ANKARA SCIENCE UNIVERSITY, RESEARCHER
Vol. 3, No. 1 July 2023
e-ISSN: 2717-9494
Research Article/Araştırma Makalesi
Doi: https://doi.org/10.55185/researcher.1208237

Estimating Medical Waste Generation Utilizing Penalized Regression Models

Burcu DEVRİM İÇTENBAŞ[1,*]
[1]Ankara Science University, Faculty of Engineering and Architecture, Department of Industrial Engineering, Ankara, Türkiye; ORCID: 0000-0002-8148-4945
* Corresponding Author: burcu.ictenbas@ankarabilim.edu.tr

Received: 21 November 2022; Accepted: 27 December 2022

Reference/Atıf: B.D. İçtenbaş, "Estimating Medical Waste Generation Utilizing Penalized Regression Models", Researcher, vol.3, no.1, pp.13-18, July 2023, doi: 10.55185/researcher.1208237

Abstract

Medical Waste (MW) amount that has a significant impact on health and environment is increasing as a result of industrialization as well as population density. There is a need an accurate estimation waste generation amount that will be useful information to select the appropriate disposal methods and to organize the recycling and storage. Some researchers have applied conventional statistical algorithms and many kinds of Machine Learning (ML) algorithms to predict MW amount. However, to the best of our knowledge, penalized regression methods such as Ridge, Lasso, and Elastic Net regressions have not been used to predict the MW amount. 18-years real data were obtained from İstanbul Metropolitan Municipality Department Open Data Portal with the input variables namely number of hospitals, number of health personal, number of bed available at the hospital, crude birth rate and gross domestic product per capita. 80% of the total database being used for developing the models, whereas the rest 20% were used to validate the models. In order to compare their performances, 5-fold cross-validation was applied and performance measures (MAE, RMSE and R-squared) were calculated in this study. Of the penalized regression methods, the Lasso regression provided better performance than those of other models with RMSE, MAE, and R-squared of 349.56, 596.52, 0.96, respectively, whereas the second-best Ridge regression poorer accuracy with RMSE, MAE, and R-squared 1039.091, 878.25,0.88, respectively. Thus, in our case, Lasso regression can be considered better than the Ridge regression and Elastic Net regression due to the lowest RMSE and MAE values and highest R-squared. The results reveal that the proposed Lasso regression is better than the other penalized regression models to predict the MW amount.

Keywords: medical waste, penalized regression models, prediction, sustainability

1. Introduction

Medical Waste (MW) amount is increasing as a result of industrialization as well as population density [1,2]. All the medical institutions from the hospitals to veterinary centers cause the MW. The management of the MW is a critical problem that it may yield public health risks and environmental pollution risks since it is accepted as a hazardous waste type [2,3,4]. Selecting appropriate disposal methods and organizing recycling and storage is a prominent issue especially in developing countries so there is a need an accurate estimation of this type of waste generation that will be useful information for these processes [1,5-8].

There have been several studies to estimate the MW generation such as Multiple Linear Regression [1,9-13], time series methods [1,14] and machine learning algorithms [2,3,15,16]. Multiple Linear Regression (MLR) is the most commonly methods by the researchers to estimate the MW generation. Their models reached higher model performance (R-squared>0.80) using the critical key factors number of hospitals, number of total patients, occupancy rate but MLR has assumptions that are not easy to meet in real life. The previous studies that utilized machine learning algorithms gave better results than MLR because of managing the non-linearity between input and output variables [8,15]. On the other hand, because of the lack of historical MW database, most of the studies for estimating MW generation based on surveys and questionnaires but this may yield misleading results [3,8,9,15,17,18]. Some researchers used machine learning algorithms such as Support Vector Machine (SVM) and Artificial Neural Networks (ANN) and compared the traditional statistical techniques such as MLR [2,315,16]. Machine

learning algorithms outperformed statistical techniques because of its ability to better model non-linear relationships but they have disadvantages related to poor performance on small data [15]. Since some MW data is time-based, different ARIMA models as time series models have been used to predict the MW amount [1,14,19,20]. Requiring long time data to detect seasonality and not robust for outlier and missing value are the weaknesses of time series analysis methods [20,21].

