

# Analysis of Contributing Factors and Prediction of Urban Waste Generation Using PSO-ANN

Taufiqotul Bariyah<sup>[1]\*</sup>, Brina Miftahurrohmah<sup>[2]</sup>, Niswaton Faria<sup>[3]</sup>

Department of Informatics<sup>[1]</sup>, Department of Information System<sup>[2]</sup>, Department of Engineering Management<sup>[3]</sup>  
Universitas Internasional Semen Indonesia  
Gresik, Indonesia

taufiqotul.bariyah@uisi.ac.id<sup>[1]</sup>, brina.miftahurrohmah@uisi.ac.id<sup>[2]</sup>, niswaton.faria@uisi.ac.id<sup>[3]</sup>

**Abstract**— This research examines the factors influencing waste generation in urban areas, with a focus on East Java, which has experienced increased waste due to population growth and urbanization. Using the Spearman correlation method, it was found that unemployment ( $\rho = 0.87$ ) and population ( $\rho = 0.865$ ) are significantly related to waste generation. However, HDI ( $\rho = -0.152$ ) and population density ( $\rho = -0.169$ ) are uncorrelated with waste generation. Furthermore, waste generation predictions will be built using the Particle Swarm Optimization-Artificial Neural Network (PSO-ANN) model. The modeling results showed that the PSO-ANN architecture with one hidden layer achieved RMSE of 0.125 and MAE of 0.109, while the model with two hidden layers achieved RMSE of 0.123 and MAE of 0.105. These findings indicate that the two-hidden-layer PSO-ANN model is more effective in predicting waste generation than the single-layer model. This study recommends exploring alternative methods and additional variables to provide a more comprehensive examination and analysis of waste disposal management in the future.

**Keywords**— Waste generation, Spearman correlation, PSO-ANN, prediction model

## I. INTRODUCTION

Waste, often defined as the material remaining from human activities without further use [1-3], poses significant environmental and management challenges, particularly in urban areas. The quantity of waste generated varies significantly across regions [4] due to differences in consumption patterns influenced by demographic, socioeconomic, and environmental factors [5, 6]. The main determinants of waste generation include population dynamics (size, growth, density, and mobility), socioeconomic factors (standard of living, economic development, and technological progress), and environmental conditions (geography, climate, and seasonal variations) [7-9]. These interrelated factors collectively shape the composition and approach to waste management globally.

As one of the provinces with a large population and dense urbanization, East Java is among the provinces that generate the most waste in Indonesia. During 2020, approximately 3.8 million tons of waste were produced in East Java. As much as 44.11% is household waste [10]. The growing volume of waste certainly makes waste management a crucial issue that must be taken seriously. This high volume of waste generated poses a

severe threat to waste management, especially in large cities such as Surabaya, Sidoarjo, and Malang. This problem is increasingly complex because several key factors influence it, first, demographic factors such as population growth and high levels of urbanization. Second, the composition of urban waste is more diverse and complex than that of rural waste [11]. Third, changes in consumption patterns in urban areas that differ from those in rural areas [12]. Practical, focused waste management policies need to be crafted to meet the community's needs [13].

Various studies have shown varying results regarding the factors influencing waste generation. In Sumatra Island, population density and waste generation show a positive correlation [14], but this is not found in several developing countries [15]. Education, employment, and knowledge factors also have a significant influence, as seen in Gorontalo Regency [16]. Meanwhile, research in Vietnam shows no correlation between income and waste generation [17], although other studies have shown the opposite [18,19]. These differences in research results indicate that the factors influencing waste generation vary widely across regions and communities.

Although previous studies have identified key drivers of waste generation, findings remain inconsistent across contexts. This disparity highlights the gap in universal predictive models and the need for localized, adaptive approaches. This study addresses this gap by systematically analyzing the socioeconomic and demographic factors affecting waste generation in East Java and developing an optimized predictive model using Particle Swarm Optimization-Artificial Neural Network (PSO-ANN) to improve the accuracy of urban waste management. In this context, the PSO-ANN approach offers a promising solution for more accurately modeling and predicting waste generation. This method combines the advantages of particle optimization in finding the best solution with the ability of artificial neural networks to handle complex data patterns.

## II. LITERATURE REVIEW

### A. Particle Swarm Optimization

Particle Swarm Optimization (PSO) is an optimization technique that mimics social behavior, allowing the model to efficiently explore the solution space by adapting and sharing information between particles [20, 21]. On the other hand, Artificial Neural Networks can learn from data and identify

hidden patterns, making them highly effective in processing non-linear data, which is often found in waste generation analysis [22]. By combining PSO and ANN, this model will not only increase the precision of waste generation prediction but also provide a better understanding of the interactions among factors that influence waste generation, such as demographics, economics, and community behavior [23].

