustainable development goal – to reduce the negative effect on the environment and human health by deploying financially reasonable, technically feasible, socially and legally acceptable solutions (Abdel-Shafy and Mansour, 2018). Toxins emanating from landfills adversely affect the quality of life of nearby residents and other living organisms. The degradation time for some waste is more than 500 years. This long degradation time coupled with ongoing growth of the human population, rapid urbanization, and a booming economy means that the volume of waste is increasing much faster than natural decomposition (Chamas et al., 2020). Major improvements in this field have been achieved, but the large quantities of waste accumulating in landfills highlight that key challenges remain (Khandelwal et al., 2019; OECD, 2020).

Waste increase is visible worldwide, and it seems likely that this trend will continue in the future. According to the World Bank's report (Kaza et al., 2018), global waste will increase approximately 70% by 2050 if urgent action is not taken. The total waste generated in the EU-27 increased by 5% between 2010 and 2018, with 502 kg of MSW per capita generated in 2019 – highlighting that the EU-27 is failing to implement its policy goal of waste reduction. A number of studies have postulated a convergence between waste generation and booming economy, improved living standards, growing urbanization and globalisation, which is clearly observed for developing counties (Chen et al., 2020; Das et al., 2019; Koop and van Leeuwen, 2017). Some countries such as Bulgaria, Spain, Hungary, and Romania have reported a significant reduction of MSW generation and some countries have recycled, composted, and anaerobically digested more than 50% of their municipal waste (Eurostat, 2021). In contrast, however, Norway, Czechia, Slovakia, and Iceland were top generators of municipal waste per capita in 2019, with significant growth over a period of 15 years. Lithuania is among those countries that experienced growth in MSW generation per capita from 387 to 472 kg since 2005.

* Corresponding author.
E-mail address: kristina.sutiene@ktu.lt (K. Sutiene).

https://doi.org/10.1016/j.wasman.2022.01.004
Received 20 October 2021; Received in revised form 19 December 2021; Accepted 7 January 2022
Available online 13 January 2022

EU strategic challenges include: (a) to reduce waste generation and (b) make beneficial recycling that would contribute to the goal of the circular economy (Lee et al., 2017). The latter objective will require high-quality secondary waste to be returned to the production process. In this context, the waste management sector will have to become the responsible unit including making new business models, focusing on municipal waste prevention, targeted monitoring, waste generation forecasting, by deploying digitization technologies, and recovery of resources from landfills. In particular, waste forecasting and subsequent decision-making will be key to meeting the goals of sustainable environment protection. This may be a reason for the growing body of literature on forecasting MSW generation and its composition, covering approaches from simple linear models to highly sophisticated ones.

The recent comprehensive review (Abdallah et al., 2020) summarised 85 research studies, published between 2004 and 2019, focusing on the application of artificial intelligence (AI) in various fields of solid waste management due to its advancement and limitations of conventional computational techniques. Based on this study, the most popular AI systems for MSW forecasting and optimisation included artificial neural networks (ANNs) (Adamović et al., 2017; Ali and Ahmad, 2019; Azarmi et al., 2018), support vector machine (SVM) (Abbasi et al., 2013; Dai et al., 2020), linear regression (LR) (Chhay et al., 2018; Popli et al., 2021), decision trees (DT) (Cha et al., 2020; Solano Meza et al., 2019) and genetic algorithms (GA) (Alberdi et al., 2020; Oliveira et al., 2019). Many authors have focused on the use of different ANN algorithms, for instance, feed-forwards, feedback propagation models or deep neural networks (Akanbi et al., 2020; Antanasijević et al., 2013; Ayeleru et al., 2021; Huang et al., 2020; Oguz-Ekim, 2021). Promising results have been demonstrated by combining several ML models. For example, the combination of ANN and fuzzy logic based techniques (specifically, the ANFIS model clustered with grid partitioning using a linear type output membership function) has been used to predict provincial waste quantities in South Africa, which resulted in the lowest forecasting error MAPE = 12.67% observed in the study (Adeleke et al., 2020). Comparatively, GA-ANN (genetic algorithm – ANN), with four input neurons and five hidden layer neurons, was successfully employed for forecasting the MSW task in New Delhi, achieving RMSE = 95.7 and R^2 = 0.87 accuracy results (Soni et al., 2019). In the study of (Song et al., 2017b), the support vector regression has been introduced to adjust the forecasting residuals of the grey model used to forecast construction and demolition waste in China regions. The combination of two models ensured that the relative percentage errors were less than 0.1 compared to stand-alone grey model. However, all these different models demonstrate their own weaknesses and strengths and it is not possible to distinguish the superior approach. On the other hand, a literature review suggests that artificial intelligence is becoming more popular because of its flexibility and proven forecasting results compared to conventional statistics and time series. This could be explained in several ways. First, historical time series of MSW generation and its composition are very dynamic in nature, and therefore only nonlinear models are able to capture those patterns hidden in the data. Moreover, municipal waste generation varies from region to region, as it is influenced by many interrelated factors, such as demographic and socioeconomic factors, infrastructure, culture, location, climate, tourism, and many others. Notably, the quality, completeness and availability of MSW monitoring data and exogenous factors differ from country to country (Turcott Cervantes et al., 2018). Therefore, the choice of modelling technique and the use of exogenous factors highly depends on the country, regional situation, and development. Specifically, insufficient data in the field of MSW forecasting and management has been stressed by many studies and are seen as a major obstacle affecting the development of intelligent MSW management systems (Chen and Chang, 2000; Dyson and Chang, 2005).

This study attempts to overcome data incompleteness by incorporating relevant geographic, demographic and socioeconomic factors and contributing to the field by proposing a hybrid k-nearest neighbours (kNN) approach for forecasting MSW generation and its composition in any municipality of Lithuania. kNN was chosen as a base model because of its flexibility to incorporate different data types and ability to impute missing data. In addition, we found this technique a more transparent and understandable model for policy-decision-making in the waste management field than sophisticated black-box models. To the best of our knowledge, only a few studies have considered solid waste management problems in Lithuania. Exponential smoothing was used the 10 administrative regions of Lithuania to provide the trends of waste generation flows in general (Stankeviciene and Buzinske, 2021). Another study presented a forecasting model for MSW generation with respect to the economic activity of the city using simple conventional approaches, such as regression modelling and time series analysis (Rimaitytė et al., 2012). This study aims to contribute not only to proposing AI-based technique for forecasting MSW generation and its composition but also filling this gap and providing forecasts of MSW generation and its composition in Lithuanian municipalities. In our case, 16% of the country's regions do not provide any data on MSW generation or only provide data for the last one or two years. Notably, the H-kNN forecasting method proposed in the paper is not limited to the Lithuanian case and could be applied for MSW forecasting in other countries or regions, especially when the history of observed MSW generation is short and contains missing values.

