Journal of Computer Science Advancements, 2(5) - October 2024 297-306

E - ISSN 3024-899X
P - ISSN 3026-3379
Journal of Computer Science Advancements

Forecasting Waste Generation with Increment Linear Regression Technique: A Case Study of SIMASKOT Application

Puguh Jayadi [1], Nasrul Rofiah Hidayati [2], Saifulloh [3], Suhardi Hamid [4], Salehuddin Shuib [5], Siti Nurbaya Ismail [6]

[1,2,3] *Universitas PGRI Madiun, Indonesia*
[4,5,6] *Universiti Teknologi MARA, Malaysia*

Corresponding Author: Puguh Jayadi, Email: puguh.jayadi@unipma.ac.id

Article Information:
Received September 22, 2024
Revised September 27, 2024
Accepted October 17, 2024

ABSTRACT

This research aims to develop a prediction system for urban waste generation using the Incremental Linear Regression method on SIMASKOT. This method is applied to deal with the limitations of historical data, where the prediction results from the previous year are used as training data to predict the next year. The problem faced is the lack of sufficient data to create accurate and reliable prediction models in the long term. The purpose of this study is to improve the accuracy of waste generation prediction using an incremental regression approach. The experimental methodology involves the use of waste generation data from several waste categories during the period 2019 to 2022, which is then used to predict data until 2026. Model evaluation was carried out using the metrics Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2. The results show that this incremental prediction model is able to provide more accurate predictions than conventional models, especially for more volatile waste categories such as wood twigs and metals. The conclusion of this study shows that the Incremental Linear Regression technique is effective to be used in waste generation prediction, and can be integrated in long-term prediction-based monitoring applications.

Keywords: *Incremental Linear Regression, Incremental Models, Limited Data, Long-Term Prediction, Waste Generation Prediction, Waste Monitoring*

Journal Homepage https://journal.ypidathu.or.id/index.php/jcsa

How to cite: Jayadi, P., Hidayati, R, N., Saifulloh, Saifulloh., Hamid, S., Shuib, S., & Ismail, N, S. (2024). Forecasting Waste Generation with Increment Linear Regression Technique: A Case Study of SIMASKOT Application. *Journal of Computer Science Advancements*, 2(5). 297-306 https://doi.org/10.70177/jsca.v2i5.1369

Published by: Yayasan Pendidikan Islam Daarut Thufulah

INTRODUCTION

Predicting science or data forecasting has become an important research focus in many sectors, including economics, industry, and the environment. One of the main challenges in prediction is data limitations, which often occur with data related to waste or waste (El Bilali dkk., 2022). Many waste management systems rely on limited historical data to predict future waste generation, which necessitates the use of predictive techniques that can work effectively with small and incomplete data (Anesti dkk., 2024; Meroni dkk., 2021; Yousuf dkk., 2022). In this context, research focusing on developing limited dataset-based prediction models is crucial to achieve accurate results.

Many prediction methods have been developed to handle large data sets, such as Support Vector Regression (SVR), Artificial Neural Networks (ANNs), and Random Forest Regression (Choi & Lee, 2023; Rekha Sankar & Panchapakesan, 2024; H. Yang & Schell, 2021). This method is well-known for its ability to handle high data complexity, especially when data is abundant. However, most studies using the method rely on large and high-quality datasets for the training of prediction models (Dong dkk., 2021; Zhang dkk., 2024). When faced with a limited dataset, the accuracy of the model often decreases significantly, and the method becomes less efficient.

Research on forecasting for small datasets has found that linear-based regression methods and simple prediction models are more efficient compared to neural network or ensemble-based approaches (Zainuddin dkk., 2019). Linear regression has the advantage of being easy to implement and can handle small datasets well, especially if implemented with specific optimization strategies, such as incremental models. However, some studies mention that there are still limitations in terms of flexibility and ability to capture non-linear patterns from data.

