Visualization of Clustering and Forecasting the Availability of Waste Disposal Land in North Penajam Paser City and the IKN Sector Using Advanced Analysis Methods and Regression Tests.

Dwi Arief Prambudi S.Kom., M.Kom 1), Nia Febrianti, S.T., M.T 2) Mufid Ubadillah 3)

*ORCHID IDs:https://orcid.org/0000-000X-XXXX-XXXX

[1)]*Information Systems, Kalimantan Institute of Technology, Indonesia*
[2)]*Environmental Engineering, Kalimantan Institute of Technology, Indonesia*
[3)]*Information Systems, Kalimantan Institute of Technology, Indonesia*

*Dwiariefprambudi@lecturer.itk.ac.id*1),niafebrianti@lecturer.itk.ac.id2),10201058@student.itk.ac.id2)*

Abstract

Waste production will continue to increase along with population growth. With the increasing population, social needs such as consumption will also increase and will affect the capacity of the Final Processing Site which is increasing. According to data from the Central Statistics Agency of Penajam Paser Utara in the 2010-2022 population census, the population of Penajam Paser Utara continues to increase with an average of 3,576 people per year. This population growth will be followed by a large level of waste production in this area. And Penajam Paser Utara only has one Final Processing Site (TPA). This study uses a linear regression algorithm method to predict the amount of waste growth and estimate the availability of landfill land in Penajam Paser Utara until 2029. The estimated results show that the total waste production reaches 306,242.4 tons with a population of 1,378,141 people. Further analysis indicates that the estimated remaining landfill land is experiencing a deficit of 187,161.6 m². In conclusion, the land area at Buluminung TPA will not be sufficient to accommodate the projected volume of waste until 2029. Evaluation of the linear regression model using RMSE shows an error rate of 40.27, which provides an overview of the accuracy of the model's predictions.

Keywords:*Data Mining, Estimation, Linear Regression, Waste, Landfill,*

1. Introduction

Good waste management is essential to create a healthy environment and support the quality of life of the community. A healthy area is characterized by how well the existing waste management is. Over time, population growth has a direct impact on waste production, which affects the capacity of the Final Disposal Site (TPA). Waste consists of various types such as plastic, paper, cans, glass, polystyrene, wood, leaves, and others. Inadequate waste management can cause serious environmental problems. According to Law Number 18 of 2008 concerning waste management, TPA is a place where waste is processed and returned to the environmental media safely for humans and the environment.[1]

Penajam Paser Utara, which will become the new capital city of Indonesia, has experienced a significant increase in population and waste production. Based on interviews and data from the Environmental Agency, the Buluminung landfill in Penajam Paser Utara has been operating since 2012 with a land area of 1.7 hectares. This land is divided into two zones, namely zone 1 and zone 2, with areas of approximately 1.1 hectares and 0.7 hectares respectively. Initial estimates indicate that the Buluminung landfill can operate for 26 years, but the unexpected increase in population requires a re-estimation of the availability of land for waste disposal. Data from the Central Statistics Agency shows that population growth in Penajam Paser Utara is an average of 3,576 people per year, which results in an increase in waste production.

According to data from the Waste Management Information System (SIPSN) of the Ministry of Environment and Forestry, the amount of waste production in Penajam Paser Utara in 2021 reached 33,766.52 tons with a population of 180,657 people. In 2022, waste production increased to 35,033.98 tons with a population of 183,043 people. Therefore, a solution is needed to estimate the availability of accurate waste disposal land in order to support effective waste management planning.[3]