2. Methodology and background

2.1. Study area

Lithuania has 60 municipalities, a total area of 65,300 square kilometres and a population of 2.794 million. In 2019, 472 kg of municipal waste were generated per capita, on average, in Lithuania (6% less than the EU average). Since 2016, Lithuania has initiated some practical reforms towards a circular economy. One of the most challenging, but necessary, waste management actions was the introduction of the deposit system (Tamulyte, 2021). In 2018, 92% of all packaging (glass, plastic, or aluminium) released to the market was returned for recycling. Comparatively, Lithuania has one of the highest recycling growth rates among European countries over a 15-year period (European Environment Agency, 2019). Another reform was the introduction of an electronic system for tracking waste movement from creation to transportation and elimination. Despite these initiatives, the amount of MSW per capita generated has increased, particularly in recent years. This is surprising, given that the population (one of the most important factors that impact waste quantity) has been decreasing since 1992 mainly because of high emigration rates and declining birth rates. Fig. 1 shows MSW per capita generated in different Lithuanian municipalities, highlighting comparatively significant heterogeneity in different regions, ranging from 190.46 to 913.81 kg.

Spatial analysis of the MSW generated reveals interesting insights. First, the four regions categorised as red exhibited the highest MSW per capita generation, which could be explained by the fact that these areas attract high numbers of local and foreign tourists. The waste amounts ranged from 400 to 500 and were close to the average MSW observed in Lithuania is reported in regions that are located near the largest cities (Vilnius, Kaunas, Siauliai) or comparatively densely populated areas. The MSW quantities of less than 400 kg are observed in most Lithuanian regions, with the lowest numbers collected in sparsely populated areas. Specifically, Anyksciai could be seen as our locally distinct region, which is actively visited by tourists but estimated with the minimum MSW quantity in 2019. Notably, it is important to restate that the waste collected in the country varies significantly among its local regions, even in a small country, as is the case of Lithuania. Therefore, the proposed forecasting approach should provide MSW quantities in the divisions of the country in order to enable effective MSW management.

Composition and classification by material of municipal waste observed in Lithuania are depicted in Fig. 2, where each category is

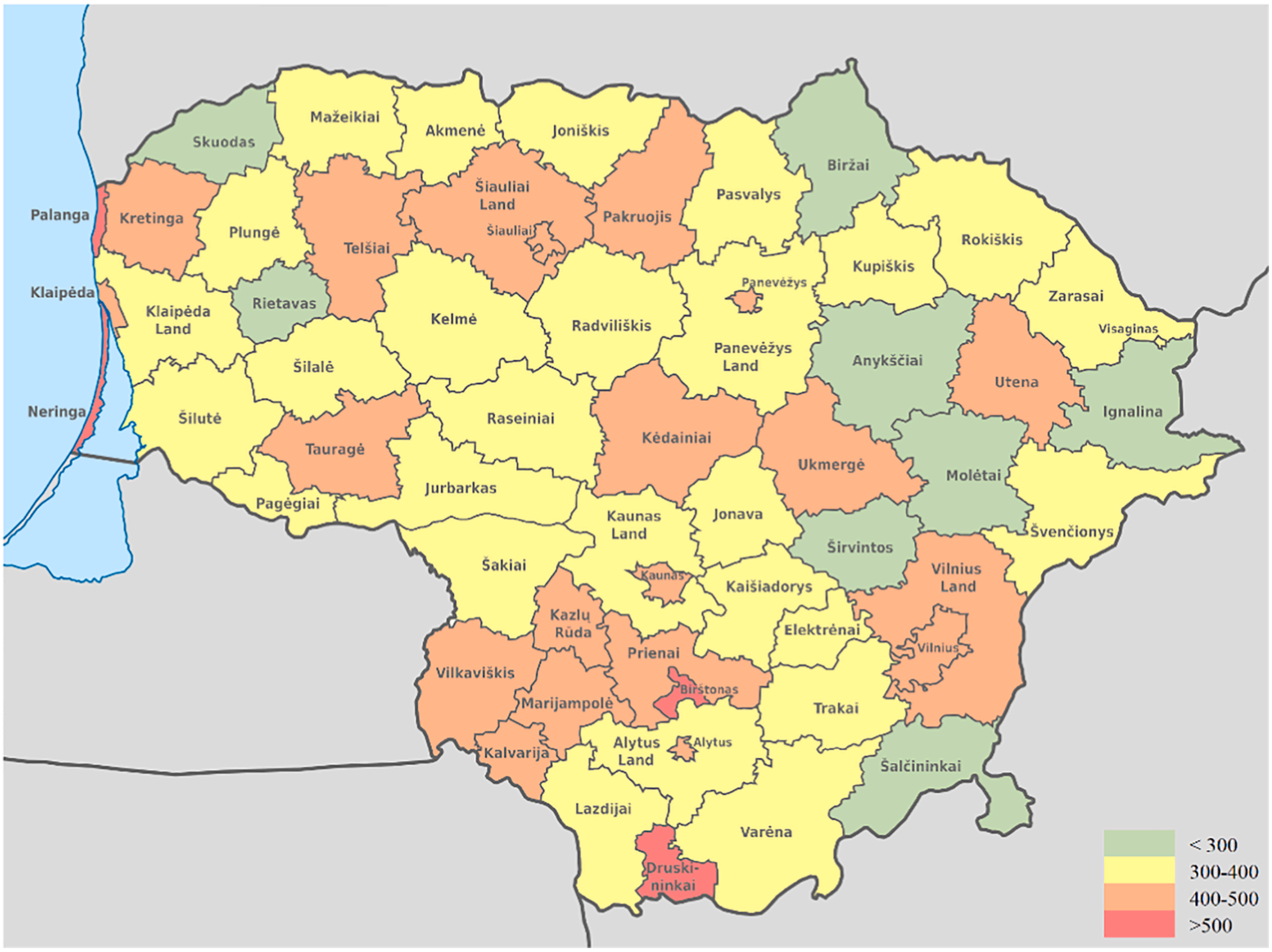


Fig. 1. Generation of MSW (kg per capita) in the municipalities of Lithuania in 2019.

given as a percentage from MSW per capita generated.