Although some prediction models have been successfully applied to small datasets, such as in k-nearest neighbors (k-NN) or ridge regression techniques, the results are still limited by the assumption of linearity and sensitivity to noise data (Ide dkk., 2024; Y. Yang dkk., 2019). Additionally, conventional approaches often fail to capture long-term data trends if the model is not updated with the latest data (Liu, 2021). This is very important in the prediction of waste generation, where the trend of waste generation can fluctuate based on external factors that change from year to year.

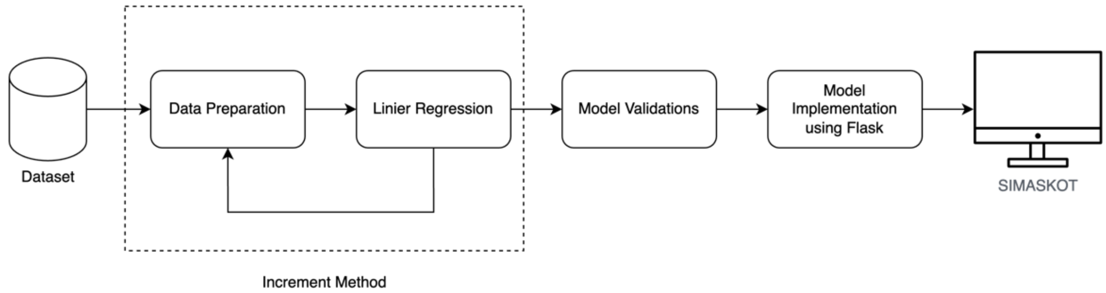
Therefore, this study proposes the use of the Incremental Linear Regression method, which allows the prediction model to be gradually updated with the latest data. Each year, the predictions generated by the model will be used as the basis for training to predict the following year. This approach has the advantage of handling limited data and allows the model to continue to be refined as new prediction data becomes available. The use of incremental linear regression is expected to overcome the limitations of conventional linear regression models that are static and not adaptive to changes in data trends.

This study hopes to improve the accuracy of waste generation prediction by utilizing predictions as training data so that the model can capture long-term changes and be adaptive to limited data dynamics. In addition, this research is expected to contribute to the development of a better waste monitoring system, especially in the context of urban waste management. Thus, the results of this study are expected to be used as a basis for sustainable waste monitoring applications, such as the SIMASKOT application, which can utilize prediction models for more efficient and responsive waste management planning.

RESEARCH METHODOLOGY

The method used in this study consists of several interrelated stages, as described in Figure 1, starting from data preparation to the implementation of a prediction model in a Fask-based municipal waste monitoring application (SIMASKOT). The first stage is Data Preparation, where the available datasets must be carefully processed and prepared (Hou dkk., 2022). At this stage, the data cleaning process is carried out to ensure that the dataset is free of *missing values* or inconsistent data. After the data is ready, normalization is carried out so that the linear regression model can work more effectively in identifying patterns.

Figure 1. Research Methods



After the preparation process is completed, it is continued with Linear Regression, where a prediction model is built based on a linear relationship between independent variables (years) and dependent variables (amount of waste generated). The model is trained using historical data to predict the amount of waste generated in the future. After the model is trained, the Model Validations stage is performed to evaluate the model's performance using various metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 (determination coefficient) (Choi & Lee, 2023; Kaczmarek-Majer & Hryniewicz, 2019; Rekha Sankar & Panchapakesan, 2024; Yagli dkk., 2020). This validation aims to ensure that the model can predict with an adequate level of accuracy, as well as measure the resulting prediction errors. In the final stage, the Model Implementation using Flask is carried out to integrate the prediction model into the waste generation monitoring application. The application allows users to monitor and predict the amount of waste generation based on historical data and future predictions generated by linear regression models.

The dataset used in this study comes from annual waste generation data that covers various categories of waste: food waste, plastic, metal, fabric, glass, and other categories.