**B. Related Research**

Previous studies have identified demographic, socioeconomic, environmental, and behavioral factors as determinants of waste generation. However, most of these are descriptive and contextual, yielding inconsistent results across regions, making it difficult to produce reliable, adaptive predictive models. Commonly used analytical methods, such as linear regression or simple correlation, are unable to capture the non-linear relationships and complex interactions between standard variables related to waste generation. The use of Artificial Intelligence (AI), such as ANN (Analytical Network Analysis), yields insignificant results compared to other optimization techniques, such as PSO (Social Network Analysis), for improving predictive performance, particularly in Indonesia.

The application of PSO-ANN has been proven effective in the context of waste management, as shown by [24], which indicates that PSO-ANN can help formulate targeted management policies by considering the complex factors that influence waste generation. Further research suggests that PSO-ANN can be utilized to predict waste generation with high accuracy, which is crucial in waste management planning [25]. Case studies in various countries [26] show that this model can improve the efficiency of waste management in urban areas. PSO produces better predictions compared to conventional methods [27], while [28] and [29] emphasized the importance of multidimensional data integration to improve prediction accuracy. Therefore, PSO-ANN is expected to bring positive changes to waste management by providing a more powerful tool for predicting and responding to challenges in this field [30].

The combination of these approaches strengthens the argument that PSO-ANN is a valuable tool for addressing the increasingly complex challenges of waste management. The first step in determining policies for waste management and the environment in the targeted area is analyzing the dominant factors driving waste generation [31]. These dominant factors serve as a basis for forecasting future waste generation, financial planning, and the design of waste management facilities. Effective waste management enhances the community's standard of living and contributes to sustainable development.

This study addresses this gap by systematically analyzing the locally relevant dominant factors in East Java and developing a PSO-ANN-based predictive model. This model is better at adapting to data complexity and spatial variation. ANN's ability to recognize non-linear patterns, combined with PSO's effectiveness in exploring optimal solutions, yields more accurate outputs than conventional approaches and enables practical implementation in data-driven waste management policies.

**III. METHODOLOGY**

This study aims to develop a waste generation prediction model using an ANN optimized with PSO as illustrated in Fig. 1. The main objective is to identify the determinants of waste generation through statistical analysis and build a predictive model with ANN-PSO optimization to improve accuracy. The results are expected to form the basis for recommendations for data-based waste management policies.

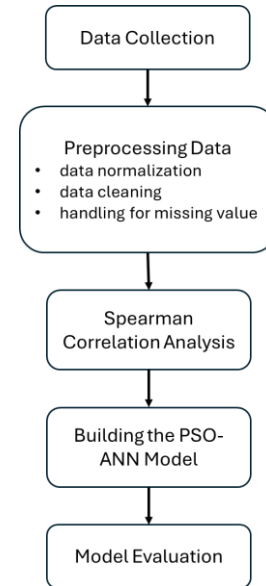


Fig. 1. Research Methodology.

The research begins with the identification of the dominant variables that influence waste generation, based on literature reviews and preliminary analysis. Furthermore, data on these variables are collected from relevant sources. The acquired data then undergoes preprocessing, including outlier handling, normalization, and feature selection, to ensure data quality. Next, the data is split into training and test sets at a specific ratio for model training and validation. The next stage involves determining the ANN architecture, including selecting the number of layers and neurons, and then optimizing using the PSO algorithm to obtain optimal parameters, such as the learning rate, batch size, and the optimal number of neurons. The next step is to evaluate the built model's performance using test data using RMSE, MAPE, and MAE. The evaluation results are further analyzed to assess the model's reliability in predicting waste generation. This research is anticipated to generate a precise and robust prediction model and to provide policy recommendations based on analysis findings.

**A. Data Collection and Preprocessing**

This research used secondary data from Badan Pusat Statistik (BPS) and the Sistem Informasi Pengelolaan Sampah Nasional (SIPSN). This study uses data gathered from 2018 to 2023. The following data were collected in this study.

- 1) *Amount of waste generation*: average waste generation daily (tons per year).
- 2) *Unemployment*: number unemployed of population over 15 years old.
- 3) *Population* (thousand people).

- 4) *Income*: average monthly net income of informal workers based on the highest education level (Rp).
- 5) *Low-income households*: number of low-income families by regency/city in East Java (Thousand People).
- 6) *Population density* (per km<sup>2</sup>).
- 7) *Human Development Index* (HDI).

One of the significant challenges in using real-world datasets is handling outliers and missing values. Outliers are data points that are distant from the others, and missing values indicate missing data for an unknown reason. Both outliers and missing values can significantly influence analysis and modeling of the data if not appropriately handled. The data collected then proceeds to the preprocessing stage. At this stage, the data is examined for outliers and noise, then normalized, and checked for incomplete data (missing values). Data cleaning and handling of incomplete data is carried out to increase the veracity of the prediction results in the next stage. Incomplete data can reduce the accuracy of the prediction model [32].