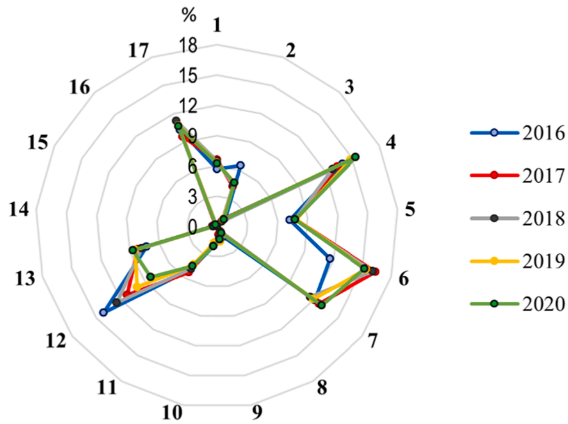
A total of 17 MSW categories are defined by the Ministry of Environment of the Republic of Lithuania. The main constituents of waste are biodegradable food and kitchen waste, other municipal biodegradable waste, textile waste, and plastics, including packaging, followed by inert waste, other non-hazardous waste, and other uncategorised municipal waste. Not individually mentioned waste categories comprising less than 8% of MSW. The heterogeneity of MSW solid waste is the major challenge in sorting and recycling. Considering the changes of composition over the considered period, it can be observed that in 2016, the share of inert waste in the total amount of solid waste was approximately 14%, which has been significantly decreased to almost 8% in 2020. Another significant change has been observed for PET packaging waste per year since 2016 due to the deposit system for bottles and cans introduced in Lithuania.

The main problem is that not all Lithuanian municipalities provide long-term data on the amount of municipal waste and its composition. Most centres, being responsible for regional waste management, store data about MSW composition mainly since 2016, which is too short period to make the forecasting waste generation using artificial intelligence methods reliable. Weekly or monthly recordings of MSW could be very useful to the development of a forecasting model, employing time series methods (e.g., ARIMA model). However, weekly observations are available only for two municipalities, i.e., Vilnius and Siauliai, without providing any details on MSW composition. Specifically, waste categories are recorded in all municipalities quarterly, but only from 2016 onwards. Thus, the lack of complete statistical information is a major issue that limits the choice of forecasting approach.

2.2. MSW generation in the context of socioeconomic and demographic factors

To assess the relation between the generation of municipal waste and exogenous factors, such socioeconomic and demographic indicators, a list of possible factors were identified based on the literature review (Araiza-Aguilar1 et al., 2020; Intharathirat et al., 2015; Lin et al., 2021; Liu et al., 2019). These factors included population, living and working, tourism, economic developments, production, and services. The statistical data were retrieved from the webpage of the Department of Statistics of Lithuania (STD, Statistics Lithuania, 2021) and summarised in Appendix A. The final dataset was limited by data availability and access, which restricted us from taking full advantage of these factors potential in forecasting MSW generation. Correlation analysis was used to estimate the strength of the relationship between selected factors and waste levels measured in terms of the country's MSW per capita and total MSW to identify those factors particularly related to the waste quantities generated in Lithuania (see Fig. 3).

Notably, Fig. 3 reveals that many factors, with some exceptions, strongly positively correlate with MSW per capita but negatively with total MSW quantities. This could be explained by the fact that there were several waves of emigration in Lithuania during the considered period; therefore, the population fluctuated significantly, and the general trends show a consistent decrease of Lithuanian population. Consequently, this was the reason to eliminate total MSW from the following analysis. Another finding from Fig. 3 suggests that factors such as public utilities, unemployment rate, and population density, negatively correlate with MSW per capita. In comparison, the inflation rate and agricultural



No.	Abbreviation	Description
1	Paper	Paper and cardboard, including packaging
2	Green	Green waste
3	Wood	Wood waste, including packaging
4	Bio	Biodegradable food and kitchen waste
5	Textile	Textile waste
6	Biodegr	Other municipal biodegradable waste
7	Plastic	Plastics, including packaging
8	PET	PET packaging waste
9	Comb	Combined packaging waste
10	Metal	Wastes from metals, including packaging
11	Glass	Glass waste, including packaging
12	Inert	Inert wastes (ceramics, concrete, stones, etc.)
13	O_nonhazard	Other non-hazardous waste accidentally accepted at the RNHWL facility
14	O_elec	Other non-hazardous waste, such as electronic equipment accidentally accepted into a RNHWL facility
15	O_bat	Other non-hazardous waste, such as battery and accumulators accidentally accepted at the RNHWL facility
16	O_hazard	Other hazardous waste accidentally accepted at the RNHWL and MSW facility
17	O_mun	Other municipal wastes (e.g., hygienic waste, footwear, rubber)

Note: RNHWL - regional non-hazardous waste landfill

Fig. 2. Distribution of MSW by waste category over 2016–2020.

production are examples of non-informative factors that do not significantly explain either total MSW, nor MSW per capita. However, these findings are not necessarily true for the separate MSW categories in the composition, since the composition has a tendency change in time to a greater extent than socioeconomic factors. This could be revealed by estimating the relationship between MSW per capita and its categories (see Table 1).

Table 1 shows that the correlation coefficient ranges from −0.8588 (glass waste, including packaging) to 0.8012 (biodegradable food and kitchen waste), which is an interesting finding. Strong positive relationship is also estimated for categories, such as combined packaging waste and other non-hazardous waste, while the second largest negative correlation is observed for inert waste. In cases when the estimated coefficient is close to 0, we could infer that the generation of a certain MSW category fluctuated over time by changing the direction against the quantities of MSW. In summary, the evidence shows that it is not enough to provide forecasts for MSW per capita in general but, more importantly, to estimate future MSW levels for each category in the composition separately.

2.3. Hierarchical kNN approach for MSW forecasting

Historically, there were cases when data on the generated MSW quantities were not registered in a certain municipality of Lithuania. To solve this problem, missing values could be determined based on neighbouring regions. For this purpose, we propose the approach H-kNN, which is based on k-nearest-neighbour (kNN) with hierarchical setting. In the literature, there have been several similar attempts to modify kNN by introducing several levels for solving prediction problems from other areas. For example, the authors proposed a diagnostic system, which is based on the hierarchical structure of kNN classifiers, for detecting the onset of degradation and classifying the type of defect (Baraldi et al., 2016). Another example comes from medical image analysis, where the classification of skin lesions has been investigated using a hierarchical decision tree, with a different kNN trained for each decision node (Fisher et al., 2020). This suggests that the modification of classical kNN allows improving the prediction results and still keeps the functionality to impute missing values if needed.

