Technovasia IAM International Journal © 2025
Volume **1**, Number **2**, May **2025** pp. **49-55**

PREDICTION OF WASTE VOLUME BASED ON ANNUAL TRENDS USING DATA MINING

Made Aka Suardana[1,*] and Vitri Tundjungsari[2]

[1,2]Faculty of Computer Science
Esa Unggul University
Jl Arjuna Utara, Jakarta Barat, Indonesia
*Corresponding author: madeakas99@student.esaunggul.ac.id

Received February 2024; accepted April 2025

ABSTRACT
Waste management has become a critical issue in many large cities due to the increasing volume of waste generated. This study aims to predict waste volume based on annual trends using data mining techniques with the XGBoost model. The data analyzed includes total waste from 2003 to 2017, focusing on plastic, construction, and food waste types. The prediction results show that the XGBoost model can follow seasonal fluctuations in the data, although there are discrepancies in certain periods, such as approaching 2040. Overall, the waste volume trend shows a significant increase, especially in plastic and construction waste, which aligns with the growth in plastic consumption and infrastructure development. Although there has been an increase in the amount of waste recycled, its proportion to the total waste remains low, indicating that recycling programs are not yet optimal. Based on this study, more effective policies and strengthened recycling infrastructure are needed to improve the proportion of recycled waste, as well as data-driven waste management planning to support waste reduction and more sustainable management in the future.
Keywords: Data Mining, Waste Management, XGBoost, Recycling, Waste Volume Prediction

1. Introduction

The issue of waste management has become a serious concern in various major cities around the world. With the increasing population and urbanization, the volume of waste generated each year continues to grow significantly. The inability to accurately predict waste volume often leads to inefficient waste management, such as waste accumulation at disposal sites, transportation overload, and suboptimal use of waste processing facilities [1-4]. As a result, environmental impacts like water, soil, and air pollution are becoming more severe.

Effective waste management planning requires reliable predictions of the waste volume that will be generated in the future [5-7]. These predictions are crucial to ensure proper resource allocation, such as the number of collection vehicles, adequate disposal site locations, and efficient waste processing capacity. However, waste volume patterns are often influenced by various factors such as seasons, lifestyle changes, and economic activities, making manual predictions or conventional methods less accurate [8-10].

One approach that can be used is data mining techniques [11-15]. By understanding the patterns and trends of waste production, both the government and the public can manage waste more efficiently and systematically, thus reducing the potential adverse environmental impacts and preventing waste accumulation in the future.

2. Methodology

The algorithm used in this study is XGBoost (Extreme Gradient Boosting). XGBoost was chosen for its ability to handle large and complex datasets and its computational efficiency [10-14]. This algorithm is known for its boosting mechanism, which improves prediction accuracy by minimizing errors gradually. Additionally, XGBoost features regularization that helps prevent overfitting, which is crucial for ensuring the model generalizes well to new data. The algorithm also supports parallel computing, making the training process faster compared to other methods.

In this study, the researcher utilizes Google Colab as the platform for the data mining process. Google Colab was selected because it provides a cloud-based environment with access to GPUs, accelerating the execution of models like XGBoost. Moreover, Google Colab supports Python programming and comes with a wide range of pre-installed data science libraries, such as Pandas, NumPy, and Scikit-Learn, making data analysis very convenient. The platform's ease of collaboration and real-time file sharing is also another significant advantage.

Figure 1 Steps in Data Mining with XGBoost

2.1. Data Collection

In the first stage, the data used in this study is obtained through data collection. The required data is raw data that includes waste management. Data collection is the process of gathering unprocessed data for further analysis. The data used in this study was taken from the Kaggle site, with data related to waste management. The collected data is then stored in a shared drive that can be accessed through the following link: (source: Link).

2.2. Data Preprocessing

After that, the next step is data preprocessing. In this study, data preprocessing includes several important stages to prepare the data for analysis. First, data cleaning is performed to remove outliers and irrelevant entries that could skew the results or lead to misleading insights, ensuring the dataset remains accurate and relevant. Next, handling missing data is addressed by filling in missing values using appropriate methods, such as imputation techniques, to maintain the integrity of the dataset and prevent information loss. Finally, normalization is applied to scale the data uniformly, ensuring that all features contribute equally to the analysis and making it easier to process and interpret. These steps are crucial for preparing the data to be clean, complete, and ready for further analysis.

