



Predicting Waste Production Trends in Palu City Using Linear Regression Analysis

Moh Randi Labambe^{1*}, Rizka Ardiansyah², Septiano Anggun Pratama¹, Wirdayanti¹

¹ Department of Information Technology, Faculty of Engineering, Tadulako University, Jl. Soekarno Hatta No KM 9 Palu 94148, Central Sulawesi, Indonesia

² Information Systems Department, Faculty of Engineering, Tadulako University Jl. Soekarno Hatta No KM 9 Palu 94148, Central Sulawesi, Indonesia

*Randymhmd73@gmail.com

Abstract . The aim of this research is to obtain predicted results for the volume of waste in Palu City. In Helping the Environmental Agency One of the complex aspects of the waste problem in Palu City is the lack of waste facilities, there is no basic reference for predicting the movement of waste carried out by the Environmental Agency. If this is left unchecked then the waste in Palu City will never be completely resolved. The algorithm chosen is a linear regression algorithm which can help make good predictions, one of which can create waste volume traffic which is useful for knowing the rise and fall of waste volume in each area. in Palu City, so it is a concern. For the community, it is important to protect the environment from waste. Therefore, the Linear Regression Method is used to predict the value of the dependent variable if the independent variable has a value that is known to describe the level of waste pollution. Waste production in the Palu City area to provide information to the Palu City Environmental Service regarding waste production which continues to increase every year. Trash Trends that occur in Palu City using the previous dataset, based on Trend results showing the accumulated volume of waste in Palu City. The highest waste volume occurred in 2017 to 2021, around 350,000 (kg/person/month), the lowest volume occurred in 2021, around 200,000 (kg/person/month). shows that the analysis carried out is as good as possible. The system creation process begins with creating a flowchart, collecting waste volume data, determining an algorithm that can manage waste data, and determining an algorithm that can predict data that shows trends.

Keywords : Linear Regression, Linear Regression Analysis, Waste Volume Prediction, Mean Absolute Error (MSE), Mean Absolute Percentage Error (MAPE)

(Received 2024-05-02, Accepted 2024-06-09, Available Online by 2024-06-11)

1. Introduction

Waste is one of the sources of problems in Indonesia which to this day we are still trying to manage it well. Accumulating rubbish in every area of Palu City also overcomes obstacles related to increasingly narrow land and the effects of unpleasant odors that disturb the people of Palu City, which is still the government's main target in overcoming this. [1] One aspect of the complex waste problem in Palu City is the lack of waste facilities, and there is no basic reference for predicting waste movement by the Environmental Service. If this is allowed to happen then waste management in the city of Palu will never be properly resolved, this could also hamper the efforts of the Palu city government in pursuing Adipura, which so far is the main problem of the Palu City Government regarding waste management which has not been effective and efficient. According to the Environmental Agency, the waste processing process

in the city of Palu is transporting waste from the TPS using a fleet that is delivered directly to the TPA. [2] Transporting rubbish in Palu City every day from afternoon until the end, according to the Environmental Service, the local government only provides two fleet units in each village and not all villages have a rubbish transport fleet, resulting in rubbish in every area of the village. not yet available. [3] has been lifted evenly. Likewise, the trash cans in every sub-district in Palu City were partially damaged. Therefore, the government has made a policy of collecting rubbish in every home yard and business premises [4] .

Therefore, I use a linear regression algorithm to predict waste collection in the next few years to help the government overcome it so that in the future the volume of waste in Palu City will decrease and it will become a city free from waste [5] . Linear regression is a very versatile and powerful algorithm that is ideal for a variety of prediction tasks. Its simplicity, interpretability, and efficiency [6] .

2. Method

2.1. Prediction or Experience

Prediction or forecasting can be described as an activity of predicting what will happen in the future, this activity is carried out using past or current data or information mathematically or statistically [7] . The prediction aims to find out, see and describe the growth of waste in the city of Palu so that later the government can see forecasts using statistics in the results of this research to create awareness and attention in overcoming the waste problem. In designing a prediction or forecasting method, there are three stages that must be gone through, namely:

1. Analyze historical data to obtain pattern and trend data.
2. Choose the available method according to your needs. The choice of method influences the prediction results. Prediction results are measured by calculating the smallest error.
3. Historical data processing using the selected method. If necessary, changes are made as needed.