İstanbul has strong influence on the environment and the health since it has Over the 15 million population, but MW generation database is still lacking the other factors that may affect the MW generation like medical waste type, social economic and health institutions type [1,22]. Penalized regression methods that extensions of linear regression models can be an effective tool for estimating the MW generation to discover the relations between the input and output parameters where there is limited data also not having assumptions like MLR. Besides these methods can deal with the multicollinearity as a general problem in MLR. These methods have been used many areas successfully [23-27]. To the best of our knowledge, penalized regression methods as Ridge Regression, Lasso Regression and Elastic Net Regression have not been employed for estimating the MW generation. The aim of this study is to employ and compare these penalized regression methods estimating the MW generation for Istanbul. First, 18 years for actual data for MW amount with input parameters namely crude birth rate, number of hospitals, number of bed available at the hospital, and Gross Domestic Product (GDP). Next, to predict MW generation these penalized regression methods namely Ridge Regression, Lasso and Elastic Net have been employed and their performances compared with R-squared, RMSE and MSE as a performance measures.

The paper structured as follows: Materials and methods were explained in Section 2. Section 3 discussed the results of the models. Finally, conclusions, limitations of the study and future directions were presented in Section 4.

2. Materials and Methods

2.1 Data collection

The dataset used in this study combined the two tables [28] and [29] the years between 2004-2021 in MS Excel worksheet. The input parameters as number of hospital (NH), number of bed (NB), crude birth rate (CBR), number of health personal (NHP) and gross domestic product (GDP) were selected based on the previous studies and data availability. While MW is a dependent variable, remaining variables are the independent variables. Table 1 provides variables and their types.

Table 1. The variables and their types

Variables	**Type of Variables**
Number of hospitals	Numeric
Number of beds	Numeric
Crude birth rate	Numeric
Number of health personnel	Numeric
GDP	Numeric
Medical Waste	**Numeric**

2.2 Data processing

Data pre-processing is the most key step the modelling process because the real data may contain errors or outliers and missing values [30,31]. The missing values have been filled for NH, NB, GDP, NHP and CBR because they have missing values in this study. The training set is used for training model allocated as 80 % of the samples and the testing set allocated to 20 % of the samples have been used for evaluating samples.

2.3 Penalized Regression Methods

Multiple Linear Regression

Ordinal Least Squares (OLS) aims to estimate $\beta_1,\beta_2,\dots,\beta_p$ by minimizing the residual sum of squares (RSS).

$$\text{RSS}= \sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2 \quad (1)$$

Ridge Regression

Ridge Regression with penalty called L2-norm is a linear regression model was proposed by Hoerl and Kennard proposed in 1970 and designed to handle the multicollinearity. The aim of the ridge regression is to determine the coefficients to minimize the sum of squares by employing the penalty to these coefficients in Eqn. 6.5 where λ ≥0 denotes a tuning parameter and $\lambda\sum_{j=1}^{p}\beta_j^2$ denotes a shrinkage penalty.

$$\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{j=1}^{p}\beta_j^2 = \text{RSS}+\lambda\sum_{j=1}^{p}\beta_j^2 \quad (2)$$

The selecting the ideal value of λ is important since least squares produce only one set of coefficients while ridge regression generate a different set of coefficient estimates for different values of λ [32].

Lasso Regression

Lasso regression use a penalty term called L1-norm that denotes the sum of absolute coefficients lead coefficient estimates of insignificant parameters equal to zero that means more simpler and more accurate models. The aim of the lasso regression is to find the lasso coefficients by the minimizing the quantity

$$\sum_{i=1}^{n}(y_i-\beta_0-\sum_{j=1}^{p}\beta_j x_{ij})^2+\lambda\sum_{j=1}^{p}|\beta_j| = \text{RSS}+\lambda\sum_{j=1}^{p}|\beta_j| \quad (3)$$

The term replaced by the has been changed by in the lasso penalty [32].