2.3. Applying XGBoost Algorithm

The XGBoost model is trained using the processed data to predict waste volume by iteratively minimizing errors.

2.4. Model Evaluation

The model's performance is measured using accuracy and error rate metrics. Model parameters are tuned if necessary to achieve optimal results.

2.5. Result Visualization

The prediction results are visualized in Google Colab to provide a clear overview of waste volume patterns and trends.

3. Results and Discussion

From the data mining results using XGBoost, prediction graphs of total waste, waste trends based on waste type, and the comparison of recycled vs discarded waste are obtained. Below is the total waste prediction graph.

3.1. Prediction of total waste generated

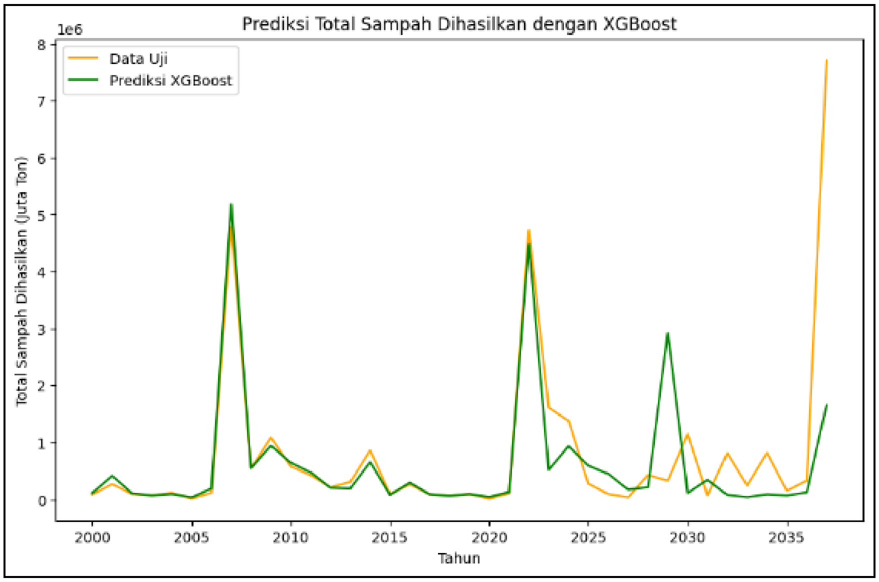

Figure 2 Prediction of Total Waste Generated

Based on the prediction graph using the XGBoost model, it is observed that the model's prediction (marked with the green line) shows a fluctuation pattern similar to the actual data (orange line) in most periods. This indicates that the XGBoost model with applied tuning is able to capture seasonal trends and certain peaks in the historical data. The model successfully follows most of the fluctuations in the test data, especially at certain peaks such as around 2005, 2015, and other years. This suggests that the model is fairly good at predicting the up-and-down patterns over a specific time period. However, despite the model's ability to follow the general trend, there are some areas where the prediction does not reach the actual peak, particularly approaching the 2040 period. Additionally, the predicted results generated by the model appear more stable for most of the time, although there are some sharp fluctuations in the actual data that the model does not fully capture.