Several methods can be used to identify outliers in the data. In this study, boxplots will be used to identify observations outside the interquartile range (1.5 times the interquartile range or 3 times the interquartile range if extreme outliers are found)—and to predict missing values using a regression or correlation model based on other variables. However, the risk of overfitting may arise if the model is too complex or there are many missing values. Another approach is to create multiple "plausible" complete data sets by inputting missing values numerous times based on the estimated distribution. This method is considered more robust to missing-value mechanisms.

#### B. Spearman Correlation Analysis

The preprocessed data were then analyzed using the Spearman correlation technique. Spearman's Correlation is used to determine the interrelation between two variables when the data scales of both variables are at least ordinal [33]. This correlation analysis is included in the nonparametric analysis because the data are not normally distributed. Like the Pearson Correlation, the relationship between two variables must go through a series of testing stages. The testing stages carried out are the same as those in Pearson's Correlation: determining the hypothesis, determining the alpha level, calculating the test statistic, determining the critical region, making decisions, and drawing conclusions. If Pearson Correlation must meet the requirements of data distribution, namely normal distribution, Spearman Correlation does not require it [34]. In this Spearman correlation, the critical area used is to compare the test statistics with the  $\rho$  table (for Spearman correlation).

Spearman correlation coefficients are calculated to determine the correlation between each variable and waste generation. The correlation coefficient  $\rho$  provides an overview of the strength and direction of the relationship among these variables. Furthermore, the results will be interpreted to identify which factors are significantly associated with waste generation.

#### C. Building the PSO-ANN Model

After identifying the dominant variables that significantly influence waste generation, the next step is to predict waste generation using the PSO-ANN model. The process begins by configuring a neural network, which includes defining the architecture, such as the quantity of hidden layers, the number of neurons, and the activation functions to be implemented. Before the model can be trained, the available data must be split into training and test sets. Using a ratio of 80% for data training and 20% for data testing. Training data will be used to "teach" the model. In contrast, testing data will assess the model's performance on unfamiliar data, ensuring it generalizes effectively. In addition, the PSO algorithm is used to optimize the ANN's parameters and determine the most effective ones. Following this, the model is trained on the prepared training data and then validated on test data to evaluate its performance.

#### D. Model Evaluation

This section discusses the predictive model's performance assessment. This assessment is crucial to understanding the model's estimated waste output. Testing is performed using validation data (20%) that was not used during model learning. This method guarantees an unbiased evaluation of the model's ability to extrapolate its understanding to novel, unfamiliar data.

To quantitatively measure the accuracy and dependability of the model, three crucial performance indicators are utilized:

- 1) *Root Mean Square Error* (RMSE): This metric is used to evaluate the accuracy of predictive models. It measures the difference between the model's predicted values and the actual values. RMSE gives greater weight to large deviations, so a lower RMSE value indicates better model estimation accuracy.
- 2) *Mean Absolute Error* (MAE): MAE measures the median of the absolute values of the deviations between the model's forecast results and the actual observations. This metric is less susceptible to outliers than RMSE. A smaller MAE value indicates greater prediction precision.
- 3) *Mean Absolute Percentage Error* (MAPE): MAPE presents the median error in a percentage format. This indicator is handy for assessing the model's relative accuracy compared to the actual value. A lower MAPE indicates better model performance in estimating percentages.

A careful analysis of these three indicators enables a comprehensive evaluation of the PSO-ANN model's performance. If an error occurs, the study will assess the extent to which it affects the model's reliability and predictive accuracy of waste generation.

## IV. RESULTS AND DISCUSSION

This section will cover the findings of the analysis conducted to identify the relationships among variables that affect waste generation using Spearman's correlation, as well as the implementation of the PSO-ANN algorithm to improve prediction accuracy. This analysis was carried out in two main stages: a statistical examination of data distribution and

correlation patterns, followed by a machine learning-based modeling stage using PSO-ANN. The goal is not only to understand the underlying data relationships but also to test the effectiveness of predictive algorithms in real-world socio-environmental contexts. Before conducting a correlation analysis, a crucial initial step is to understand the distributional characteristics of each research variable. Every variable has been standardized using Min-Max scaling to equalize the data to a scale of 0 to 1. Distribution visualization is done using a boxplot (Fig. 2).

From the boxplot analysis in Fig. 2, several important observations can be drawn:

- 1) *Distribution Asymmetry (Skewness)*: Most variables show asymmetric distributions. Variables such as 'Waste Generation,' 'Population Density,' and 'Unemployment' tend to have distributions skewed to the right (positive skew), indicating a concentration of data at lower values and a 'tail' towards higher values. In contrast, 'HDI' and 'Income' distributions tend to be more symmetric with some deviations. This condition indicates that the assumption of univariate normality is not fully met for at least one variable.
- 2) *Presence of Outliers*: The boxplot identifies the presence of outliers in several variables. Specifically, 'Waste Generation,' 'HDI,' 'Population,' 'Income,' and 'Unemployment' each show several data points that fall far outside the interquartile limits (IQR), indicating extreme values. These outliers can significantly affect the estimation of statistical parameters, including correlation coefficients. These outliers may also influence the accuracy of machine learning models used in the next stage, as extreme values can distort the learning process and increase prediction errors, particularly in RMSE and MAE metrics. Therefore, their presence should be taken into account when interpreting model performance results.