Elastic Net Regression

Elastic net combines the ridge regression and lasso regression that to shrink coefficients as ridge regression and to set some coefficients to zero like lasso regression. Also, it has computational advantage over ridge and lasso regression. The aim of the Elastic Net regression is to find the elastic net coefficients by the minimizing the quantity [33].

$$\sum_{i=1}^{n}(y_i - \beta_0 - \sum_{j=1}^{p}\beta_j x_{ij})^2 + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}\beta_j^2 = \text{RSS} + \lambda_1 \sum_{j=1}^{p}|\beta_j| + \lambda_2 \sum_{j=1}^{p}\beta_j^2 \quad (4)$$

2.4 Performance Measures

Three performance metrics have been used such as MAE, RMSE, and R^2 as shown in Equation (5-7) to compare the penalized regression models [8,15]:

$$\boldsymbol{MAE} = \frac{\sum_{i=1}^{n}|y_i - x_i|}{n} \quad (5)$$

$$\boldsymbol{RMSE} = \sqrt{\sum_{i=1}^{n}\frac{(y_i - x_i)^2}{n}} \quad (6)$$

$$\boldsymbol{R^2} = 1 - \frac{\sum_{i=1}^{n}(y_i - x_i)^2}{\sum_{i=1}^{n}(y_i - \bar{x}_i)^2} \quad (7)$$

2.5 Cross-validation

A resampling procedure is used called Cross-validation (CV) is to produce equal random subsets of samples for training data and testing data when a small data [7]. In this study, the data is divided into five equal size subsamples and one part is denoted as the validation set, the resting four subsamples denoted as the training data as called five-fold CV. Till each sub-sample is used as validation set, the procedure is repeated for instance five times for five-fold CV in this study. Finally, to allege the optimal hyperparameter values, the average accuracy of five validation set is used.

3.Results and Discussion

This section provides the experimental results and discusses the performance of three regression methods such as Ridge, Lasso and Elastic Net Regression considering MAE, RMSE, R-squared as an evaluation metrics.

Table 2. Results of penalized regression models

Penalized Regression Models	**MAE**	**RMSE**	**R^2**
Ridge Regression	427.67	650.37	0.95
Lasso Regression	349.56	596.52	0.96
Elastic Net Regression	2109.32	2310.40	0.42

With regards to the MAE, Lasso Regression achieves the best performance with 349.56, Ridge Regression and Elastic Net Regression second and third with the 427.67 and 2109.32 respectively. The same order of performance is achieved with respect to R^2 and RMSE with values 0.96, 596.52 and 0.42, 2310.40 respectively. The results showed that Ridge Regression and Lasso Regression had good performances that both can be used predicting MW amount. Lasso Regression outperformed other penalized regression models with the minimum RMSE, MAE, also the higher R -squared performance measures with this small dataset. Lasso regression is successful by reaching the good ML performances [2] as well successful than some studies [3,34].

Conclusion

Prediction of MW amount is a vital information for medical waste management systems in the future especially megacities as İstanbul that have profound impact on the environment. Ridge Regression, Lasso and Elastic Net have been employed to predict MW generation for Istanbul. The methods have been compared with the performance measures MAE, RMSE and R-squared. Among the penalized regression models, Lasso Regression outperforms the other algorithms while Ridge Regression and Elastic Net Regression are ranked second and third.
The limitation of this study is small data so SMOTE technique will be used to create artificial data in the future study and the results will be compared.

Conflicts of Interest: The author declares no conflict of interest.

Funding: Funding information is not applicable / No funding was received.

Ethics committee approval (if needed)

This study does not require ethical approval.

References

[1] Ceylan, Z.; Bulkan, S.; Elevli, S. Prediction of medical waste generation using SVR, GM (1,1) and ARIMA models: a case study for megacity Istanbul. *J Environ Health Sci Engineer.* **2020**, *18*:687–697. https://doi.org/10.1007/s40201-020-00495-8.

[2] Jahandideh, S.; Jahandideh, S.; Asadabadi, E.B.; Askarian, M.; Movahedi, M.M.; Hosseini, S.; Jahandideh, M. The use of artificial neural networks and multiple linear regression to predict rate of medical waste generation. *Waste Manag.* **2009, 29***(11)*:2874-9. doi: 10.1016/j.wasman.2009.06.027.

[3] Golbaz, S.; Nabizadeh, R.; Sajadi, H.S. Comparative study of predicting hospital solid waste generation using multiple linear regression and artificial intelligence. *J Environ Health Sci Engineer.* **2019,** *17*:41–51. https://doi.org/10.1007/s40201-018-00324-z.