The dataset used in this study came from https://sipsn.menlhk.go.id/. This data covers several years, with each year providing complete information on the composition and amount of waste generated from these categories. However, this dataset is relatively limited, both in terms of the amount of data and the duration of time covered, so an effective and efficient prediction approach is needed. Therefore, prediction models must be designed in such a way that they can capture patterns from existing data and use them to predict future waste generation, even with the limited historical data available. In this context, Incremental Linear Regression is the approach of choice because of its ability to utilize prediction data from the previous year as a training basis to predict the next year, thus extending the limited use of the dataset.

Linear regression, as the core of the prediction technique used, is a method that looks for the best line representing the linear relationship between independent variables and dependent variables. In simple linear regression, the equation used is as follows:

$$y = \beta_0 + \beta_1 x + \ldots + \beta_n x + \epsilon \tag{1}$$

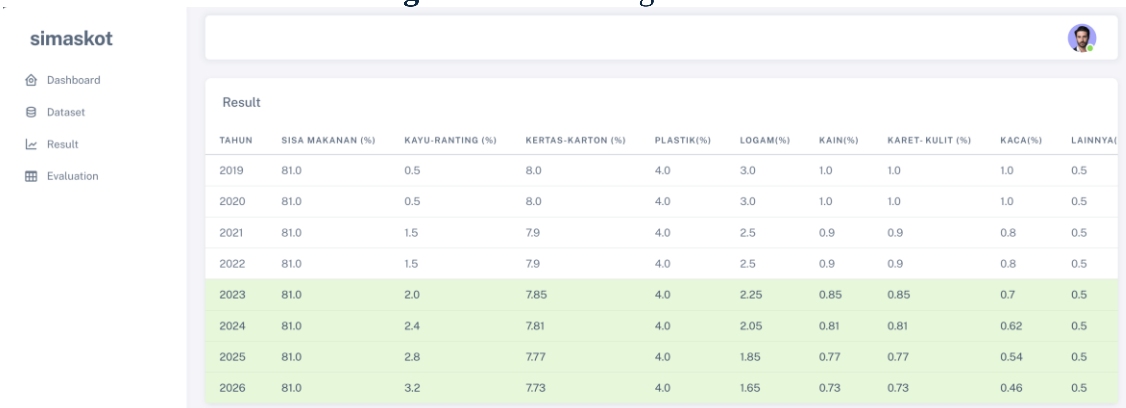
In equation 1, y is the dependent variable (waste generation), x is the independent variable (year), β_0 is the intercept, β_1 is the slope, and ϵ is an error or residual. The model works by looking for optimal values of $\beta 1$ and $\beta 1$, which minimizes prediction errors. The Incremental Linear Regression technique applied in this study expands the basic concept of linear regression with an incremental approach. Each year, the model is trained using data from previous years, and the prediction results for that year are then added to the dataset to be used as training data in the following year. This approach allows the model to be continuously updated and refined as new predictions are made so that the model can adapt to the latest data trends without having to retrain from scratch. This is an important contribution to this research, where Incremental Linear Regression can provide an effective solution for limited dataset-based waste generation prediction cases.

This methodology not only allows for more accurate predictions but also ensures that the resulting model can be implemented directly into Flask-based monitoring applications. This application not only predicts waste generation but also provides an easy-to-use interface to visualize prediction results over a longer period. Along with the collection of new data, the model is continuously updated automatically through an incremental process, which makes this approach highly efficient for real-world applications in urban waste management.

RESULT AND DISCUSSION

In this section, the application of Incremental Linear Regression, which is used to predict waste generation based on historical data from 2019 to 2022, will be discussed. This method leverages predictions from the previous year as training data for the following year, allowing the model to be continuously updated with each iteration. With this technique, the prediction model is expected to be able to overcome the limitations of available data while maintaining the accuracy of the predictions for years to come. In this experiment, waste generation prediction was carried out for 2023 to 2026 by utilizing historical data from waste categories, such as food waste, twigs, cardboard, metal, and plastic.