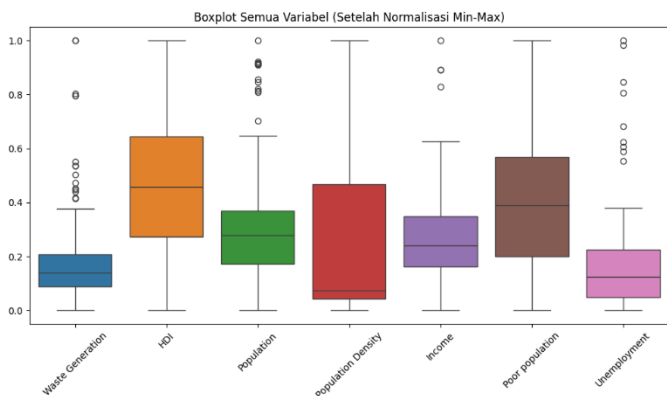


Fig. 2. Boxplot of normalized variable distribution.

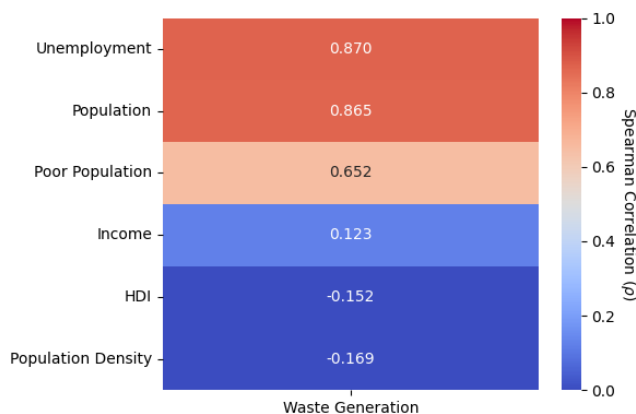


Fig. 3. Correlation value of each independent variable to the dependent variable.

The non-normal distribution of the data and the presence of outliers, as identified in the boxplot, have significant implications for selecting correlation analysis methods. Given these conditions, where many variables exhibit skewness and the presence of outliers, this analysis will use Spearman's Correlation ( $\rho$ ) as the primary method to measure the relationship between variables. The selection of Spearman's Correlation is based on its nonparametric nature, which does not require normality assumptions and is less sensitive to outliers because it uses ranks rather than raw values. This approach is thought to offer a more robust and accurate representation of the magnitude and orientation of the monotonic relationship among variables in this dataset.

The results of the Spearman correlation analysis of numerous socioeconomic variables and waste production in Fig. 3 show that unemployment and population variables exhibit strong positive linear relationships with waste generation, with correlation coefficients of 0.870 and 0.865, respectively. That result suggests that increases in both variables significantly contribute to increased waste volume. In addition, the low-income household variable shows a moderate positive correlation ( $\rho = 0.652$ ), indicating that an increase in the number of poor households is associated with higher waste volumes. However, the impact is smaller than that of unemployment and population. The lowest correlation is observed with the income variable, which has a value of 0.123. HDI and population density have a higher correlation than income, with respective values of 0.152 and 0.169. However, the difference between HDI and population density, in terms of the income variable and the other three variables with high correlations, lies in their negative correlation with waste generation. Specifically, the higher the HDI and population density in an area, the lower the waste production in that area. However, both have similarities, but they do not correlate with waste generation. These results highlight the importance of focusing on unemployment and population as the primary socioeconomic factors contributing to waste generation. Policymakers may benefit from targeting interventions or programs that address these variables in urban planning and sustainability strategies. Thus, the variables selected for further analysis are unemployment, population, and low-income households.