Let us briefly introduce the classical kNN approach, focusing on its regression formulation (Song et al., 2017a). Suppose we have observational pairs $(X_1, Y_1), \cdots, (X_n, Y_n)$, where $X \in R^d$ is a set of variables (or features) that describe the instance, and $Y_i \in R$ is a target variable, which is under consideration for a particular problem. For every new instance, the value of Y is forecasted based on the most similar observations defined as k nearest neighbours. To find the similarities among neighbours, it is necessary to compute the distance between pairs of

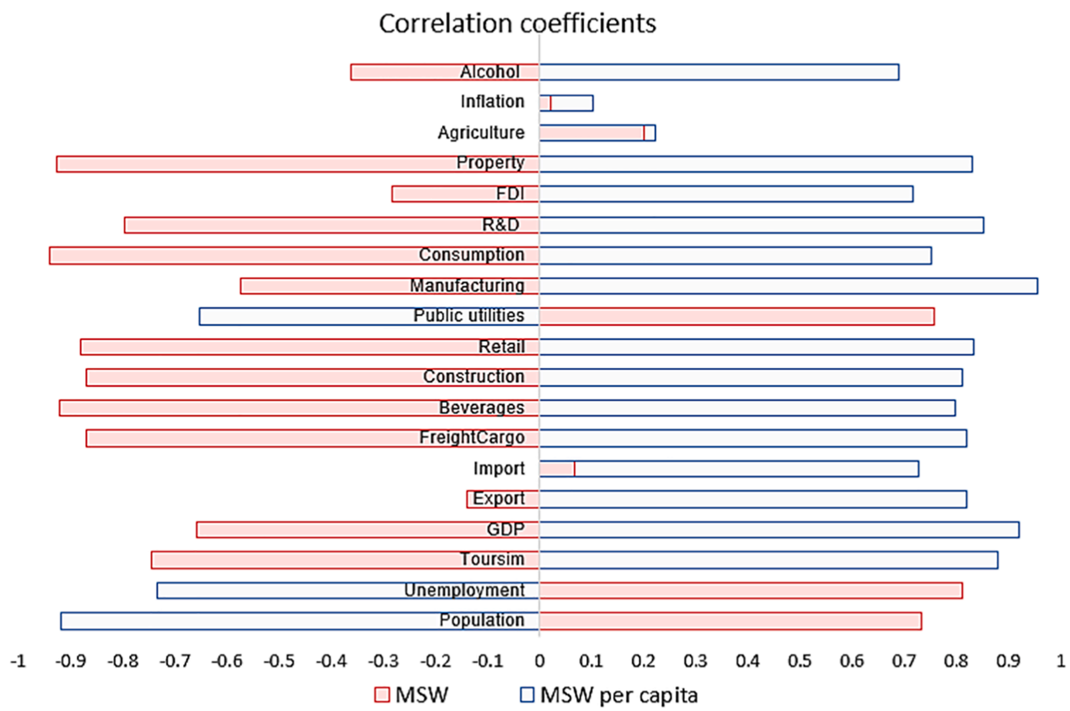


Fig. 3. Graphical representation of the correlation between municipal waste levels and exogenous factors.

Table 1
Correlation analysis between MSW per capita and its composition.

MSW category	Correlation coefficient
Paper	0.0329
Green	−0.2327
Wood	−0.2959
Bio	0.8012
Textile	0.3635
Biodegr	0.1945
Plastic	0.2215
PET	−0.0886
Comb	0.6648
Metal	0.2757
Glass	−0.8588
Inert	−0.7886
O_nonhazard	0.7077
O_elec	−0.3157
O_bat	−0.5334
O_hazard	−0.3586
O_mun	0.5773

observations. For real-valued observations, the most popular similarity measure is Euclidean distance. Alternatively, other examples of commonly used distance measures include Manhattan, Hamming, Minkowski or cosine metrics (Chomboon et al., 2015). Using this approach, the multiple neighbours k based on their similarity are combined to obtain forecasts of Y by taking the average of their values. For a MSW forecasting problem, X defines any relevant information that could be attributed to the municipality waste generation, while Y describes the amount of MSW per capita generated.

In this paper, the hierarchical implementation of kNN includes several iterations with calculations performed at different depth levels (see Fig. 4). Suppose L defines a set of hierarchical levels. In the first hierarchical level L_1, similar neighbours, say k_1, are selected based on values of X_i, for example, geographical distance. Then, in the second level L_2, the number of neighbours, suppose $k_2 \leq k_1$, are chosen from the candidates included in the previous level but considering other variables, for example, population density. In the same manner, the number of similar neighbours $k_3 \leq k_2 \leq k_1$ are selected in the third level L_3 using other variables. In general, the depth of the hierarchical scheme is unbounded and mainly depends how many variables are included in the dataset and how they are allocated among different levels.

As an alternative, instead of using historical values of multiple exogenous factors, one could consider including the meaningful principal components defined as a linear combination of optimally weighted observed variables.

In further analysis, the allocation of variables X among hierarchical levels and the selection of k nearest neighbours should be determined. For this purpose, commonly used error metrics, such as the mean absolute percentage error (MAPE), mean absolute error (MAE), and root mean square error (RMSE), could be chosen (Hyndman and Athanasopoulos, 2021). Specifically, the choice of MAPE should be carefully considered since this metric is skewed and does not exist for zero observations (Hyndman and Koehler, 2006).

For this particular study, we implemented a scheme of H-kNN with two hierarchical levels arranging variables as follows:

L_1 level: Straight geographical distance, D, between two municipalities;

L_2 level: Demographic and socioeconomic factors that correlate with MSW per capita and are available for each municipality: population density, GDP per capita, private property, foreign investment per capita, and tourism.

3. Results and discussion

3.1. MSW forecasting using kNN and H-kNN methods

Experiments with the selected Lithuanian municipalities and their known amount of annual MSW quantities have been performed to evaluate the performance of the proposed H-kNN by estimating forecasting error. For the demonstration purposes, let us focus on three municipalities, Varena, Druskininkai and Kaunas, with the aim of determining the amount of municipal waste according to the described methodology. As it has been already presented (see Section 2.3), in the first hierarchical level L_1, the geographical distance between the

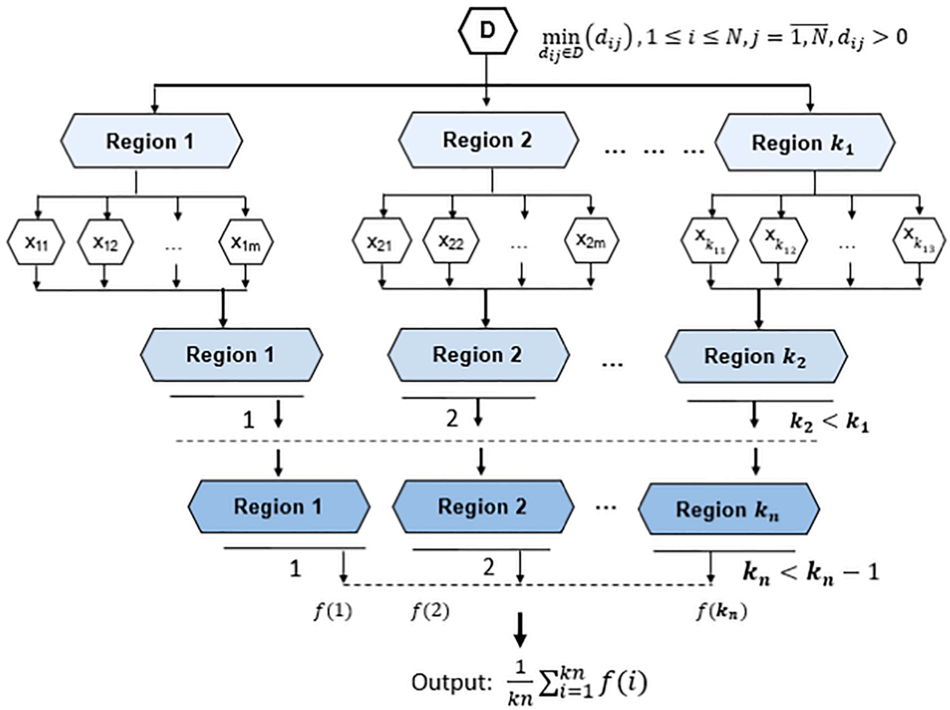


Fig. 4. The general scheme of the H-kNN algorithm for MSW prediction.

selected region and its neighbouring municipalities is computed. For example, Fig. 5 demonstrates the case of Varena, where nine nearest geographical neighbours are considered, i.e., $k_1 = 9$.