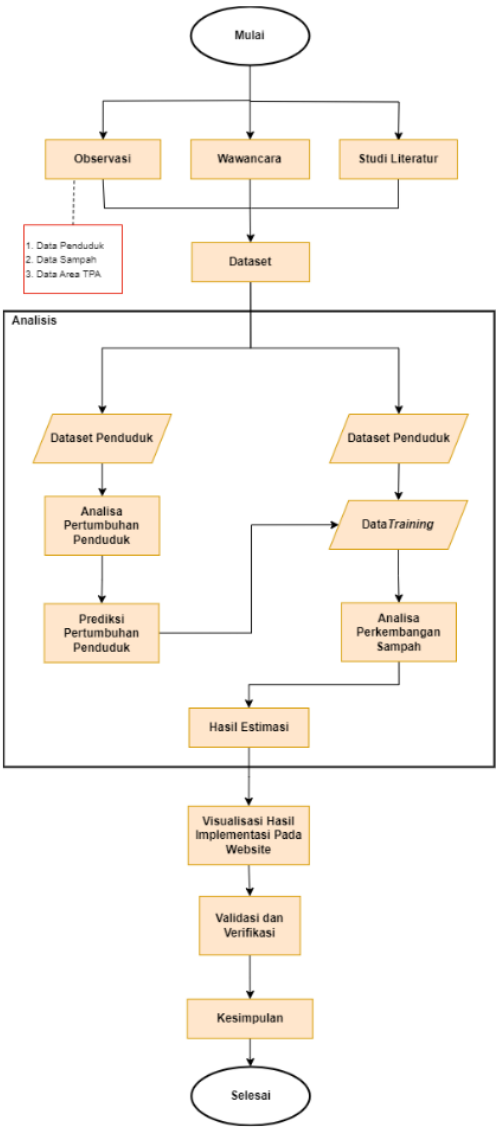
Estimation is one of the techniques of data mining used to estimate the population value based on sample data. This study uses the linear regression method to estimate the need for landfill land in the future. By analyzing population data and waste volume each year, the relationship between population growth and waste volume can be understood. The analysis tools used are Python and RapidMiner, which help in predicting the amount of waste in the future.[2]

Therefore, this research is expected to provide new knowledge that is useful for sustainable waste management in North Penajam Paser, so that waste disposal areas can be managed more efficiently and effectively.[5]

2. Method

This section contains a brief description of the research methods used during the research. The contents of this method include the research outline, research flow diagram, and research procedures.

2.1. Research Flow Chart



This study aims to estimate the availability of landfill sites in Penajam Paser Utara using a linear regression algorithm. This study is divided into six stages. The first stage is observation and interviews to collect data on population, waste, and landfill area through interviews with the Environmental Service and data access from SIPSN and BPS Penajam Paser Utara. The collected data is then selected, focusing on the last six years (2018-2023). Furthermore, researchers analyze population growth to predict population data for 2024-2029. After that, predictions of waste volume development are carried out using linear regression. The prediction model is evaluated using Root Mean Square Error (RMSE) to assess accuracy. Based on the predictions of population growth and waste volume, the estimated availability of landfill sites for 2029 is then calculated.

2.2. Dataset

At this stage, relevant datasets are collected in the research and will be used as future analysis. Data is obtained through several sources such as interviews, BPS and SIPSN websites. The data produced are:

Table 1. North Penajam Paser Population Dataset

Year	Total population
2018	159386
2019	160912
2020	178681
2021	180656
2022	183043
2023	196556

The dataset contains the year and population in units of people with a range of years 2018 to 2023.

Table 2. North Penajam Paser Waste Generation Dataset

Year	Daily Waste Generation	Annual Waste Generation
2018	136.60	44,495.40
2019	107.91	39,386.69
2020	109.29	39,889.94
2021	92.51	33,766.52
2022	95.98	35,033.98
2023	98.28	35,873.30

The dataset contains the year, the amount of daily waste generation and the amount of annual waste generation, the generation referred to refers to the amount or volume of waste produced.

Table 3. North Penajam Paser Landfill Dataset

Location	Land area	Height of Garbage Pile
Blumium	1.7 Ha	1.5 tons/m^2

The dataset contains information about the location of the landfill, the area of land available, and the height of the waste pile at that location.

2.3. Population Growth Analysis

At this stage is the stage of population growth analysis in North Penajam Paser. The analysis will be based on population data from 2018 to 2023. Using the equation.

$$x = \frac{jumlah\ penduduk\ tahun\ n - jumlah\ penduduk\ tahun\ awal}{jumlah\ penduduk\ tahun\ awal} X100\%$$

Then, after getting the population growth results, the next step is to find the average increase in population per year using the equation.

$$r = \frac{x1 + x2 + x3 + x4 + x5}{n}$$

Population increase per year, variable 'x' refers to the amount of population growth each year for population growth prediction, and 'n' is the time period covering the years analyzed in the dataset.

2.4. Population Growth Prediction

After knowing the population growth and average in the previous year, the next step is to predict population growth in 2024 to 2029, using the equation.

$$Pn = Pa\ (1 + r)^n$$

With the caption 'Pn' refers to the Population Number of the Year 'n', while 'Pa' is the Population of the Initial Year. Then 'r' indicates the average population growth each year in percentage (%). Finally, 'n' is the projection time interval in years, which indicates the time period over which we will project population growth.[4]

2.5. Waste Development Estimate

At this stage, an estimation of waste development will be carried out using a simple linear regression algorithm. Based on the results of the population growth prediction of Penajam Paser Utara for the next 6 years, an estimation of waste development can be carried out using the equation.

$$y = a - bx$$

With the following information:
y= Variable Response (Dependent)
x= Variable Predictor (Independent)
a= Constant (Intercept)
b= Regression Coefficient (Response)

Before making an estimate, first find the constant value (a) and the regression coefficient value (b), using the equation.