TABLE I. VARIATIONS IN MODEL PERFORMANCE BASED ON DIFFERENT HYPERPARAMETERS FOR ONE HIDDEN LAYER

n <sup>th</sup> Particle	Hyperparameter	RMSE	MAPE	MAE
1	Number of Neuron: 78 Learning rate: 0.041 Batch size: 32	0.125	34.5%	0.109
2	Number of Neuron: 38 Learning rate: 0.093 Batch size: 32	0.135	37.6%	0.115
3	Number of Neuron: 64 Learning rate: 0.611 Batch size: 32	0.127	32.5%	0.107
4	Number of Neuron: 58 Learning rate: 0.038 Batch size: 64	0.128	31.8%	0.110
5	Number of Neuron: 85 Learning rate: 0.052 Batch size: 64	0.134	37.7%	0.112
6	Number of Neuron: 44 Learning rate: 0.046 Batch size: 64	0.130	33.5%	0.113

Next, an analysis was performed to assess the impact of independent variables on waste generation using the PSO-ANN method. Previously, correlation analysis showed that only three variables—unemployment, population, and low-income households—were highly or moderately correlated with waste generation. Therefore, these variables were selected as inputs for the PSO-ANN model to improve prediction accuracy and gain a deeper understanding of the relationships among these variables and waste generation. The outcomes of the PSO-ANN modeling with one hidden layer, as presented in TABLE I, demonstrate differences in model performance across various hyperparameters, specifically the number of neurons, learning rate, and batch size. Six particles will be used in this analysis.

Based on the results of the experiments, each particle with a different hyperparameter combination produces a different level of accuracy. The 1st particle with 78 neurons, a learning rate of 0.041, and a batch size of 32 shows quite good performance with an RMSE of 0.125, a MAPE of 34.5%, and an MAE of 0.109. The 3rd particle with 64 neurons and a higher learning rate (0.611) achieves an RMSE of 0.127 and the lowest MAPE of 32.5%, indicating that a larger learning rate can help improve model accuracy. Meanwhile, the 2nd particle with 38 neurons and a learning rate of 0.093 produces a slightly higher error (RMSE 0.135, MAPE 37.6%). On the other hand, the 4th and 6th particles, with a batch size of 64, show relatively stable results, with MAPEs of 31.8% and 33.5%, respectively. The 5th particle with 85 neurons has the highest MAPE of 37.7%, indicating that a larger number of neurons does not always guarantee better performance. The combination of the 3rd and 4th particles yields optimal results. At the same time, the effect of batch size on accuracy is not very significant compared to variations in learning rate and number of neurons. The 3rd particle is the most optimal overall because it has the lowest MAE, the second-best MAPE, and a low RMSE. This balanced combination of values shows that the third particle performs best at predicting data, even though its learning rate (0.611) is very high, which risks causing training instability (overshooting).

TABLE II. VARIATIONS IN MODEL PERFORMANCE BASED ON DIFFERENT HYPERPARAMETERS FOR TWO HIDDEN LAYERS

n <sup>th</sup> Particle	Hyperparameter	RMSE	MAPE	MAE
1	Number of Neuron: 97 Learning rate: 0.029 Batch size: 32	0.123	35.2%	0.105
2	Number of Neuron: 47 Learning rate: 0.005 Batch size: 32	0.124	34.2%	0.108
3	Number of Neuron: 38 Learning rate: 0.033 Batch size: 32	0.132	31.4%	0.109
4	Number of Neuron: 44 Learning rate: 0.046 Batch size: 64	0.130	33.5%	0.113
5	Number of Neuron: 54 Learning rate: 0.020 Batch size: 64	0.128	34.2%	0.109
6	Number of Neuron: 87 Learning rate: 0.073 Batch size: 64	0.129	32.4%	0.108

The results of the PSO-ANN model with two hidden layers, presented in TABLE II, show variations in model performance across different hyperparameter settings. Based on the experiments, the six particles showed quite diverse performance in terms of prediction accuracy. The 1st particle with 97 neurons, a learning rate of 0.029, and a batch size of 32 delivered the best results, with the lowest RMSE of 0.123 and the lowest MAE of 0.105, although its MAPE was relatively high at 35.2%. Meanwhile, the 3rd particle had the lowest MAPE value of 31.4% but was accompanied by a higher RMSE of 0.132. The 4th and 5th particles, with a batch size of 64, demonstrated stable performance, with an RMSE of approximately 0.130 and a MAPE of 33-34%.

Meanwhile, the 6th particle, with the second-largest combination of neurons (87) and the highest learning rate (0.073), produced an RMSE of 0.129 and an MAPE of 32.4%, indicating that increasing the learning rate and the number of neurons do not constantly significantly improve accuracy. That statement suggests that the optimal parameter combination does not necessarily arise from a single fixed configuration. Instead, it requires a balance between absolute accuracy (RMSE and MAE) and relative percentage error (MAPE). Overall, these results indicate that the number of neurons and learning rate affect model performance, but not necessarily in a linear manner, with the best MAPE achieved by particles with the fewest neurons. Although error metrics are within the acceptable range, further evaluation, such as residual analysis, is recommended to examine potential systematic bias in the model. Uneven distribution of residuals might indicate underfitting, overfitting, or missing explanatory variables.