Figure 2. Forecasting Results


simaskot

Dashboard

Dataset

Result

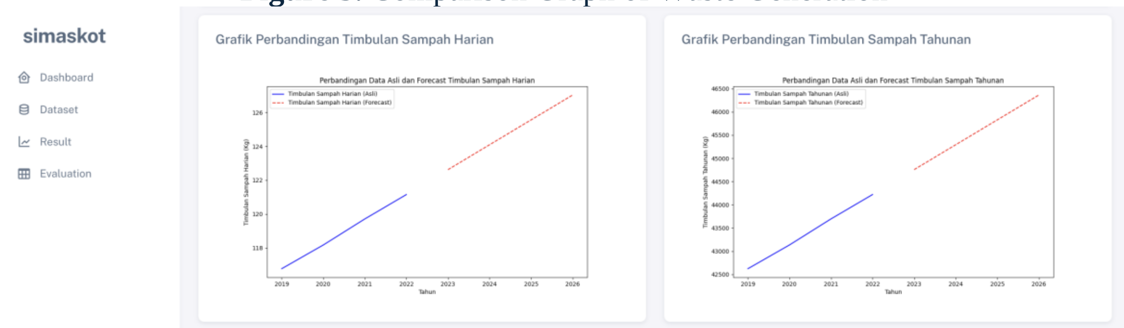
Evaluation

Result

TAHUN	SISA MAKANAN (%)	KAYU-RANTING (%)	KERTAS-KARTON (%)	PLASTIK(%)	LOGAM(%)	KAIN(%)	KARET- KULIT (%)	KACA(%)	LAINNYA(
2019	81.0	0.5	8.0	4.0	3.0	1.0	1.0	1.0	0.5
2020	81.0	0.5	8.0	4.0	3.0	1.0	1.0	1.0	0.5
2021	81.0	1.5	7.9	4.0	2.5	0.9	0.9	0.8	0.5
2022	81.0	1.5	7.9	4.0	2.5	0.9	0.9	0.8	0.5
2023	81.0	2.0	7.85	4.0	2.25	0.85	0.85	0.7	0.5
2024	81.0	2.4	7.81	4.0	2.05	0.81	0.81	0.62	0.5
2025	81.0	2.8	7.77	4.0	1.85	0.77	0.77	0.54	0.5
2026	81.0	3.2	7.73	4.0	1.65	0.73	0.73	0.46	0.5

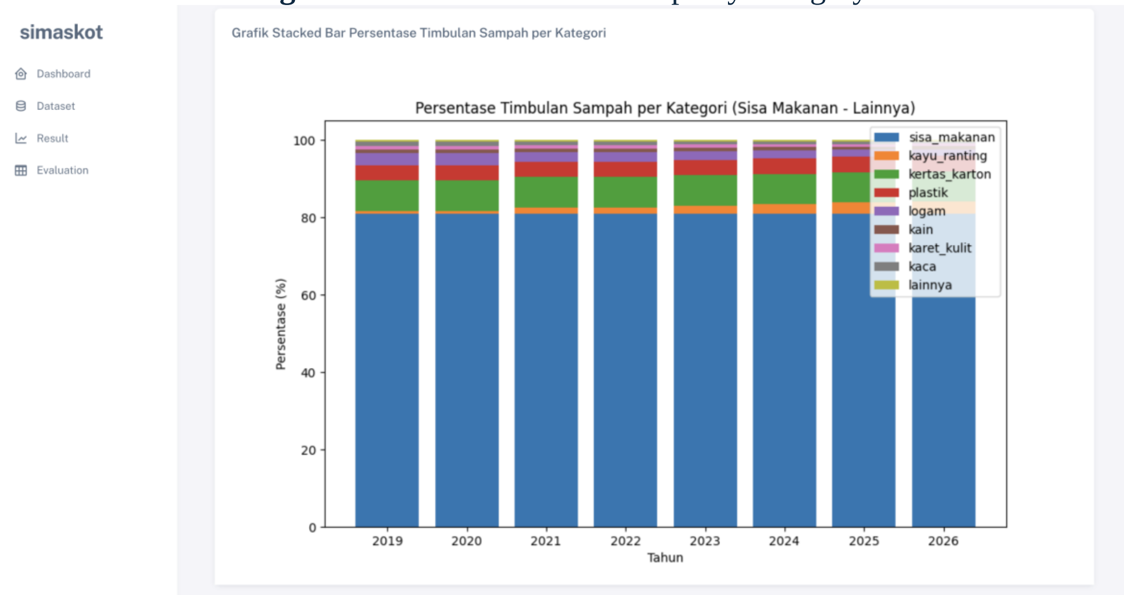
The prediction results for 2023 to 2026 in Figure 2 show significant changes in the composition of waste. In the twig category, the percentage increased from 0.5% in 2019 to 2.0% in 2023 and continues to rise to reach 3.2% in 2026. The cardboard category showed a slight decline, from 8.0% in 2020 to 7.85% in 2023, and continued to decline to 7.73% in 2026. Another significant change was seen in the metals category, where the percentage fell from 3.0% in 2019 to 2.25% in 2023 and continued to decline to 1.65% in 2026. In contrast, the food waste category remained constant at 81% during the period. This prediction provides an overview of the dynamics of future waste distribution, which can help with more efficient waste management planning.