2.2. Flow System Diagram

A system flow diagram is a flow diagram that shows the sequence diagram of a program. The system that will be created is to predict the population of Palu City and the volume of waste in Palu City using the Linear Regression algorithm.



Figure 1 . System flow diagram

- a. The system starts from problem identification, the aim of this stage is to get an overview of the problem .
- b. After successfully identifying the problem, the next step is to determine a solution to the problem.

What is done at this stage includes conducting a literature study related to the problems raised to find out what kind of problem it is along with an overview of the solutions that have been widely developed.

- c. After successfully identifying the problem, the next step is the Data Collection Stage obtained from the Palu City Environmental Service through interviews and data collection observations [8].
- d. After successfully collecting data, at this stage the aim is to select and select the data and attributes that need to be used in waste management research.
- e. After successful data selection, at this stage it is used to manipulate the raw data that has been obtained and data transformation techniques help reduce complex dimensional variables to lower dimensions but still contain important information, thus producing one input for data processing [9].
- f. After successful data transformation, this stage aims to produce a more complete and meaningful data set, so that it can be used for analysis or decision making. At this stage an approach will be taken to model the relationship between a dependent variable and one or more independent variables.
- g. After that, testing the Mean Square Error (MSE) and Mean Absolute Error (MAE) is one of the methods used to emit a forecasting model or regression model [10].
- h. A good evaluation process involves dividing the data into training data and testing data, training the model on the training data, and then testing the model's performance on the testing data to ensure that the model is not overfitting and can generalize well to an unprecedented point [11].

2.3. Dataset

A dataset is a collection of objects and the properties or characteristics of the objects themselves. The dataset used in this research is waste generation data from 2017 to 2021. In 2021, the volume of waste in the Palu City area, with the amount of waste each year amounting to 71,206,025.00 tons. In 2020, the volume of waste in the Palu City area, with the amount of waste each year amounting to 67,959.90 tons. In 2019, the volume of waste in the Palu City area with the amount of waste each year was 71,204.38 tons. In 2018, the volume of waste in the Palu City area was 70,373.83 tons each year. In 2017 the volume of waste in the Palu City area with the amount of waste each year was 69,310.40 tons.

That is the total volume of waste according to data from the Environmental Agency. The dataset used can be added or changed according to predictions by the Environmental Agency and the author is permitted to use the volume of waste data in this research.

2.4. Linear regression

Linear regression is a statistical method used to model the relationship between a dependent variable and one or more independent variables. In this research I chose the linear regression method to predict the rise and fall of waste production on a graph as a research result for the government's attention. The reason I took the linear regression method is that it is easy to understand and implement, even for people who do not have a strong statistical background, linear regression can produce accurate predictions for the amount of waste, especially if the relationship between the independent variable and the dependent variable is indeed linear. the value of a dependent variable is based on the values of the independent variable [12]. Linear regression is appropriate to use in this research, where this predicted value is of concern and urgency for the government in the context of handling waste in the Palu City area. Linear regression can be interpreted as a statistical model that analyzes the linear relationship between a dependent variable and a certain choice of independent variables, as for the linear regression model:

$$Y = mX + b$$

Y is the attachment variable we are trying to predict

X is the dependent variable used to make predictions.

m is the slope of the regression line that represents the influence of X on Y

b is a constant called the Y intercept. If X = 0, Y is equal to

2.5. Mean Squared Error

Mean Square Error (MSE) testing is one of the methods commonly used to transmit the performance of forecasting models or regression models [13]. MSE is carried out in two parts, namely Training data and Testing data. Training data is used to calculate MSE as test data to get a general picture of model performance. And for data testing, namely testing the model that will be created using the Matrix that will be used [14]. The calculation is done by taking each pair of predicted values and actual values, calculating the difference, then squaring them. After all differences are squared, calculate the average of all differences squared [15]. The test indicator is that the smaller the MSE value, the better the MSE test results, while MAE measures the average absolute difference between the actual value and the predicted value. The smaller the MAE value, the better the model's performance in predicting the actual value [16]. MAE has a more intuitive interpretation than Mean Squared Error (MSE), because MAE directly measures the error in the same units as the observation data [17].

If the test results are close to the minimum value of 0, the smaller the MSE value indicates the greater the match between the predicted value and the actual value. The Mean Squared Error (MSE) formula is as follows:

$$MSE = \frac{\sum_{t=1}^n (A_t - F_t)^2}{n}$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

A_t = Actual Value of the request

F_t = Prediction result value

n = number of data

From the MAE and MSE calculations, the actual value is reduced by the predicted value then divided by the actual value which is used as an absolute value and multiplied by 100. The grouping of MAPE and MSE value categories can be seen in the following table [18].