In the second hierarchical level L_2, the number k_2 of nearest neighbours is determined from k_1 neighbours, considering socioeconomic and demographic factors. The identified municipalities are highlighted in the table displayed in Fig. 5, when $k_2 = 3$. Therefore, the forecast value of MSW quantity in Varena municipality is determined by averaging MSW quantities observed in the regions of Alytus, Elektrenai and Lazdijai identified by H-kNN approach, i.e.

$$MSW_{Varena} = \frac{402 + 304 + 292}{3} \approx 333kg/capita$$

The total MSW for Varena is calculated by multiplying MSW by the population size in the region, i.e.,

$k_1 = 9$

Municipalities	Distance (D)
Alytus Land	28
Alytus	40
Druskininkai	45
Trakai	52
Salcininkai	53
Birstonas	56
Prienai	62
Elektrenai	62
Lazdijai	69
Vilnius	70
Kaisiadorys	72

$k_2 = 3$

Municipalities
Alytus
Elektrenai
Lazdijai

Municilapities	MSW kg per capita
Alytus	402.1259
Elektrenai	303.5763
Lazdijai	291.8812

Fig. 5. The radius of nine nearest neighbours of Varena municipality.

$$TotalMSW_{Varena} = MSW_{Varena} \times POP_{Varena} = 333 \times 21,811 \approx 7,252.158 tons$$

Comparing the actual level of MSW in Varena municipality, which is 7,574.256 tons, and the predicted value, i.e., 7,252.158 tons, using the H-kNN, we obtain error deviation of 4.44%.

The analysis of error sensitivity in terms of MAPE to the number of nearest neighbours is demonstrated for the selected municipalities, such as Varena, Druskininkai and Kaunas, using the classical kNN with $k = \overline{1,9}$ and the proposed approach H-kNN, considering the number of neighbours $k_1 = \overline{1,9}$ and $k_2 = \overline{1,9}$, respectively (see Table 2). The last column of the table shows MAPE values averaged over all k.

From the results provided in Table 2, it can be determined that the average value of MAPE for Varena municipality resulted in 22.43% using kNN, which on average has been reduced to 14.89% using H-kNN approach. Comparatively, the lowest observed forecasting error for kNN is 2.322% with $k = 1$, which could be too risky to rely on the information of a single nearest neighbour. The smallest error observed for H-kNN is MAPE = 0.11% when $k_2 = 4$. Similar conclusions could be drawn for other considered municipalities. The results in Table 2 suggest that MAPE values significantly depend on the number of neighbours to be chosen, therefore this parameter should be carefully considered. More specifically, starting from $k_2 = 6$ the error tends to increase significantly. Thus, it could be summarised that the optimal number of nearest neighbours may vary for different municipalities and should be chosen based on error minimisation. This case shows that H-kNN approach, in comparison to kNN, reduced error estimates by 7.54, 14.57, and 15.92 percentage points for Varena, Druskininkai and Kaunas, respectively.In the similar manner, it possible to forecast future values of MSW composition using H-kNN approach. In this case, some time series model (e.g. ARIMA) is used to forecast future values of socioeconomic or demographic factors, and then H-kNN is used to forecast future MSW values. In principle, another advantage of the proposed approach is highlighted here, showing that the method can be applied considering different scenarios of socioeconomic development. For demonstration purposes, the obtained results are provided for the same municipalities in Fig. 6, when $k_1 = 9$ and $k_2 = 3$.

It is obvious from Fig. 6 that the forecasts of MSW composition are provided with high accuracy for many waste categories, which suggests a good performance of H-kNN approach for this particular problem. Nevertheless, the errors vary to some extent for some specific categories among municipalities. This could be explained by the different MSW compositions observed in neighbouring regions. For example, the true percentage of Biodegr category is observed comparatively different, i.e., 0.01%, 1.00%, and 24.05% in Druskininkai, Varena, and Kaunas, respectively. This might be a reason for worse forecasting accuracy for this category, resulting from the H-kNN approach. Fig. 6 also depicts forecasts of MSW composition for two-year ahead. It can be seen that some MSW categories are subject to increase, while others are expected to decrease. However, we can observe the categories, for example, Wood, Textile or Methal, where those tendencies differ among municipalities. All experiments with the H-kNN model were implemented in Python and run on a PC (Dell Allienware) with NVidia GeForce GTX 980 M with 4 GB GDDR5 graphics and 4 GB of GDDR5 standard memory. Average model training time 0.934 s.

To summarise, after analysing the forecasting results of MSW and, particularly, its composition for each Lithuanian municipality, we claim that the inaccuracies could be explained by the changing behaviour of residents. For instance, the forecasted value of recycling MSW categories, such as paper and plastic, including packaging, was up to twice as high as actually observed. Presumably, this could be caused by the fact that the Lithuanian population is becoming increasingly involved in the waste sorting process, and, consequently, reducing the waste amounts in landfill.