Based on the findings above, the most suitable particle is the 1st particle, as it has the lowest RMSE (0.123), indicating the least overall model prediction error. In addition, the 1st particle also has the lowest MAE (0.105), which means the smallest absolute average mistake among all experiments. Although the MAPE of the 1st particle is 35.2%, which is not the lowest (the 3rd particle has a MAPE of 31.4%), this value is still within a reasonable range and not too far from the best MAPE. For the

3rd particle, which has the smallest MAPE, the RMSE value is higher (0.132), and the MAE is also not as good as the 1st particle. In the context of predictive model optimization, the first particle provides the best balance between absolute and relative errors, making it the optimal configuration among the six experiments.

When comparing the two architectures, the two-hidden-layer configuration (1st particle) slightly outperforms the best one-hidden-layer model (3rd particle) in terms of RMSE and MAE. However, the one-hidden-layer model yields a lower MAPE, which may indicate improved relative predictive capability. Those results suggest that while deeper networks can enhance prediction performance, the benefit might be marginal, depending on the metric used.

To further assess the robustness of the models, residual analysis was performed. Residuals, defined as the difference between actual and forecast values, help identify trends in prediction errors. An ideal residual distribution should be random, with no observable patterns or trends. In this study, the residuals of the best-performing two-hidden-layer model (1st particle) were plotted and examined. The residuals appeared evenly distributed around zero, with no indication of heteroscedasticity or systematic bias. This finding supports the model's generalization capability and suggests it did not suffer from significant underfitting or overfitting.

From a practical perspective, these findings have important implications for urban environmental policy. Forecasting waste production using socioeconomic factors, such as unemployment and population, provides policymakers with a data-driven foundation for designing targeted interventions. For instance, regions with high unemployment rates may need additional waste management infrastructure. Despite the promising results, the current study has limitations. The dataset is relatively small and does not capture spatial-temporal dynamics, which could be crucial for enhancing prediction accuracy. Future work should incorporate additional contextual data, such as seasonal patterns and policy variables, and experiment with hybrid modeling approaches to further improve generalization.

Although the PSO-ANN framework demonstrates competitive performance, the error levels indicate room for improvement through more complex model architecture, regularization, or feature engineering. Future studies may also explore ensemble models or hybrid approaches to further reduce prediction error.

## V. CONCLUSION

This research emphasizes the impact of unemployment, population, and poverty on waste generation, with strong positive correlations observed through Spearman's analysis. Using these variables as inputs, the PSO-ANN model was applied and evaluated with different configurations. The model with two hidden layers yielded the best performance (RMSE 0.123, MAE 0.105). Those results suggest that while deeper networks can enhance prediction performance, the benefit might be marginal, depending on the metric used. However, overall error rates suggest limited accuracy, particularly in unemployment and population data, due to the presence of

outliers. These outliers likely disrupted the learning process, underscoring the need for more effective data preprocessing techniques.

The findings of this study contribute significantly to the formulation of more effective urban environmental policies. The model's ability to predict waste generation based on key socioeconomic variables, such as the unemployment rate, population density, and income, enables local governments to allocate resources more efficiently. For example, areas with high unemployment rates tend to generate more household waste due to limited access to paid waste management services or low environmental awareness. Therefore, the government can design policies that prioritize the construction of new waste management facilities in areas with high unemployment rates, which are often characterized by high population density and low compliance with waste reduction practices. This can include increasing the number of temporary landfills (TPS) or launching recycling training programs that involve local communities to join waste bank while reducing waste generation. Furthermore, educational strategies and economic incentives for densely populated households to compost independently, or plastic reduction campaigns in areas with predicted high inorganic waste generation, thus supporting more targeted and sustainable facility planning, budget allocation, and education programs.

While the prediction results demonstrate promising performance, this study does have notable limitations. The dataset is relatively small and inadequately captures the spatial-temporal dynamics that often drive fluctuations in waste generation, such as seasonal differences and the impact of specific local government policies. This underscores the importance of further research that integrates additional contextual data, including seasonal patterns, regional waste management policy data, and environmental variables such as rainfall and temperature. Furthermore, experimenting with hybrid modeling approaches, such as combining machine learning with spatial econometric methods or time-series analysis, could further enhance the generalizability of predictive models. These efforts are expected not only to improve the technical accuracy of the models but also to strengthen the relevance of data-driven policies for sustainable urban environmental management.

Future work should focus on improving data quality, applying robust preprocessing (e.g., outlier handling), and exploring advanced models such as ensemble methods or deeper architectures. Incorporating additional variables (e.g., policy factors, urbanization) and enhancing model validation strategies are also recommended to improve generalizability and predictive power.

## ACKNOWLEDGMENT

The authors would like to convey their sincerest appreciation to the Directorate General of Higher Education (RISTEKDIKTI) for providing funding through the Penelitian Dosen Pemula (PDP) scheme. In addition, the author would also like to thank Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) of Universitas Internasional Semen Indonesia (UISI) for providing direction, guidance, and

facilities during the research process. The researcher would also like to thank all parties involved in this research, as without their help and support, this research would not have been completed correctly.