3.2. Trend of total waste based on waste type

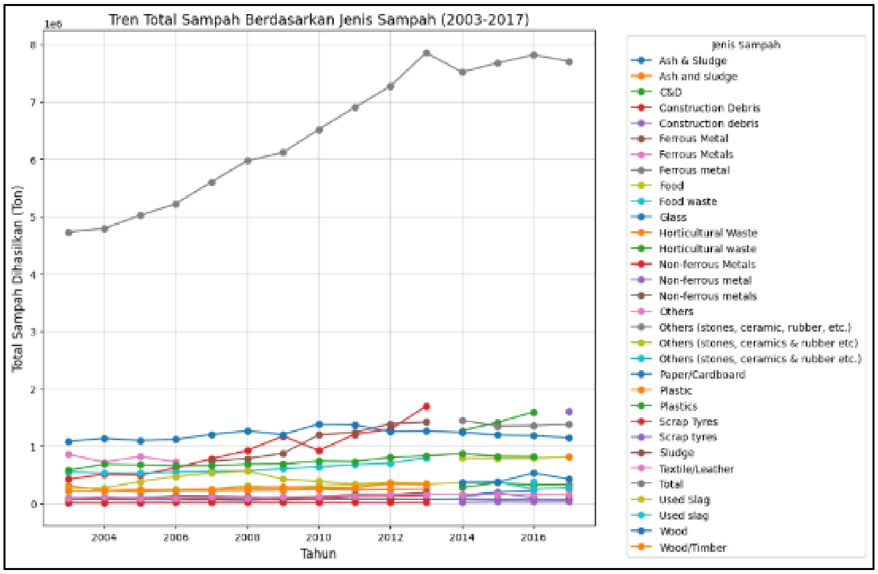

Figure 3 Trend of Total Waste Based on Waste Type

The graph above shows the total waste trends by type from 2003 to 2017. In general, the total waste generated increased significantly until 2013, before stabilizing or slightly decreasing until 2017. This trend can be linked to population growth, urbanization, and high economic activity during this period. The type of waste that showed a significant increase is plastic, which aligns with the increased consumption of plastic-based products. Construction waste also increased sharply, likely due to massive infrastructure development.

Additionally, food waste continued to grow consistently, reflecting changes in consumption patterns. This is aligned with research investigated by [15]. Meanwhile, other categories such as ferrous metals, non-ferrous metals, and used slag showed stable trends with no significant increase. This indicates that waste from metal materials tends to be more controlled, possibly due to recycling programs or the economic value of these materials. This phenomenon suggests that plastic and construction waste management require more attention. An in-depth study of the factors contributing to the increase in specific types of waste is also important to support more effective waste management policies. This trend reflects the need for sustainable strategies in waste management, focusing on reduction, recycling, and efficient management of organic waste.

3.3. Comparision of Recycled vs. Disposed Waste

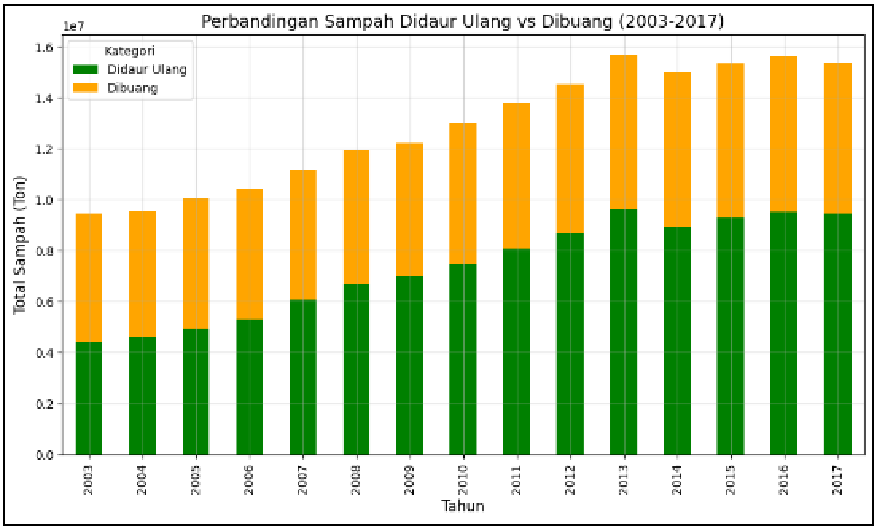

Figure 4 Comparison of Recycled vs. Disposed Waste

This graph provides an overview of the comparison between recycled and disposed waste during the period from 2003 to 2017. The total volume of waste continued to increase throughout the period, reflecting population growth, urbanization, and higher consumption. Although recycled waste shows a gradual but stable increasing trend, disposed waste still dominates the total waste generated. A significant increase in the amount of disposed waste is observed between 2007 and 2013, which could serve as an evaluation point for waste management policies during that period. Despite efforts to improve recycling-based management, the proportion of waste recycled compared to the total waste remains relatively small. This indicates that recycling programs need to be strengthened through infrastructure development, public education, and providing economic incentives to encourage broader participation. Progress in recycling is a positive indicator, but major challenges still remain in reducing waste production, particularly for waste types that are difficult to recycle. Both the government and society need to collaborate in creating more effective policies that focus on sustainable waste management strategies, such as waste reduction, enhanced recycling, and managing waste based on the circular economy.