UMK value	Category
<10%	Very accurate
10-20%	Good
20-50%	Reasonable
>50%	Not accurate

3. Results and Discussion

3.1. Waste Volume Trends

After testing, there was an increase in overall waste volume from 2017 to 2021. The most significant increase occurred in 2020, with waste volume reaching 350,000 Kg/Person/Month [19]. This increase is likely caused by several factors, such as:

- a. Increasing population growth in an area can result in more waste.

- b. Increased consumption Increased consumption of goods and services, especially products that produce a lot of waste, such as plastic and packaging.
- c. Lack of public awareness Lack of public knowledge and awareness about good waste management.

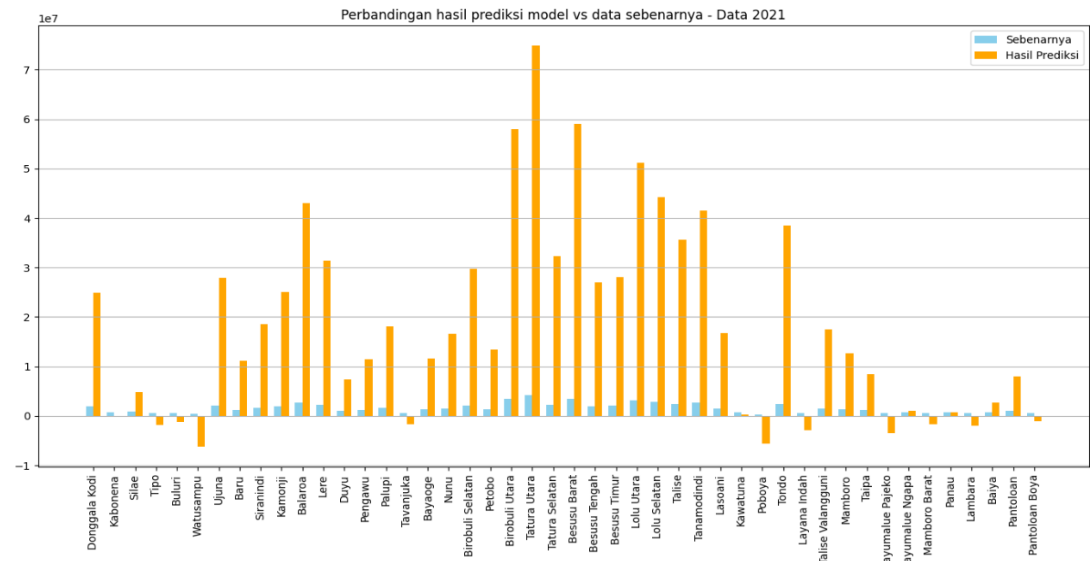


Figure 2. Waste Trends

The image above shows the waste trend that occurred in Palu City using the previous dataset. Based on the results of the trend above, it shows the high volume of waste in Palu City. The highest waste volume occurred in 2017 to 2021, around 350,000 (kg/person/month), the lowest volume occurred in 2021, around 200,000 (kg/person/month).