[4] Shinee, E.; Gombojav, E et al. Healthcare Waste Management in the Capital City of Mongolia. *Waste Manag.* **2008,** *28*: 435-444.

[5] Nie, L.; Qiao, Z.; Wu, H. Medical Waste Management in China: A Case Study of Xinxiang. *J Environ Prot Ecol.* **2014**, *5*: 803-810. http://dx.doi.org/10.4236/jep.2014.510082.

[6] Uysal, F.; Tinmaz, E. Medical waste management in Trachea region of Turkey: suggested remedial action. *Waste Manag Res.* **2004**, Oct;*22(5):*403-7. doi: 10.1177/0734242X04045690.

[7] Birpinar, M.E.; Bilgili, M.S.; Erdoğan, T. Medical waste management in Turkey: A case study of Istanbul. *Waste Manag.* **2009**, *29(1)*:445-8. doi:10.1016/j.wasman.2008.03.015.

[8] Nguyen, X.C.; Nguyen, T.T.H.; La, D.D. et al. Development of machine learning-based models to forecast solid waste generation in residential areas: A case study from Vietnam.*Resour Conserv Recycl.* **2021**,167:105381. https://doi.org/10.1016/j.resconrec.2020.105381.

[9] Bdour, A.; Altrabsheh, B.; Hadadin, N.; Al-Shareif, M. Assessment of medical wastes management practice: a case study of the northern part of Jordan. *Waste Manag.* **2007**, *27*:746–59. https://doi.org/10.1016/J.WASMAN.2006.03.004.

[10] Sabour, M.R.; Mohamedifard, A.; Kamalan, H. A mathematical model to predict the composition and generation of hospital wastes in Iran. *Waste Manag.* **2007**,*27*:584–7. https://doi.org/10.1016/J.