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{n(\Sigma xy)(\Sigma x^2) - (\Sigma x)(\Sigma xy)}{n(\Sigma x^2) - (\Sigma x)^2}$$

With the following information:
Σ= Total Amount
y= Total Value Sum
x= Amount of Waste
n= Data Length

After getting the results (a) And (b), then it has been able to apply the linear regression equation model. With this model, it can predict the amount of waste produced in the period 2024 to 2029. Furthermore, evaluate the model using Root Mean Square Error (RMSE), using the equation.

$$RMSE = \sqrt{\frac{\sum_{t-1}^{n}(At - Ft)^2}{n}}$$

With the following information:
At = Actual Data Value
Ft = Forecast Result Value
n = Number of Data
Σ= Total Value Sum

RMSE will determine how close the prediction model is to the original data, the smaller the RMSE value, the better the model's performance in predicting waste development.

2.6. Estimation Results

With the results of the linear regression calculation projection on the amount of waste that has been done, the next step is to estimate the availability of land for waste disposal. By dividing the predicted results of the total volume of waste that has been produced by the height of the waste pile at the TPA. This stage calculates how much land is needed to accommodate the volume of waste produced in the next few years. Then, the area of land for landfills that have been used until 2029 will be reduced from the total land area of landfills.

3. Results and Discussion

This section explains the results and discussion of the research stages carried out.

3.1. Population Growth Analysis

Population, the analysis will be based on previous year's population data. Population growth analysis is carried out by utilizing calculations using excel and data processing with the Python programming language.

a. Manual Calculation

In analyzing population growth with Excel, the analysis uses equations 1 and 2 to find the average, and the average population growth result is 4.36%.

Table 4. Population Growth Analysis

Year	Population (Souls)	Population growth	Percentage (%)
2018	159386	0	
2019	160912	1526	0.96
2020	178681	17769	11.04
2021	180656	1975	1.11
2022	183043	2387	1.32
2023	196556	13513	7.38
Σ		37170	21,807
Average		7434	4.36

b. Python Calculations

By using the same equation in manual calculations, the calculation results in Python have results equivalent to manual calculations.

```
Analisis Pertumbuhan Penduduk Tahunan :
Tahun 2019: Pertumbuhan 1526 orang (0.96%)
Tahun 2020: Pertumbuhan 17769 orang (11.04%)
Tahun 2021: Pertumbuhan 1975 orang (1.11%)
Tahun 2022: Pertumbuhan 2387 orang (1.32%)
Tahun 2023: Pertumbuhan 13513 orang (7.38%)

Total Pertumbuhan Penduduk: 37170 orang
Total Persentase Pertumbuhan Penduduk: 21.81%
Rata-rata Peningkatan Penduduk: 7434.00 orang per tahun
Rata-rata Persentase Peningkatan Penduduk: 4.36% per tahun
```

Figure 1. Population Growth Analysis

In Figure 1, the average population growth results obtained from the Python calculation are 4.36%, equivalent to the results of manual calculations.

3.2. Population Growth Prediction

At this stage, population growth predictions are made, based on previous year's population data and its average. Population growth predictions are made by utilizing calculations using excel and data processing with the Python programming language.

a. Manual Calculation

In predicting population growth with Excel predictions using equation 3, the results of the population growth prediction are as follows.

Table 5. Population Growth Analysis

Year	Projection Time Interval (n)	Population Prediction

2024	6	205899
2025	7	214876
2026	8	224245
2027	9	234022
2028	10	244225
2029	11	254874
Σ		1378141

Table 5 shows the total predicted population until 2029 reaching 1,378,141 people.

b. Python Calculations
By using the same equation in manual calculations, the calculation results in Python have slight differences between the manual calculation results and Python. This difference is caused by different rounding methods for rounding numbers.

```
Tingkat pertumbuhan penduduk tahunan (r) : 0.0436
Prediksi Pertumbuhan Penduduk (2024-2029:)
   Tahun  Prediksi Penduduk
0   2024             205129
1   2025             214077
2   2026             223414
3   2027             233159
4   2028             243329
5   2029             253943
```

Figure 2. Population Growth Prediction

In Figure 1, the average population growth results obtained from the Python calculation are 4.36%, equivalent to the results of manual calculations.