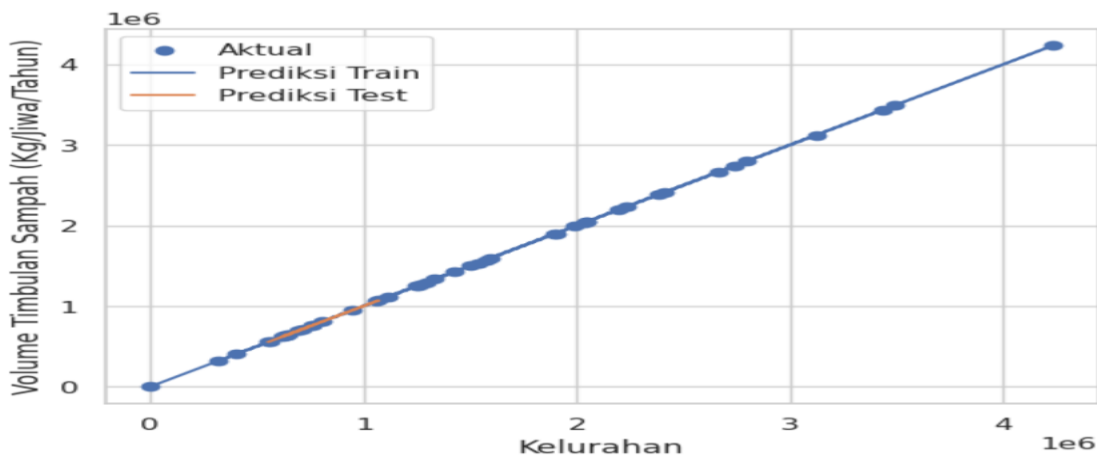


Figure 3. Prediction Graph

In the image above, the prediction graph is compared with the actual value using the Prediction Train and Prediction Test models for the functions of these two models to help obtain and obtain model accuracy. Where the X axis = Village and on the Y axis = Volume of Waste Generation (Kg/Person/Year), the actual value is round, the Predicted Value is a blue line and for the prediction test the line is orange.

Test score: 0.9248228556076293
 Train score: 0.9274866966401237

The results were obtained by comparing the Actual Values and Predicted Values using two prediction models, namely Train Prediction and Test Prediction.

3.2. Discussion

The results of this research indicate that the analysis carried out was as good as possible. The system creation process begins with creating a flowchart, collecting waste volume data, determining an algorithm that can manage waste data, and determining an algorithm that can predict data that shows trends. The algorithm chosen is a linear regression algorithm which can help make good predictions, one of which can create waste volume traffic which has the benefit of knowing the rise and fall of waste volume in each area in Palu City, so that it becomes a concern for the community that it is important to protect the environment from waste. After that, this research uses the MSE and MAE methods for data modeling so that it can see good prediction results and accuracy.

This research has several limitations that need to be considered so that they can be improved in future research, namely scope limitations, data limitations, generalization limitations, affordability limitations, and time limitations. By considering these limitations, it is hoped that further research can reveal additional factors that might influence prediction results.

4. Conclusion

After testing, there was an increase in overall waste volume from 2017 to 2021. The most significant increase occurred in 2020, with waste volume reaching 350,000 kg/person/month. Forecasting waste volume using the linear regression method can be said to be quite good. This is based on the results of the MSE, MAE tests on all existing historical data. However, good prediction results can only be obtained with data that has a significant range from one value to another, or does not experience interference or drastic decline. The linear line itself will only point up or down, so that the results obtained will only be an increase or decrease, without experiencing ups and downs and vice versa. This is most likely due to the period being a variable used in linear regression calculations, so adding other variables in making predictions will most likely further strengthen the prediction value obtained. The results obtained from this research are only estimates based on data over a certain period of time. However, developers can include this information as a consideration in making decisions when reducing and maximizing sectors that are experiencing improvement. Based on the forecasting results using the Linear Regression algorithm for MAE and MSE calculations, it can be concluded that the predictions are 100% accurate. The prediction results can provide an idea for the government to always pay attention to areas that produce the most waste so that it can be handled better [20].

5. Thank-you note

The researcher would like to thank Tadulako University for the extraordinary opportunity to participate in the Independent Study Campus Program (MBKM), especially in the Independent Study Program. Researchers appreciate the support, guidance and inspiring learning environment provided by lecturers, staff and the entire academic community. This experience is a valuable journey that provides new insights, skills and opportunities for researchers' self-development. Thank you for Tadulako University's dedication in providing an innovative educational platform and providing space for researchers to develop independently.

References

- [1] . G., RH Laluma, and A. Prasetya, "Prediction of Waste Management Volume and Rates in Bandung City Using the Linear Regression Method," *Techno-Socio Ekon.*, vol. 15, no. 1, p. 49, 2022, doi: 10.32897/techno.2022.15.1.1195.
- [2] A. Mathematics, "Prediction of Talangagung Landfill Waste Volume Using a Dynamic Systems Approach," pp. 1–23, 2016.
- [3] I. Irwan, FN Ichsan, N. Gistituati, and S. Marsidin, "Education Policy Analysis Regarding Learning Implementation During the Covid 19 Emergency Period," *J. Manaj. Educator.*, vol. 9, no. 2, pp. 89–95, 2021, doi: 10.33751/jmp.v9i2.4238.
- [4] A. Hidayat, MT Ziyad, and C. Juliane, "Predicting Waste Volume in West Java Using the Linear Regression Method," vol. 10, no. 1, 2023.