4. Conclusions

The results of this study indicate that the total waste trend has significantly increased over time, particularly driven by plastic and construction waste. Although efforts have been made to increase the amount of waste recycled, its proportion to total waste remains low, highlighting the need for strengthening recycling policies, infrastructure, and public education. Predictions using the XGBoost model provide an insight into the potential future increase in waste, but the significant spike in 2035 needs to be further validated to ensure the model's accuracy. Therefore, strategic and data-driven measures are needed to address the challenges of waste management, including optimizing recycling programs, reducing waste production, and planning for long-term sustainability

REFERENCES

[1] R. Alsabt, W. Alkhaldi, Y. A. Adenle, and H. M. Alshuwaikhat, “Optimizing waste management strategies through artificial intelligence and machine learning - An economic and environmental impact study,” *Clean. Waste Syst.*, vol. 8, no. April, p. 100158, 2024, doi: 10.1016/j.clwas.2024.100158.

[2] G. Rashmi and S. S. Kumar, “Prediction of Solid Garbage Waste Generation in Smart Cities using Naive Bayes Algorithm,” *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 2S, pp. 53–56, 2019, doi: 10.35940/ijitee.b1031.1292s19.

[3] B. Fang *et al.*, *Artificial intelligence for waste management in smart cities: a review*, vol. 21, no. 4. Springer International Publishing, 2023. doi: 10.1007/s10311-023-01604-3.

[4] K. N. Sami, Z. M. A. Amin, and R. Hassan, “Waste Management Using Machine Learning and Deep Learning Algorithms,” *Int. J. Perceptive Cogn. Comput.*, vol. 6, no. 2, pp. 97–106, 2020, doi: 10.31436/ijpcc.v6i2.165.

[5] D. Palanikkumar, R. R. Kumar, M. Masud, M. M. Alnfiai, and M. Abouhawwash, “Bayes-Q-Learning Algorithm in Edge Computing for Waste Tracking,” *Intell. Autom. Soft Comput.*, vol. 36, no. 2, pp. 2425–2440, 2023, doi: 10.32604/iasc.2023.033879.

[6] G. W. Cha *et al.*, “Development of a prediction model for demolition waste generation using a random forest algorithm based on small datasets,” *Int. J. Environ. Res. Public Health*, vol. 17, no. 19, pp. 1–15, 2020, doi: 10.3390/ijerph17196997.

[7] D. B. Olawade *et al.*, “Smart waste management: A paradigm shift enabled by artificial intelligence,” *Waste Manag. Bull.*, vol. 2, no. 2, pp. 244–263, 2024, doi: 10.1016/j.wmb.2024.05.001.

[8] M. Chaerul, M. Puturuhu, and I. Artika, “Optimasi Rute Pengangkutan Sampah dengan Menggunakan Metode Nearest Neighbour (Studi Kasus: Kabupaten Manokwari, Papua Barat),” *J. Wil. dan Lingkung.*, vol. 10, no. 1, pp. 55–68, 2022, doi: 10.14710/jwl.10.1.55-68.

[9] X. C. Nguyen *et al.*, “Development of machine learning - based models to forecast solid waste generation in residential areas: A case study from Vietnam,” *Resour. Conserv. Recycl.*, vol. 167, no. November 2020, p. 105381, 2021, doi: 10.1016/j.resconrec.2020.105381.

[10] V. Tundjungsari, B.P.D. Putranto, M.B. Ulum, N.Anwar, “CIRCONOMY: Integrating IoT, Semantic Web, and Gamification for Circular Waste Management - Insight from an Indonesia Case Study”, JMIR Serious Games 2025, DOI: 10.2196/66781.

[11] I. F. Hanbal, J. S. Ingosan, N. A. A. Oyam, and Y. Hu, “Classifying Wastes Using Random Forests, Gaussian Naïve Bayes, Support Vector Machine and Multilayer Perceptron,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 803, no. 1, 2020, doi: 10.1088/1757-899X/803/1/012017.

[12] T. Nadu, C. Author, T. Sciences, and T. Nadu, “Artificial neural networks are compared to K- nearest Neighbors in a novel swarm optimization for smart waste management systems employing automated machine,” vol. 10, pp. 1998–2007, 2023.

[13] A. Namoun, A. Tufail, M. Y. Khan, A. Alrehaili, T. A. Syed, and O. BenRhouma, “Solid Waste Generation and Disposal Using Machine Learning Approaches: A Survey of Solutions and Challenges,” *Sustain.*, vol. 14, no. 20, 2022, doi:

10.3390/su142013578.
[14] V. Tundjungsari, "Dasar Machine Learning", Deepublish: Desember 2024, ISBN: 978-623-02-9824-0.
[15] V. Tundjungsari, “Food Waste Management Strategy with Green and Digital Technology", Technovasia, vol. 1 (2), May 2025.