3.3. Waste Development Analysis

After carrying out the population growth prediction stage, the next step is to analyze the development of waste using the results of the population growth prediction from 2024 to 2029. The analysis will go through a simple linear regression calculation process which will be implemented using 3 tools, namely Excel, RapidMiner, and Python.

a. Manual Calculation
In analyzing waste development, the analysis uses a simple linear regression model with equation 4, and to find the values (a) and (b) using equations 5, 6. The results of waste development are as follows.

Table 6. Waste Development Prediction

Year	Waste Development Prediction (Y)
2024	45734.97
2025	47736.84
2026	49826.13

2027	52006.4
2028	54281.67
2029	56656.39
Total	306242.4

The results obtained from manual calculations using Excel obtained a total predicted waste development result until 2029 reaching 306,242.4 tons.

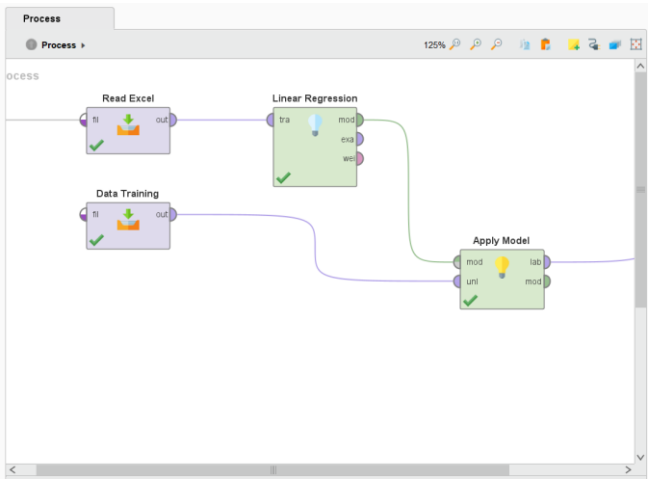


Figure 3. RapidMiner Waste Development Analysis

By using Training Data, which is data containing the population and volume of waste in 2018 to 2023. And the following results were obtained.

Row No.	Y	prediction(Y)	X
1	?	45963.151	205899
2	?	47974.971	214876
3	?	50074.641	224245
4	?	52265.747	234022
5	?	54552.324	244225
6	?	56938.853	254874

Figure 4. RapidMiner Trash Development Results

In Figure 4, the results of RapidMiner implementation in waste development analysis are obtained. There is a slight difference between the results of manual calculations and RapidMiner due to different rounding methods. Then after getting the results of waste development, the next step is to evaluate the model using Root Mean Square Error (RMSE) in RapidMiner.

Figure 5. RapidMiner RMSE Evaluation

By using the Root Mean Square Error (RMSE) model evaluation. The RMSE result is 40.274. This result will be compared with the RMSE in python.

The results obtained from manual calculations using Excel obtained a total predicted waste development result until 2029 reaching 306,242.4 tons.

b. RapidMiner Calculation
In the RapidMiner implementation, waste development analysis uses various types of operators, namely ReadExcel, Exampleset, LinearRegression, and ApplyModel.

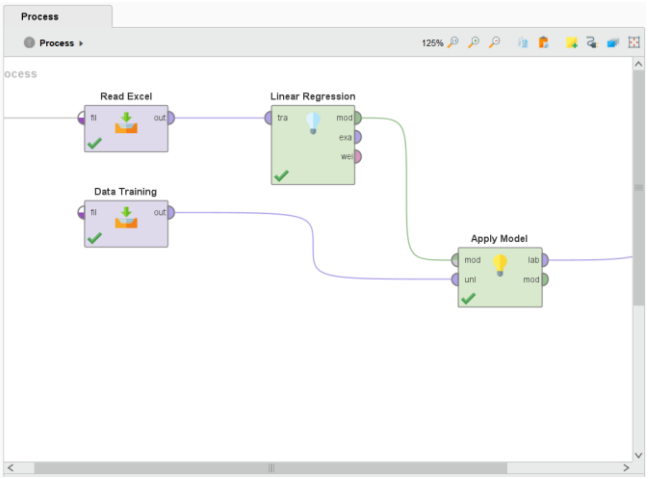


Figure 3. RapidMiner Waste Development Analysis

By using Training Data, which is data containing the population and volume of waste in 2018 to 2023. And the following results were obtained.