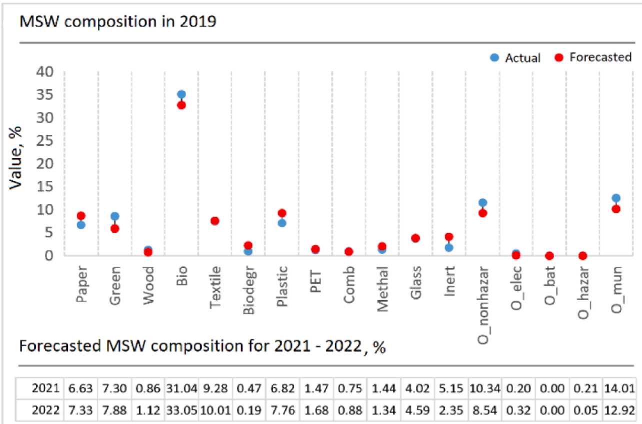
3.2. Discussion and recommendations

The results provided in the study are obtained including relevant exogenous factors, such as population density, GDP per capita, private property, foreign investment per capita, and tourism. The inclusion of other possible factors was limited by their availability and accessibility. It is no surprise that population has been identified as the most important driving factor for MSW generation, i.e., the higher the population is, the larger solid waste is accumulated, which is in line with other studies (Chhay et al., 2018; Kawai and Tasaki, 2016; Younes et al., 2015). Other than population, the list of relevant factors could be expanded by adding more data that represent globalisation, rapid urbanization, booming economy, and living standards. For example, the study of Turcott Cervantes et al. (2018) reported forty sets of variables could impact municipal solid waste management. The final list included 377 different factors, with 49.3% of them linked to technical aspects of waste management. Additionally, the authors highlighted a number of data problems (such as scarce or non-existent information, lack of transparency and homogeneity) that restrict the use of models from being applied globally. Lithuania is no exception with a comparatively small number of exogenous factors linked to MSW generation rates in this study. In order to expand the number of factors included, we suggest that relevant information is collected at the earliest possible moment. In addition, future research could focus how proxy indicators could be used to describe the development of each municipality and enhance forecasts of MSW and, more importantly, its composition.

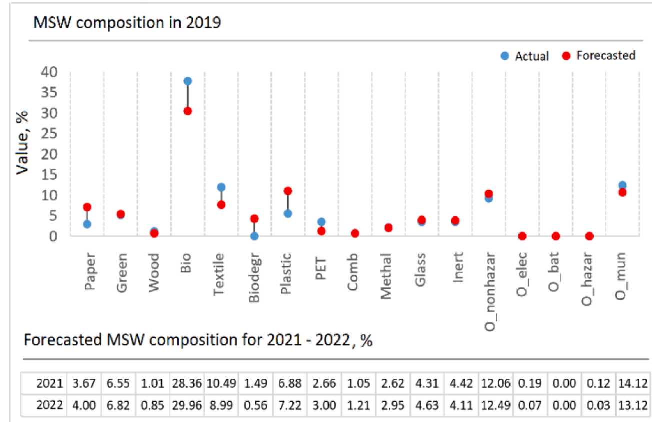
In terms of the forecasting model itself, we were restricted to demonstrate the flexibility of an H-kNN algorithm in terms of selection of the optimal number of nearest neighbours in different hierarchical levels or the number of levels in general by performing a cross-validation technique. The limitation comes from a comparatively short history of MSW generation data collected in the regions of Lithuania. We recommend carefully choosing the optimal number of nearest neighbours k based on the literature and adopting these for the H-kNN algorithm (Ghosh, 2006; Tembusai et al., 2021). Another recommendation is that only a small number of neighbours in the H-kNN algorithm should be selected if neighbouring municipalities are heterogeneous in terms of MSW generation; otherwise, it negatively affects the forecasting accuracy.

Table 2
MAPE of forecasted MSW quantities in different Lithuanian municipalities using kNN and H-kNN approaches, in percentage value.

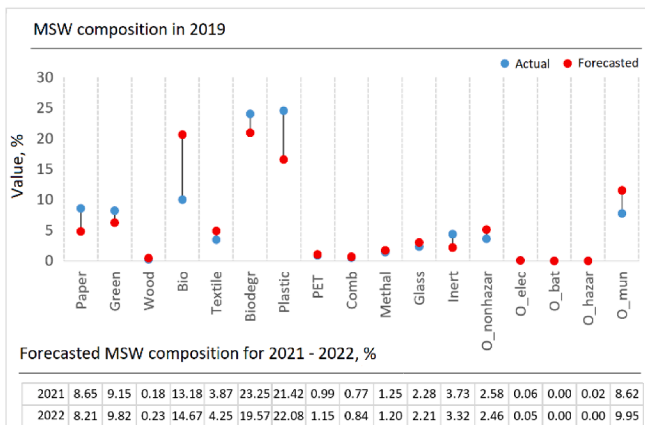
	kNN									
k	1	2	3	4	5	6	7	8	9	Average
Varena	2.32	6.74	25.47	33.65	19.65	35.88	28.96	26.99	22.22	22.43
Druskininkai	40.05	44.23	38.65	39	35.8	24.3	25.19	25.19	24.28	32.97
Kaunas	32.68	8.23	8.64	15.69	15.02	35.08	32.75	28.01	28.91	22.78
	H-kNN									
k_2	1	2	3	4	5	6	7	8	9	Average
Varena	13.17	0.3	4.25	0.11	12.67	20.26	31.49	29.53	22.22	14.89
Druskininkai	2.91	20.77	0.08	9.72	27.13	30.59	30.35	22.32	21.77	18.40
Kaunas	2.2	6.99	14.23	10.12	1.56	2.26	6.49	9.36	8.51	6.86



a) Varena municipality



b) Druskininkai municipality



c) Kaunas municipality

Fig. 6. MSW composition forecasts for testing period (2019 year) and two-year ahead in three Lithuanian regions.

4. Conclusions

In this study, a hierarchical kNN approach is proposed to forecast MSW generation rates and composition in Lithuanian municipalities. Previous research has shown that MSW generation and composition is dependent on many exogenous factors and, therefore, the inclusion of these factors could be beneficial for providing waste generation forecasts. We expanded the kNN approach with hierarchical setting by first including spatial information and then socioeconomic factors that are observed regionally. Notably, the use of exogenous factors could be arranged in different hierarchical levels depending on the application domain. In fact, the choice of kNN as a base model is motivated by its transparent nature and ability to deal with missing values, as is the case in Lithuania and many other countries. Overall, In terms of MAPE, the H-kNN model outperformed kNN by 18.22%, on average, including all Lithuanian municipalities.

Funding

This work was supported by the Lithuanian Business Support Agency (Grant agreement number 01.2.1-LVPA-V-835–03-0005).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.wasman.2022.01.004.

References

Abbasi, M., Abduli, M., Omidvar, B., Baghvand, A., 2013. Forecasting Municipal Solid Waste Generation by Hybrid Support Vector Machine and Partial Least Square Model. International Journal of Environmental Research 7, 27–38. https://doi.org/10.22059/ijer.2012.583.

Abdallah, M., Abu Talib, M., Feroz, S., Nasir, Q., Abdalla, H., Mahfood, B., 2020. Artificial Intelligence Applications in Solid Waste Management: A Systematic Research Review. Waste Management 109, 231–246. https://doi.org/10.1016/j.wasman.2020.04.057.

Abdel-Shafy, H.I., Mansour, M.S.M., 2018. Solid Waste Issue: Sources, Composition, Disposal, Recycling, and Valorization. Egyptian Journal of Petroleum 27 (4), 1275–1290. https://doi.org/10.1016/j.ejpe.2018.07.003.