Figure 3. Comparison Graph of Waste Generation



The comparison graph in Figure 3 between the original data and the predicted data for daily and annual waste generation shows a consistent upward trend from 2019 to 2026. In the daily waste generation graph, the original value in 2022 of 121.15 kg is predicted to increase to more than 126 kg in 2026. Similarly, the annual waste generation, which reached 44,219.8 kg in 2022, is projected to continue to increase to close to 46,000 kg in 2026. This graph shows that the prediction model can capture the trend of increasing waste volumes, although the historical data used is very limited. This increase in waste generation can be attributed to increased consumption patterns, as well as changes in the proportion of waste categories such as wood, twigs, and metals.

Figure 4. Prediction Results Graph by Category



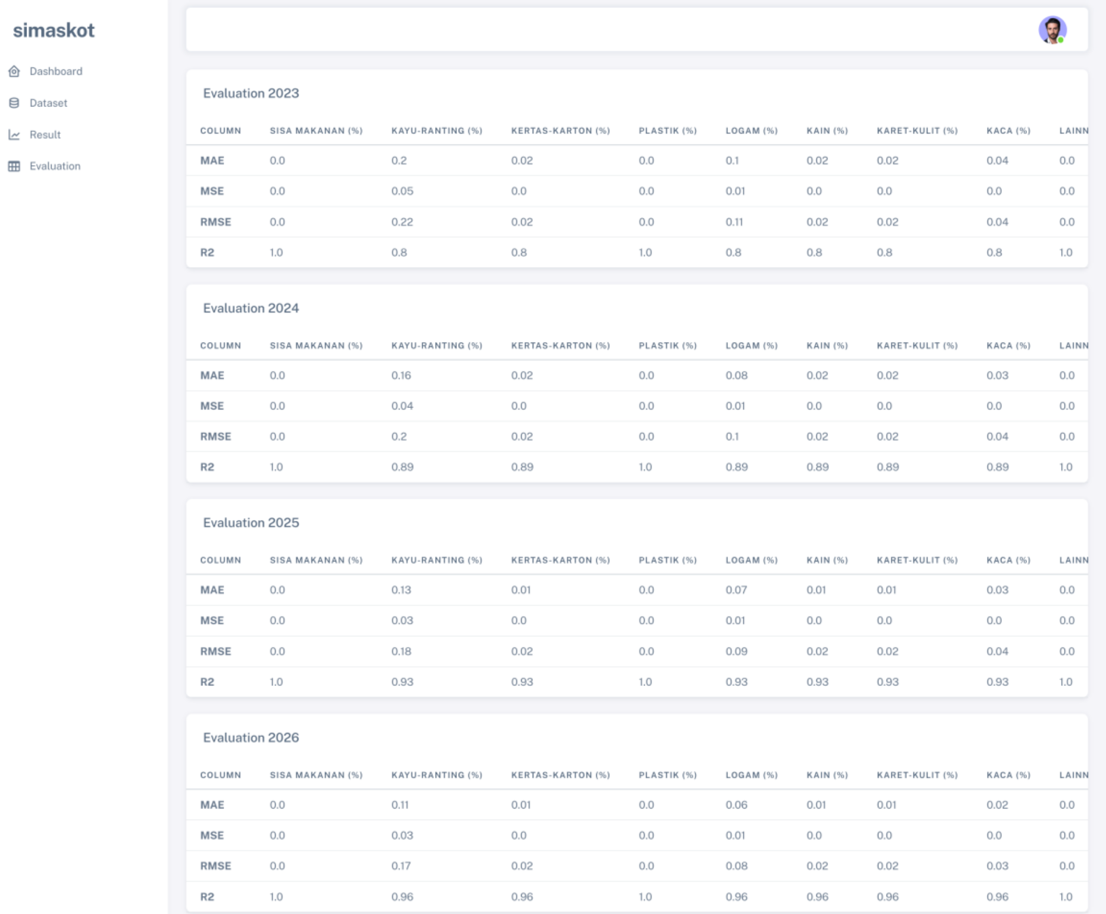
In addition, the bar graph in Figure 4 showing the percentage of waste generation per category confirms that food waste remains the dominant category throughout the observed period, with a proportion of more than 80% annually. However, other categories, such as timber twigs and metals, show quite dynamic proportions change from year to year. This reflects that the pattern of waste composition can be influenced by various external factors, including changes in waste management policies or people's consumption habits. Thus, predicting the composition of waste per category can help policymakers in planning more effective waste management strategies, especially in dealing with waste categories that are predicted to increase.