- [5] JA Usuary, HR Ngemba, S. Hendra, and A. Kasim, "Wina Sentosa Bottled Water Distribution System Using Web-Based Distribution Requirement Planning and Trend Moment Algorithms," no. March 2023, pp. 1–9.
- [6] EE Nugraheni, "Prediction of Capacity of Cipeucang Final Disposal Site (TPA) South Tangerang City in 2031," *UIN Syarif Hidayatullah*, p. 156, 2017.
- [7] M. Audina, "Prediction and Analysis of Final Waste Disposal Sites (Tpa) in Padang City," *J. Buana*, vol. 2, no. 2, p. 423, 2018, doi: 10.24036/student.v2i2.93.
- [8] I. Djakaria, S. Guritno, and SH Kartiko, "Visualization of Iris Data Using Principal Component Analysis and Kernel Principal Component Analysis Visualization of Iris Data Using Principal Component Analysis and Kernel Principal Component Analysis," *J. ILMU BASIC*, vol. 11(1), no. 1, pp. 31–38, 2010.
- [9] MAH Putra, M. Mutiani, J. Jumriani, and MRN Handy, "The Development of a Waste Bank as a Form of Community Participation in Waste Management," *Kalimantan Soc. Stud. J.*, vol. 2, no. 1, p. 22, 2020, doi: 10.20527/kss.v2i1.2460.
- [10] S. Kosasi, S. Pontianak, J. Merdeka, N. 372 Pontianak, and K. Barat, "Application of the Backpropagation Neural Network Method to Predict School Exam Scores," *J. Teknol.*, vol. 7, no. 1, pp. 20–28, 2014, [Online]. Available: <https://journal.akprind.ac.id/index.php/jurtek/article/view/1012>
- [11] SS Mukrimaa *et al.*, " *J. Researcher. Educator. School Teacher. Basics*, vol. 6, no. August, p. 128, 2016.
- [12] AA Hardy, AA Rohmawati, and SS Prasetyowati, "Prediction of Bandung City Traffic Classification Using Machine Learning and Spatial Analysis," *J. Media Inform. Budidharma*, vol. 6, no. 4, p. 1861, 2022, doi: 10.30865/mib.v6i4.4538.
- [13] NR Feta and AR Ginanjar, "Comparison of the Kernel Function of Support Vector Machine Methods for Classification Modeling Against Comparison of the Kernel Function of Support Vector Machine Method for Modeling Classification of Soybean Plat Disease," *J. Ilm. Computer Science, Science and Technology. Apply.*, vol. 1, no. 1, pp. 33–39, 2019.
- [14] NS Fauziah and RD Dana, "Implementation of the Naive Bayes Algorithm in Classifying the Welfare Status of the Gunungsari Village Community," *Blend Sains J. Tek.*, vol. 1, no. 4, pp. 295–305, 2023, doi: 10.56211/blendsains.v1i4.234.
- [15] SHS Herho, "Data Visualization Tutorial Using Seaborn," *Itb*, pp. 1–75, 2019.
- [16] M. Daenie and H. Rizqi, "Implementation of Waste Management Policy in the City of Semarang," *Public Serv. Gov. J.*, vol. 2, no. 01, p. 96, 2021, doi: 10.56444/psgj.v2i01.1969.
- [17] GN Ayuni and D. Fitrihanah, "Application of the Linear Regression method to predict property sales at PT XYZ," *J. Telemat.*, vol. 14, no. 2, pp. 79–86, 2019, [Online]. Available: <https://journal.ithb.ac.id/telematics/article/view/321>
- [18] Normah, B. Rifai, S. Vambudi, and R. Maulana, "Vtuber Development Sentiment Analysis Using the SMOTE-Based Support Vector Machine Method," *J. Tek. Comput. AMIK BSI*, vol. 8, no. 2, pp. 174–180, 2022, doi: 10.31294/jtk.v4i2.
- [19] WAL Sinaga, S. Sumarno, and IP Sari, "The Application of Multiple Linear Regression Method for Population Estimation of Mount Malela District," *JOMLAI J. Mach. Learn. Artif. Intel.*, vol. 1, no. 1, pp. 55–64, 2022, doi: 10.55123/jomlai.v1i1.143.
- [20] Yuli Mardi, "Data Mining: Classification Using the C4 Algorithm. 5 Data mining is part of the Knowledge Discovery in Database (KDD) process stage. Edik Informatics Journal," *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.