Adamović, V.M., Antanasijević, D.Z., Ristić, M.Đ., Perić-Grujić, A.A., Pocajt, V.V., 2017. Prediction of Municipal Solid Waste Generation Using Artificial Neural Network Approach Enhanced by Structural Break Analysis. Environmental Science and Pollution Research 24 (1), 299–311.

Adeleke, O., Akinlabi, S.A., Jen, T.C., Dunmade, I., 2020. Prediction of Municipal Solid Waste Generation: An Investigation of the Effect of Clustering Techniques and Parameters on ANFIS Model Performance. Environmental Technology 1–14. https://doi.org/10.1080/09593330.2020.1845819.

Akanbi, L.A., Oyedele, A.O., Oyedele, L.O., Salami, R.O., 2020. Deep Learning Model for Demolition Waste Prediction in a Circular Economy. Journal of Cleaner Production 274, 122843. https://doi.org/10.1016/j.jclepro.2020.122843.

Alberdi, E., Urrutia, L., Goti, A., Oyarbide-Zubillaga, A., 2020. Modeling the Municipal Waste Collection Using Genetic Algorithms. Processes 8 (5), 513. https://doi.org/10.3390/pr8050513.
Ali, S.A., Ahmad, A., 2019. Forecasting MSW Generation Using Artificial Neural Network Time Series Model: A Study from Metropolitan City. SN Applied Sciences 1, 1338. https://doi.org/10.1007/s42452-019-1382-7.
Antanasijević, D., Pocajt, V., Popović, I., Redžić, N., Ristić, M., 2013. The Forecasting of Municipal Waste Generation Using Artificial Neural Networks and Sustainability Indicators. Sustainability Science 8 (1), 37–46. https://doi.org/10.1007/s11625-012-0161-9.
Araiza-Aguilar1, J., Rojas-Valencia, M., Aguilar-Vera, R., 2020. Forecast Generation Model of Municipal Solid Waste Using Multiple Linear Regression. Global Journal of Environmental Science and Management 6, 1–14.
Ayeleru, O.O., Fajimi, L.I., Oboirien, B.O., Olubambi, P.A., 2021. Forecasting Municipal Solid Waste Quantity Using Artificial Neural Network and Supported Vector Machine Techniques: A Case Study of Johannesburg, South Africa. Journal of Cleaner Production 289, 125671. https://doi.org/10.1016/j.jclepro.2020.125671.
Azarmi, S., Oladipo, A., Vaziri, R., Alipour, H., 2018. Comparative Modelling and Artificial Neural Network Inspired Prediction of Waste Generation Rates of Hospitality Industry: The Case of North Cyprus. Sustainability 10 (9), 2965. https://doi.org/10.3390/su10092965.
Baraldi, P., Cannarile, F., Di Maio, F., Zio, E., 2016. Hierarchical K-Nearest Neighbours Classification and Binary Differential Evolution for Fault Diagnostics of Automotive Bearings Operating Under Variable Conditions. Engineering Applications of Artificial Intelligence 56, 1–13. doi:/10.1016/j.engappai.2016.08.011.
Cha, G.-W., Moon, H.J., Kim, Y.-M., Hong, W.-H., Hwang, J.-H., Park, W.-J., Kim, Y.-C., 2020. Development of a Prediction Model for Demolition Waste Generation Using a Random Forest Algorithm Based on Small Datasets. International Journal of Environmental Research and Public Health 17 (19), 6997. https://doi.org/10.3390/ijerph17196997.
Chamas, A., Moon, H., Zheng, J., Qiu, Y., Tabassum, T., Jang, J.H., Abu-Omar, M., Scott, S.L., Suh, S., 2020. Degradation Rates of Plastics in the Environment. ACS Sustainable Chemistry & Engineering 8 (9), 3494–3511. https://doi.org/10.1021/acssuschemeng.9b06635.
Chen, D.-C., Bodirsky, B.L., Krueger, T., Mishra, A., Popp, A., 2020. The World's Growing Municipal Solid Waste: Trends and Impacts. Environmental Research Letters 15 (7), 074021. https://doi.org/10.1088/1748-9326/ab8659.
Chen, H.W., Chang, N.-B., 2000. Prediction Analysis of Solid Waste Generation Based on Grey Fuzzy Dynamic Modeling. Resources, Conservation and Recycling 29 (1-2), 1–18. https://doi.org/10.1016/S0921-3449(99)00052-X.
Chhay, L., Reyad, M.A.H., Suy, R., Islam, M.R., Mian, M.M., 2018. Municipal Solid Waste Generation in China: Influencing Factor Analysis and Multi-Model Forecasting. Journal of Material Cycles and Waste Management 20 (3), 1761–1770. https://doi.org/10.1007/s10163-018-0743-4.
Chomboon, K., Chujai, P., Teerarassamee, P., Kerdprasop, K., Kerdprasop, N., 2015. An Empirical Study of Distance Metrics for K-Nearest Neighbor Algorithm. In Proceedings of the 3rd International Conference on Industrial Application Engineering, The Institute of Industrial Applications Engineers, Japan. pp. 280–285. doi:10.12792/iciae2015.051.
Dai, F., Nie, G.-H., Chen, Y.i., 2020. The Municipal Solid Waste Generation Distribution Prediction System Based on FIG-GA-SVR Model. Journal of Material Cycles and Waste Management 22 (5), 1352–1369. https://doi.org/10.1007/s10163-020-01022-5.
Das, S., Lee, S.H., Kumar, P., Kim, K.H., Lee, S.S., Bhattacharya, S.S., 2019. Solid Waste Management: Scope and the Challenge of Sustainability. Journal of Cleaner Production 228, 658–678. https://doi.org/10.1016/j.jclepro.2019.04.323.
Dyson, B., Chang, N.-B., 2005. Forecasting Municipal Solid Waste Generation in a Fast-Growing Urban Region with System Dynamics Modeling. Waste Management 25 (7), 669–679. https://doi.org/10.1016/j.wasman.2004.10.005.
European Environment Agency, 2019. Municipal Waste Recycling Rates in Europe by Country. https://www.eea.europa.eu/ims/waste-recycling-in-europe. Accessed: November 23, 2021.
Eurostat, 2021. Municipal Waste Statistics. https://ec.europa.eu/eurostat/statistics-explained. Accessed: July 5, 2021.
Fisher, R.B., Rees, J., Bertrand, A., 2020. Classification of Ten Skin Lesion Classes: Hierarchical kNN versus Deep Net. In Medical Image Understanding and Analysis, Zheng, Y., Williams, B.M., Chen, K. (Eds.), Springer International Publishing, Cham. pp. 86–98.
Ghosh, A.K., 2006. On Optimum Choice of k in Nearest Neighbor Classification. Computational Statistics & Data Analysis 50 (11), 3113–3123. https://doi.org/10.1016/j.csda.2005.06.007.
Huang, L.i., Cai, T., Zhu, Y.a., Zhu, Y., Wang, W., Sun, K., 2020. LSTM-based Forecasting for Urban Construction Waste Generation. Sustainability 12 (20), 8555. https://doi.org/10.3390/su12208555.