Evaluation of the prediction model is carried out using several evaluation metrics, such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 (Coefficient of Determination) (Anesti et al., 2024; Tolbin, 2023). For 2023, the evaluation results show that the prediction of the twig category has an MAE of 0.2, an RMSE of 0.22, and an R^2 value of 0.8. Meanwhile, for the food waste category, the model showed excellent evaluation results with MAE and RMSE of 0 and an R^2 value of 1, which showed a perfect prediction. A similar trend is also seen in 2024, where the MAE for twig wood decreases to 0.16 with an increased R^2 value to 0.89. The results of this evaluation show that the model is increasingly able to improve its accuracy over time, especially in predicting more variable waste categories such as wood twigs and metals.

In 2025 and 2026, as shown in Figure 5, the evaluation results show a consistent improvement in the accuracy of the prediction model. In 2025, the MAE for twig wood drops to 0.13, while the R^2 value increases to 0.93. RMSE scores for most categories, including metals and paperboard, also showed significant improvement. In 2026, the

MAE for twig wood drops further to 0.11 with an R^2 value of 0.96, suggesting that the model is increasingly able to capture long-term trends in the more volatile waste category. Meanwhile, the food waste category still showed a very accurate prediction with an R^2 value of 1 throughout the evaluation period.

Figure 5. Forecasting Evaluation Results



Evaluation 2023

COLUMN	SISA MAKANAN (%)	KAYU-RANTING (%)	KERTAS-KARTON (%)	PLASTIK (%)	LOGAM (%)	KAIN (%)	KARET-KULIT (%)	KACA (%)	LAINN
MAE	0.0	0.2	0.02	0.0	0.1	0.02	0.02	0.04	0.0
MSE	0.0	0.05	0.0	0.0	0.01	0.0	0.0	0.0	0.0
RMSE	0.0	0.22	0.02	0.0	0.11	0.02	0.02	0.04	0.0
R2	1.0	0.8	0.8	1.0	0.8	0.8	0.8	0.8	1.0