#### REFERENCES

- [1] M. Rizal, "Analisis Pengolahan Persampahan Perkotaan (Studi kasus pada Kelurahan Boya Kecamatan Banawa Kabupaten Donggala)," SMARTek: Sipil, Mesin, Arsitektur, Elektro, vol. 9, no. 2, pp. 155-172, 2011.
- [2] R. F. Widyawati, E. Hariani, A. Y. Hutabarat, N. Rahmawati and S. Oktaviani, "Pelatihan Pemberdayaan Masyarakat dalam Pengelolaan Limbah Rumah Tangga untuk Menjaga Lingkungan," Kreasi: Jurnal Inovasi dan Pengabdian Kepada Masyarakat, vol. 3, no. 2, pp. 268-276, 2023.
- [3] Rondiyah, Sulistiyani and M. Rahardjo, "Faktor-Faktor Yang Mempengaruhi Kinerja Pengelolaan Sampah di Pasar Banjarsari Kota Pekalongan," Jurnal Kesehatan Masyarakat (e-Journal), vol. 2, no. 3, pp. 192-199, 2014.
- [4] S. Subekti and E. Apriyanti, "Pengelolaan Sampah Kawasan Perkotaan Kendal Kabupaten Kendal," Neo Teknika: Jurnal Ilmiah Teknologi, vol. 6, no. 1, 2020.
- [5] I. Oribe-Garcia, O. Kamara-Esteban, C. Martin, A. M. Macarulla-Arenaza and A. Alonso-Vicario, "Identification of influencing municipal characteristics regarding household waste generation and their forecasting ability in Biscay," Waste Management, vol. 39, pp. 26-34, 2015.
- [6] M. Herlambang, Pendapatan per Kapita dan Timbulan Sampah di Indonesia: Analisis Hipotesis Environmental Kuznets Curve, Universitas Padjajaran, 2020.
- [7] E. Damanhuri and T. Padi, Diktat Pengelolaan Sampah, Bandung: Teknik Lingkungan Institut Teknologi Bandung (ITB), 2011.
- [8] R. Gardiner and P. Hajek, "Municipal waste generation, R&D intensity, and economic growth nexus – A case of EU regions," Waste Management, vol. 114, pp. 124-135, 2020.
- [9] E. Sarwono, Y. I. Saputro and B. N. Widarti, "Perencanaan Pengelolaan Sampah Di Kelurahan Muara Jawa Ulu dan Muara Jawa Pesisir Kabupaten Kutai Kartanegara," Jurnal Teknologi Lingkungan, vol. 1, no. 2, pp. 27-30, 2017.
- [10] K. L. H. d. Kehutanan, "Sistem Informasi Pengelolaan Sampah Nasional (SIPSN), 2020. Data Pengelolaan Sampah & RTH (SIPSN)," Kementerian Lingkungan Hidup dan Kehutanan, 2020. [Online]. Available: <https://sipsn.menlhk.go.id/sipsn/public/data/timbulan>. [Accessed 19 August 2024].
- [11] M. A. F. Rakhman, M. Busyairi and A. Kahar, "Analisis Timbulan Dan Komposisi Sampah Perumahan Dan Non Perumahan Wilayah Kabupaten Kutai Kartanegara (Studi Kasus: Kecamatan Anggana)," Jurnal Teknologi Lingkungan UNMUL, vol. 6, no. 2, 2022.
- [12] N. A. Maharani, N. R. Juliardi A.R., A. Amalia and A. U. Farahdiba, "Analisa Kuantitas dan Komposisi Timbulan Sampah Makanan Supermarket di Surabaya," INSOLOGI (Jurnal Sains dan Teknologi), vol. 2, no. 1, 2023.
- [13] D. Nadjih, S. Saputro and M. Madani, "Identifikasi Jumlah dan Faktor Timbulan Sampah Di Kawasan Wisata Malioboro," Nuansa Akademik: Jurnal Pembangunan Masyarakat, vol. 5, no. 1, 2020.
- [14] G. Prajati and A. J. Pesurnay, "Analisis Faktor Sosiodemografi dan Sosioekonomi terhadap Timbulan Sampah Perkotaan di Pulau Sumatra," Jurnal Rekayasa Sipil dan Lingkungan, vol. 3, no. 1, 2019.
- [15] K. Kawai and T. Tasaki, "Revisiting estimates of municipal solid waste generation per capita and their reliability," Journal of Material Cycles and Waste Management, vol. 18, pp. 1-13, 2016.
- [16] U. Aulia and V. Hadju, "Analisis Faktor yang Berpengaruh Terhadap Angka Timbulan Sampah," Jurnal Kolaboratif Sains, vol. 7, no. 6, 2024.
- [17] P. T. T. Trang, H. Q. Dong, D. Q. Toan, N. T. X. Hanh and N. T. Thu, "The Effects of Socioeconomic Factors on Household Solid Waste Generation and Composition: A Case Study in Thu Dau Mot, Vietnam," in 3rd International Conference on Energy and Environment Research, ICEER 2016, Barcelona, 2016.