Row No.	Y	prediction(Y)	X
1	?	45963.151	205899
2	?	47974.971	214876
3	?	50074.641	224245
4	?	52265.747	234022
5	?	54552.324	244225
6	?	56938.853	254874

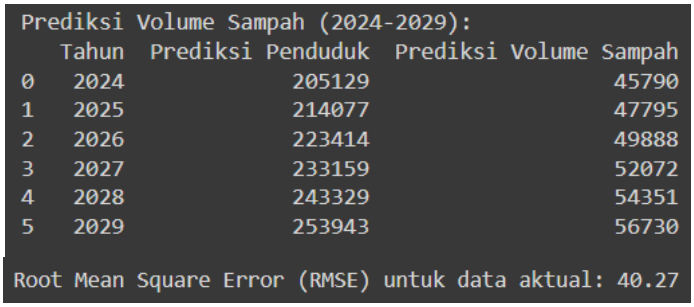
Figure 4. RapidMiner Trash Development Results

In Figure 4, the results of RapidMiner implementation in waste development analysis are obtained. There is a slight difference between the results of manual calculations and RapidMiner due to different rounding methods. Then after getting the results of waste development, the next step is to evaluate the model using Root Mean Square Error (RMSE) in RapidMiner.

Figure 5. RapidMiner RMSE Evaluation

By using the Root Mean Square Error (RMSE) model evaluation. The RMSE result is 40.274. This result will be compared with the RMSE in python.

c. Python Calculations
In the python implementation, the analysis of waste development will go through a linear regression process assisted by the 'sklearn' library and calling the 'LinearRegression' model. And evaluating the model using Root Mean Square Error (RMSE).



Prediksi Volume Sampah (2024-2029):

	Tahun	Prediksi Penduduk	Prediksi Volume Sampah
0	2024	205129	45790
1	2025	214077	47795
2	2026	223414	49888
3	2027	233159	52072
4	2028	243329	54351
5	2029	253943	56730

Root Mean Square Error (RMSE) untuk data aktual: 40.27

Figure 6. Garbage Growth Analysis in Python

In Figure 6, the results of the development of waste that occurred in Penajam Paser Utara using Python are obtained, there is a slight difference with manual calculations and RapidMiner, this is due to the method used in rounding the numbers but the difference is not too significant. And the RMSE results obtained error of 40.27 this result is in accordance with the evaluation of the model using RapidMiner.

3.4. Visualization of Implementation Results on Website

After predicting the population and estimating the development of waste volume from 2024 to 2029 in Penajam Paser Utara, the next step is to build a system that can automate the prediction and estimation process. This system is designed to facilitate the collection, analysis, and visualization of data related to population growth and waste production, as well as to produce predictions based on existing historical data.

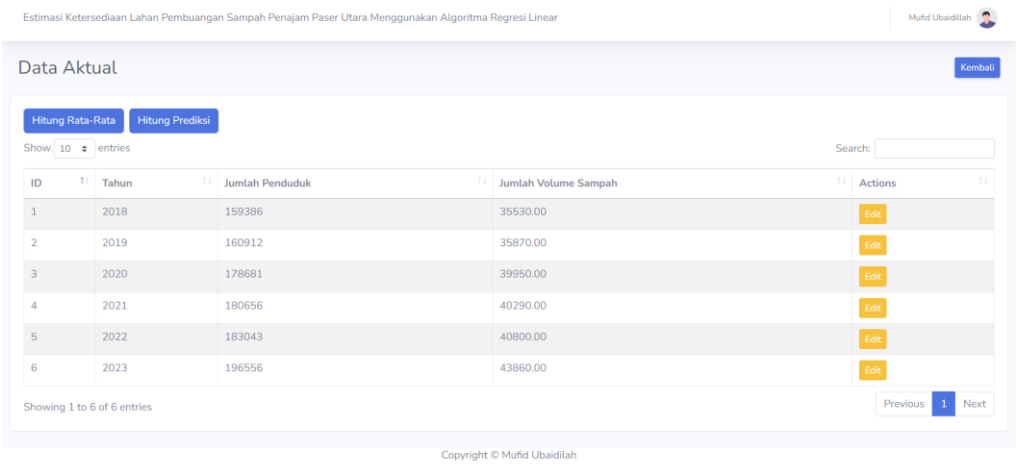


Figure 7. Population and Waste Dataset

This system will input actual population data and also waste that will be processed and will produce a visualization of the predicted results with visualization of actual data from the previous year.