Evaluation 2024

COLUMN	SISA MAKANAN (%)	KAYU-RANTING (%)	KERTAS-KARTON (%)	PLASTIK (%)	LOGAM (%)	KAIN (%)	KARET-KULIT (%)	KACA (%)	LAINN
MAE	0.0	0.16	0.02	0.0	0.08	0.02	0.02	0.03	0.0
MSE	0.0	0.04	0.0	0.0	0.01	0.0	0.0	0.0	0.0
RMSE	0.0	0.2	0.02	0.0	0.1	0.02	0.02	0.04	0.0
R2	1.0	0.89	0.89	1.0	0.89	0.89	0.89	0.89	1.0

Evaluation 2025

COLUMN	SISA MAKANAN (%)	KAYU-RANTING (%)	KERTAS-KARTON (%)	PLASTIK (%)	LOGAM (%)	KAIN (%)	KARET-KULIT (%)	KACA (%)	LAINN
MAE	0.0	0.13	0.01	0.0	0.07	0.01	0.01	0.03	0.0
MSE	0.0	0.03	0.0	0.0	0.01	0.0	0.0	0.0	0.0
RMSE	0.0	0.18	0.02	0.0	0.09	0.02	0.02	0.04	0.0
R2	1.0	0.93	0.93	1.0	0.93	0.93	0.93	0.93	1.0

Evaluation 2026

COLUMN	SISA MAKANAN (%)	KAYU-RANTING (%)	KERTAS-KARTON (%)	PLASTIK (%)	LOGAM (%)	KAIN (%)	KARET-KULIT (%)	KACA (%)	LAINN
MAE	0.0	0.11	0.01	0.0	0.06	0.01	0.01	0.02	0.0
MSE	0.0	0.03	0.0	0.0	0.01	0.0	0.0	0.0	0.0
RMSE	0.0	0.17	0.02	0.0	0.08	0.02	0.02	0.03	0.0
R2	1.0	0.96	0.96	1.0	0.96	0.96	0.96	0.96	1.0

These results show that the use of Incremental Linear Regression in waste generation prediction provides adequate results for a limited dataset. The resulting predictions are not only able to follow historical patterns but also provide a realistic picture of possible trends in the future. With evaluations that show improved accuracy from year to year, this technique can be relied upon in waste generation prediction monitoring systems, such as SIMASKOT, which require continuous monitoring and continuous model updates.

CONCLUSION

This study shows that the Incremental Linear Regression method is effective in predicting waste generation using limited historical data. By utilizing predictions from the previous year as training data for the next year, the model can improve the accuracy of predictions from year to year. The evaluation results showed that the R^2 value increased and the MAE and RMSE values decreased, especially in more varied waste categories such as wood twigs and metals. This approach proved to be particularly suitable for long-term prediction scenarios, especially in the context of municipal waste management that requires periodic model updates without retraining from scratch. With high accuracy, especially in stable categories such as food waste, this model provides an efficient solution for planning waste management capacity.

The research also opens up opportunities for further development, such as testing with larger datasets or the integration of other machine learning models. In addition, this method can be integrated into the SIMASKOT monitoring application to provide more accurate long-term predictions, allowing for more efficient waste management planning. Overall, Incremental Linear Regression makes a significant contribution to waste management, offering a sustainable and easy-to-implement predictive solution to deal with future waste generation dynamics.

REFERENCES

Anesti, N., Kalamara, E., & Kapetanios, G. (2024). Forecasting with Machine Learning methods and multiple large datasets. *Econometrics and Statistics*, S2452306224000510. https://doi.org/10.1016/j.ecosta.2024.08.003

Choi, W., & Lee, S. (2023). Performance evaluation of deep learning architectures for load and temperature forecasting under dataset size constraints and seasonality. *Energy and Buildings*, *288*, 113027. https://doi.org/10.1016/j.enbuild.2023.113027

Dong, W., Sun, H., Tan, J., Li, Z., Zhang, J., & Zhao, Y. Y. (2021). Short-term regional wind power forecasting for small datasets with input data correction, hybrid neural network, and error analysis. *Energy Reports*, *7*, 7675–7692. https://doi.org/10.1016/j.egyr.2021.11.021