Hyndman, R., Athanasopoulos, G., 2021. Forecasting: Principles and Practice. 3rd edition, OTexts: Melbourne, Australia. Accessed: July 29, 2021.
Hyndman, R.J., Koehler, A.B., 2006. Another Look at Measures of Forecast Accuracy. International Journal of Forecasting 22 (4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001.
Intharathirat, R., Abdul Salam, P., Kumar, S., Untong, A., 2015. Forecasting of Municipal Solid Waste Quantity in a Developing Country Using Multivariate Grey Models. Waste Management 39, 3–14. https://doi.org/10.1016/j.wasman.2015.01.026.
Kawai, K., Tasaki, T., 2016. Revisiting Estimates of Municipal Solid Waste Generation per Capita and their Reliability. Journal of Material Cycles and Waste Management 18 (1), 1–13.
Kaza, S., Yao, L.C., Bhada-Tata, P., Woerden, F.V., 2018. What a Waste 2.0: A Global Snapshot of Solid Waste Management to 2050. Technical Report. World Bank. URL: http://hdl.handle.net/10986/30317.
Khandelwal, H., Dhar, H., Thalla, A.K., Kumar, S., 2019. Application of Life Cycle Assessment in Municipal Solid Waste Management: A Worldwide Critical Review. Journal of Cleaner Production 209, 630–654. https://doi.org/10.1016/j.jclepro.2018.10.233.
Koop, S.H.A., van Leeuwen, C.J., 2017. The Challenges of Water, Waste and Climate Change in Cities. Environment, Development and Sustainability 19 (2), 385–418. https://doi.org/10.1007/s10668-016-9760-4.
Lee, P., Sims, E., Bertham, O., Symington, H., Bell, N., Pfaltzgraff, L., Sjögre, P., 2017. Towards a Circular Economy - Waste Management in the EU. Technical Report. European Parliamentary Research Service.
Lin, K., Zhao, Y., Tian, L., Zhao, C., Zhang, M., Zhou, T., 2021. Estimation of Municipal Solid Waste Amount Based on One-Dimension Convolutional Neural Network and Long Short-Term Memory with Attention Mechanism Model: A Case Study of Shanghai. Science of The Total Environment 791, 148088. https://doi.org/10.1016/j.scitotenv.2021. 148088.
Liu, J., Li, Q., Gu, W., Wang, C., 2019. The Impact of Consumption Patterns on the Generation of Municipal Solid Waste in China: Evidences from Provincial Data. International Journal of Environmental Research and Public Health 16 (10), 1717. https://doi.org/10.3390/ijerph16101717.
OECD, 2020. Environment at a glance: Climate Change. https://www.oecd.org/environment/environment-at-a-glance/Circular-Economy-Waste-Materials-Archive-January-2020.pdf. Accessed: June 18, 2021.
Oguz-Ekim, P., 2021. Machine Learning Approaches for Municipal Solid Waste Generation Forecasting. Environmental Engineering Science 38 (6), 489–499. https://doi.org/10.1089/ees.2020.0232.
Oliveira, V., Sousa, V., Dias-Ferreira, C., 2019. Artificial Neural Network Modelling of the Amount of Separately-Collected Household Packaging Waste. Journal of Cleaner Production 210, 401–409. https://doi.org/10.1016/j.jclepro.2018.11.063.
Popli, K., Park, C., Han, S.-M., Kim, S., 2021. Prediction of Solid Waste Generation Rates in Urban Region of Laos Using Socio-Demographic and Economic Parameters with a Multi Linear Regression Approach. Sustainability 13 (6), 3038. https://doi.org/10.3390/su13063038.
Rimaitytė, I., Ruzgas, T., Denafas, G., Račys, V., Martuzevicius, D., 2012. Application and Evaluation of Forecasting Methods for Municipal Solid Waste Generation in an Eastern-European City. Waste Management & Research 30 (1), 89–98. https://doi.org/10.1177/0734242X10396754.
Solano Meza, J.K., Orjuela Yepes, D., Rodrigo-Ilarri, J., Cassiraga, E., 2019. Predictive Analysis of Urban Waste Generation for the City Of Bogotá, Colombia, through the Implementation of Decision Trees-Based Machine Learning, Support Vector Machines and Artificial Neural Networks. Heliyon 5 (11), e02810. https://doi.org/10.1016/j.heliyon.2019.e02810.
Song, Y., Liang, J., Lu, J., Zhao, X., 2017a. An Efficient Instance Selection Algorithm for k Nearest Neighbor Regression. Neurocomputing 251, 26–34. https://doi.org/10.1016/j.neucom.2017.04.018.
Song, Y., Wang, Y., Liu, F., Zhang, Y., 2017b. Development of a Hybrid Model to Predict Construction and Demolition Waste: China as a Case Study. Waste Management 59, 350–361. https://doi.org/10.1016/j.wasman.2016.10.009.
Soni, U., Roy, A., Verma, A., Jain, V., 2019. Forecasting Municipal Solid Waste Generation Using Artificial Intelligence Models — a Case Study in India. SN Applied Sciences 1, 162. https://doi.org/10.1007/s42452-018-0157-x.
Stankeviciene, J., Buzinske, J., 2021. Trends of Municipal Waste Flows, Composition, Treatment in Lithuania and its Regions. In International Scientific Conference "Contemporary Issues in Business, Management and Economics Engineering", pp. 1–10. doi:10.3846/ cibmee.2021.599.
STD, Statistics Lithuania. URL: https://www.stat.gov.lt/en. Accessed: March 30, 2021.
Tamulyte, G., 2021. Waste Management in Lithuania. https://wilderness-society.org/waste-management-in-lithuania. Accessed: August 12, 2021.
Tembusai, Z.R., Mawengkang, H., Zarlis, M., 2021. K-Nearest Neighbor with K-Fold Cross Validation and Analytic Hierarchy Process on Data Classification. International Journal of Advances in Data and Information Systems 2, 1–8. https://doi.org/10.25008/ijadis.v2i1.1204.
Turcott Cervantes, D.E., López Martínez, A., Cuartas Hernández, M., García, L., de Cortázar, A., 2018. Using Indicators as a Tool to Evaluate Municipal Solid Waste Management: A Critical Review. Waste Management 80, 51–63. https://doi.org/10.1016/j.wasman.2018.08.046.
Younes, M.K., Nopiah, Z.M., Basri, N.E.A., Basri, H., Abushammala, M.F.M., Maulud, K.N.A., 2015. Prediction of Municipal Solid Waste Generation Using Nonlinear Autoregressive Network. Environmental Monitoring and Assessment 187, 753. https://doi.org/10.1007/s10661-015-4977-5.