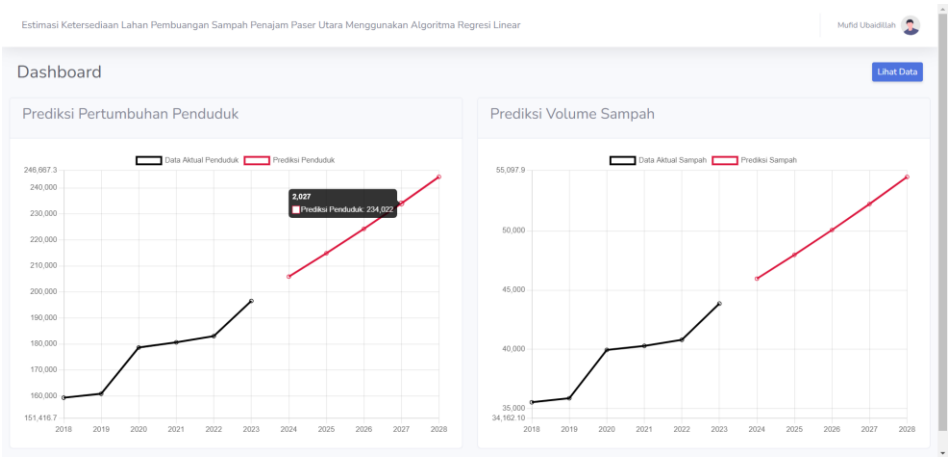


Figure 8. Visualization of Prediction Results

After inputting data and making predictions, the system will visualize the results of the data on the dashboard page. The data displayed includes actual data and predicted data as in Figure 8.

3.5. Estimation Results

After analyzing the development of waste in Penajam Paser Utara, the next step is to determine the availability of land for waste disposal at the Buluminung TPA. With a land area of 1.7 Ha (17,000 m²) and a waste pile height of 1.5 tons/m², and a predicted waste volume of 306,242.4 tons until 2029, the used land area is calculated at 204,161.6 m². By subtracting the available land area from the used land area, the result is -187,161.6 m², indicating that the Buluminung TPA land is not sufficient to accommodate waste until 2029.

4. Conclusion

Based on the results of the analysis and discussion using data mining techniques with linear regression algorithms, as well as analysis with Excel, RapidMiner, and Python, this study concludes several things. First, the estimation using a simple linear regression algorithm shows a stable and low error prediction, with a total volume of waste until 2029 reaching 306,242.4 tons and a population of 1,378,141 people. This estimate indicates a deficit of land for waste disposal of 187,161.6 m^2 at the Buluminung TPA. Evaluation of the prediction model using RMSE shows an error rate of 40.27, illustrating the accuracy of the model in predicting the amount of waste.

References

[1] Yanto, R., Studi, P., Informasi, S., Mik, ST, Nusantara, B., & Lubuklinggau, J. (2018a). Implementation of Data Mining Estimation of Land Availability for Waste Disposal Using Simple Linear Regression Algorithm. RESTI Journal (System Engineering and Information Technology), 2(1), 361–366. https://doi.org/10.29207/RESTI.V2I1.282

[2] Yustikarini, R., & Setyono, P. (2017). Evaluation and Study of Waste Handling in Reducing the Burden of Final Waste Processing Sites at Milangasri TPA, Magetan Regency. Proceeding Biology Education Conference: Biology, Science, Environmental, and Learning, 14(1), 177–185. https://jurnal.uns.ac.id/prosbi/article/view/17642

[3] Yususf SE, M. (nd). North Penajam Paser Regency in Figures 2023 (M. Yususf SE, Ed.). BPS North Penajam Paser.

[4] Saitullah, MI (2022). Correlation of Population and The High Pollution of Household Waste in Fakkie Village Pinrang Regency. Continuum: Indonesian Journal Islamic Community Development, 1(I), 8–20. https://ejurnal.iainpare.ac.id/index.php/continuum/article/view/3380

[5] Yurda, M., Yorika, R., Zonasi, P., Pengumpulan, P., Kawasan, S., & Traktor, G. (2022). Determination of Zoning Patterns of Waste Containment and Collection in the Traktor Mountain Area. Journal of PROJEKSI: Architecture and Planning, 2(2), 10–15. https://doi.org/10.51454/PROJEKSI.V2I2.862