El Bilali, A., Taghi, Y., Briouel, O., Taleb, A., & Brouziyne, Y. (2022). A framework based on high-resolution imagery datasets and MCS for forecasting evaporation loss from small reservoirs in groundwater-based agriculture. *Agricultural Water Management*, *262*, 107434. https://doi.org/10.1016/j.agwat.2021.107434

Hou, Z., Li, J., Wang, L., Zhang, Y., & Liu, T. (2022). Improving the forecast accuracy of ECMWF 2-m air temperature using a historical dataset. *Atmospheric Research*, *273*, 106177. https://doi.org/10.1016/j.atmosres.2022.106177

Ide, Y., Ozaki, S., Izutsu, S., Kotoura, T., Yamashiro, M., & Kodama, M. (2024). Hybrid real-time wave forecasting model combining Gaussian process regression and neural networks. *Ocean Engineering*, *312*, 119028. https://doi.org/10.1016/j.oceaneng.2024.119028

Kaczmarek-Majer, K., & Hryniewicz, O. (2019). Application of linguistic summarization methods in time series forecasting. *Information Sciences*, *478*, 580–594. https://doi.org/10.1016/j.ins.2018.11.036

Liu, O. Y. (2021). Assessing the contribution of urban green spaces in green infrastructure strategy planning for urban ecosystem conditions and services. *Sustainable Cities and Society*, *68*(Query date: 2024-06-05 09:16:36). https://doi.org/10.1016/j.scs.2021.102772

Meroni, M., Waldner, F., Seguini, L., Kerdiles, H., & Rembold, F. (2021). Yield forecasting with machine learning and small data: What gains for grains?

Agricultural and Forest Meteorology, *308–309*, 108555. https://doi.org/10.1016/j.agrformet.2021.108555

Rekha Sankar, S., & Panchapakesan, M. (2024). Hybrid feature selection model for accurate wind speed forecasting from numerical weather prediction dataset. *Expert Systems with Applications*, *248*, 123054. https://doi.org/10.1016/j.eswa.2023.123054

Yagli, G. M., Yang, D., & Srinivasan, D. (2020). Reconciling solar forecasts: Probabilistic forecasting with homoscedastic Gaussian errors on a geographical hierarchy. *Solar Energy*, *210*, 59–67. https://doi.org/10.1016/j.solener.2020.06.005

Yang, H., & Schell, K. R. (2021). Real-time electricity price forecasting of wind farms with deep neural network transfer learning and hybrid datasets. *Applied Energy*, *299*, 117242. https://doi.org/10.1016/j.apenergy.2021.117242

Yang, Y., Che, J., Deng, C., & Li, L. (2019). Sequential grid approach based support vector regression for short-term electric load forecasting. *Applied Energy*, *238*, 1010–1021. https://doi.org/10.1016/j.apenergy.2019.01.127

Yousuf, M. U., Al-Bahadly, I., & Avci, E. (2022). Statistical wind speed forecasting models for small sample datasets: Problems, Improvements, and prospects. *Energy Conversion and Management*, *261*, 115658. https://doi.org/10.1016/j.enconman.2022.115658

Zainuddin, N. H., Lola, M. S., Djauhari, M. A., Yusof, F., Ramlee, M. N. A., Deraman, A., Ibrahim, Y., & Abdullah, M. T. (2019). Improvement of time forecasting models using a novel hybridization of bootstrap and double bootstrap artificial neural networks. *Applied Soft Computing*, *84*, 105676. https://doi.org/10.1016/j.asoc.2019.105676

Zhang, W., Cai, Y., Zhan, H., Yang, M., & Zhang, W. (2024). Multi-energy load forecasting for small-sample integrated energy systems based on neural network Gaussian process and multi-task learning. *Energy Conversion and Management*, *321*, 119027. https://doi.org/10.1016/j.enconman.2024.119027