- [18] C. Chen, "Spatial Inequality in Municipal Solid Waste Disposal Across Regions in Developing Countries," International Journal of Environment Science and Technology, vol. 7, no. 3, 2010.
- [19] I. Arbulú, J. Lozano and J. Rey-Maqueira, "Tourism and solid waste generation in Europe: A panel data assessment of the Environmental Kuznets Curve," Waste Management, vol. 46, pp. 628-636, 2015.
- [20] Ambuj and R. Machavaram, "Optimizing energy expenditure in agricultural autonomous ground vehicles through a GPU-accelerated particle swarm optimization-artificial neural network framework," Cleaner Energy Systems, vol. 9, 2024.
- [21] S. Alzoubi, L. Abualigah, M. Sharaf, M. Sh. Daoud, N. Khodadadi and H. Jia, "Synergistic Swarm Optimization Algorithm," Computer Modeling in Engineering & Sciences, vol. 139, no. 3, pp. 2557-2604, 2024.
- [22] O. Ayeleru, L. I. Fajimi, B. Oboirien and P. Olubambi, "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, vol. 289, 2021.
- [23] T.-H. Nguyen, N.-L. Tran, V.-T. Phan and D.-D. Nguyen, "Prediction of shear capacity of RC beams strengthened with FRCC composite using hybrid ANN-PSO model," Case Studies in Construction Materials, 2023.
- [24] G. Liang, F. Panahi, A. N. Ahmed, M. Ehteram, S. S. Band and A. Elshafie, "Predicting municipal solid waste using a coupled artificial neural network with archimedes optimisation algorithm and socioeconomic components," Journal of Cleaner Production, vol. 315, 2021.
- [25] N. Elshaboury, E. M. Abdelkader, A. Al-Sakkaf and G. Alfalah, "Predictive Analysis of Municipal Solid Waste Generation Using an Optimized Neural Network Model," Processes, vol. 9, no. 11, 2021.
- [26] R. Udayakumar, R. Elankavi, V. Vimal and R. Sugumar, "Improved Particle Swarm Optimization with Deep Learning-Based Municipal Solid Waste Management in Smart Cities," Revista de Gestão Social e Ambiental, vol. 17, no. 4, pp. 1-20, 2023.
- [27] F. A. Tyas, M. Setianama, R. F. Fajriyah and A. Ilham, "Implementation of Particle Swarm Optimization (PSO) to Improve Neural Network Performance in Univariate Time Series Prediction," KINETIK: Game Technology, Information System, Computer Network, Computing, Electronics, and Control, vol. 6, no. 4, 2021.
- [28] K. M. Boehm, P. Khosravi, R. Vanguri, J. Gao and S. P. Shah, "Harnessing multimodal data integration to advance precision oncology," Nat Rev Cancer, vol. 22, no. 2, p. 114–126, 2021.
- [29] J. S. Wekesa and M. Kimwele, "A review of multi-omics data integration through deep learning approaches for disease diagnosis, prognosis, and treatment," Front Genet, vol. 14, 2023.
- [30] S. O. Fadlallah, T. N. Anderson and R. J. Nates, "Artificial Neural Network-Particle Swarm Optimization (ANN-PSO) Approach for Behaviour Prediction and Structural Optimization of Lightweight Sandwich Composite Heliostrats," Arabian Journal for Science and Engineering, vol. 46, p. 12721–12742, 2021.
- [31] S. T. Ampuyak, "Analisis Proyeksi Pertumbuhan Penduduk Dan Kebutuhan Fasilitas Persampahan di Kota Palu 2015-2025," Katalogis, vol. 4, no. 4, 2016.
- [32] A. Widiandi and I. Pratama, "Penanganan Missing Values dan Prediksi Data Timbulan Sampah Berbasis Machine Learning," Rabit Jurnal Teknologi dan Sistem Informasi Univrab, vol. 9, no. 2, pp. 242-251, 2024.
- [33] Djuniadi, M. Afiffudin and W. Lestari, Statistik Inferensial Teori, Aplikasi Dan Latihan Dengan SPSS, Semarang: Universitas Negeri Semarang, 2017.
- [34] S. D. Nurhalijah, N. Cahyati, A. Romadhona, N. Maulani and M. S. Rahayu, "Analisis Korelasi Spearman Untuk Mengetahui Hubungan Antara Penggunaan Media Sosial Dan Tingkat Produktivitas Akademis Mahasiswa Agribisnis (Studi Kasus: Universitas Sultan Ageng Tirtayasa)," Jurnal Ilmiah Wahana Pendidikan, vol. 10, no. 16, pp. 800-809, 2024.