[11] Idowu, I.; Alo, B.; Atherton, W.; Al, K.R. Profile of medical waste management in two healthcare facilities in Lagos, Nigeria: a case study. *Waste Manag Res.* **2013**, *31*:494–501. https://doi.org/10.1177/0734242X13479429.
[12] Al-Khatib, I.A.; Abu, Fkhidah. I.; Khatib, J.I.; Kontogianni, S. Implementation of a multi-variable regression analysis in the assessment of the generation rate and composition of hospital solid waste for the design of a sustainable management system in developing countries. Waste Manag Res. **2016**, 34:225–34. https://doi.org/10.1177/0734242X15622813.
[13] Çetinkaya, A.Y.; Kuzu, S.L.; Demir,A. Medical waste management in a mid-populated Turkish city and development of medical waste prediction model. *Environ Dev Sustain.* **2020**, *22*:6233–6244. https://doi.org/10.1007/s10668-019-00474-6.
[14] Chauhan, A.; Singh, A. An ARIMA model for the forecasting of healthcare waste generation in the Garhwal region of Uttarakhand. *India Int J Serv Oper Informatics*. **2017**, *8*:352. https://doi.org/10.1504/ijsoi.2017.086587
[15] Karpušenkaitė, A.; Ruzgas, T.; Denafas, G. Forecasting medical waste generation using short and extra short datasets: Case study of Lithuania. *Waste Manag Res*. **2016,** *34(4)*:378-87. doi: 10.1177/0734242X16628977
[16] Thakur, V.; Ramesh, A. Analyzing composition and generation rates of biomedical waste in selected hospitals of Uttarakhand, India. *J Mater Cycles Waste Manag*. **2018**,*20*:877–90. https://doi.org/10.1007/s10163-017-0648-7.
[17] Dissanayaka, D.M.S.H.; Vasanthapriyan, S. Forecast municipal solid waste generation in Sri Lanka. In 2019 International Conference on Advancements in Computing, Sri Lanka ,5-7 December 2019 ;210-215. doi: 10.1109/ICAC49085.2019.9103421.
[18] Meleko, A.; Adane, A. Assessment of Health Care Waste Generation Rate and Evaluation of its Management System in Mizan Tepi University Teaching Hospital (MTUTH), Bench Maji Zone, South West Ethiopia. *Ann Rev Resear*.**2018**, *1*: 555566. http://dx.doi.org/10.19080/ARR.2018.01.555566.
[19] Karpušenkaitė, A.; Ruzgas, T.; Denafas, G. Time-series-based hybrid mathematical modelling method adapted to forecast automotive and medical waste generation: Case study of Lithuania. *Waste Manag Res*. **2018**, *36*: 454 - 462.
[20] Papacharalampous, G.; Tyralis, H.; Koutsoyiannis, D. Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: A multiple-case study from Greece. *Water Resour Manag*. **2018**, *32*:5207–5239. http://dx.doi.org/10.1007/s11269-018-2155-6.
[21] Pavlyshenko, B. M.Machine-Learning Models for Sales Time Series Forecasting. *Data*. **2019**,*4(1)*:1–11. https://doi.org/10.3390/data4010015.
[22] Birpinar, M.E.; Bilgili, M.S.; Erdoğan, T. Medical waste management in Turkey: A case study of Istanbul. *Waste Manag.* **2009**, *29(1)*:445-8. doi:10.1016/j.wasman.2008.03.015.
[23] Musarrat Ijaz, Zahid Asghar & Asma Gul (2021) Ensemble of penalized logistic models for classification of high-dimensional data, Communications in Statistics- Simulation and Computation, 50:7, 2072-2088, DOI: 10.1080/03610918.2019.1595647
[24] Buyrukoğlu, S. & Yılmaz, Y. (2021). An Approach for Airfare Prices Analysis with Penalized Regression Methods . Veri Bilimi , 4 (2) , 57-61 .
[25] Greenwood CJ, Youssef GJ, Letcher P, Macdonald JA, Hagg LJ, Sanson A, et al. (2020) A comparison of penalised regression methods for informing the selection of predictive markers. PLoS ONE 15(11): e0242730. https://doi.org/10.1371/journal.pone.024273
[26] Rendall R, Pereira AC, Reis MS. Advanced predictive methods for wine age prediction: Part I - A comparison study of single-block regression approaches based on variable selection, penalized regression, latent variables and tree-based ensemble methods. Talanta. 2017 Aug 15;171:341-350. doi: 10.1016/j.talanta.2016.10.062. Epub 2016 Nov 9. PMID: 28551149.
[27] Tütmez, B. (2020). Air Quality Assessment by Statistical Learning-Based Regularization . Çukurova Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi , 35 (2) , 271-278 . DOI: 10.21605/cukurovaummfd.792412
[28] İstanbul Metropolitan Municipality Department Open Data Portal. Available online:https://data.ibb.gov.tr/en/dataset/31d85b21-32a9-4270-95d9-1712a6567ea/resource/50036dfd-aea5-4f06-832f-f7020fdaaa5a/download/ilce-yl-ve-atk-turu-baznda-atk-miktar-2021.xlsx (Accessed 10 July 2022)
[29] The official website of Turkish Statistics Institute. Available online: https://biruni.tuik.gov.tr/medas/ (accessed 10 July 2022)
[30] Ramírez-Gallego, S.; Krawczyk, B.; García, S.; Woźniak, M.; Herrera, F. A survey on data preprocessing for data stream mining: Current status and future directions. *Neurocomputing.* **2017**,*239*:39-57. https://doi.org/10.1016/j.neucom.2017.01.078.
[31] Kwak, S.K.; Kim, J.H. Statistical data preparation: management of missing values and outliers. *Korean J of anesthesiology.* **2017,** *70(4)*:407. doi: 10.4097/kjae.2017.70.4.407.
[32] James, G., Witten, D., Hastie, T., & Tibshirani, R. "An introduction to statistical learning", Vol. 112, p. 18,. New York: springer, 2013.
[33] Zhang, Z., Lai, Z., Xu, Y., Shao, L., Wu, J., & Xie, G. S. "Discriminative elastic-net regularized linear regression". IEEE Transactions on Image Processing, 26(3), 1466-1481, 2017
[34] Tesfahun, E.; Kumie, A.; Beyene, A. Developing models for the prediction of hospital healthcare waste generation rate. *Waste Manag Res.* **2016**, Jan;*34(1)*:75-80. doi: 10.1177/0734242X15607422.