JOURNAL OF DATA SCIENCE | Vol.2023:10
eISSN:2805-5160

Waste Prediction in Gross Pollutant Trap Using Machine Learning Approach

Elpina Sari[1] and Tri Basuki Kurniawan[1]

[1] Magister of Information Technology, University of Bina Darma, Palembang, Indonesia

[*]**Email**: tribasukikurniawan@binadarma.ac.id

Abstract

Urbanization is often associated with decreased rainwater quality due to many factors, such as uncontrolled pollution and waste disposal. Therefore, managing water quality impacts in urban areas must be addressed to protect our environment. One of the maintenance steps is installing gross pollutant traps (GPT). The main objective of GPT is to remove dirty pollutants that are carried into the rainwater system before the rainwater enters the main river channel. At the same time, it is essential to understand that tropical climates are always associated with high rainfall intensity in a short period. It means that the amount of waste that GPT daily captures cannot be predicted well. It causes other problems, namely the emergence of difficulties in predicting the amount of waste that must be transported and moved from the GPT location to the final waste disposal site so that often, the rubbish that the GPT has caught will pile up at the GPT location without being able to be transported properly. Because the garbage vehicles that must transport the garbage are insufficient in number and capacity, it is necessary to have a model that can accurately predict the amount of waste that may be captured by each GPT based on past data on the amount of garbage that has been captured. This research compares 3 algorithms for predicting the amount of waste trapped by GPT: Simple Linear Regression, Multiple Linear Regression, and Polynomial Regression. The results show pretty good accuracy in our model, which is the RMSE is 1000. Next, a simple application was developed to lead the implementation of a load optimization scenario to show the importance of predicting the number of rubbish traps by each GPT by calculating how many trucks should be used to carry the garbage to the final waste disposal site.

Keywords

Gross Pollutant Devices, Gross Pollutant Traps, Simple Linear Regression, Multiple Linear Regression, Polynomial Regression

Introduction

Urbanization is often associated with decreased rainwater quality due to many factors, such as uncontrolled pollution and waste disposal. It leads to an increase in the socio-economic life of an area but also brings various environmental challenges. Therefore, managing water quality impacts in urban areas must be addressed to protect our environment. One of the maintenance steps is

Submission: 10 September 2023; **Acceptance:** 4 October 2023

installing gross pollutant traps (GPT). Gross pollutants are defined as discarded materials more significant than 5 mm, including dirt and debris, and coarse sediments are particles with a grain size greater than 0.5 mm (Allison, Chiew, & McMahon, 1997). Street waste that looks like litter or excrement (waste, trash, etc.) and dead organic matter in the form of pruned branches and leaves that can be used as fertilizer and organic matter (sediment, leaves, and grass clippings) are classified as dirty pollutant and can exhibit various levels. Physical and material properties include hardness, shape, size, and density (Madhani & Brown, 2015).

The main objective of GPT is to remove dirty pollutants that are carried into the rainwater system before the rainwater enters the main river channel (Fitzgerald & Bird, 2011). In Malaysia, GPT is suggested to be located at the end of each canal to trap dirty pollutants before they enter the primary river system (DID, 2012).

At the same time, it is crucial to understand that tropical climates are always associated with conditions where high rainfall intensity occurs in a short time, which causes the amount of waste that can be captured by GPT to be unpredictable (Mohd Sidek et al., 2014). It causes other problems, namely difficulties in predicting the amount of waste that must be transported and transferred from the GPT location to the final landfill. Often, the rubbish captured by the GPT will accumulate at the GPT location without being properly transported because the waste vehicles that must transport the waste are insufficient in quantity and capacity. For this reason, it is necessary to have a model that can accurately predict the amount of waste that might be captured by each GPT based on past data on the amount of garbage that has been caught. In addition, considering past data on the amount of rainfall in the GPT location can provide predictive results with a high accuracy value.

In their research, Zahari et al. (2016) did the data collection of the garbage and the pre-processing process to obtain the primary dataset to use in their research. Based on their study, we collected our dataset using a similar procedure. Additionally, in their research, Rahmawati et al. (2021), polynomial regression and Facebook Prophet models were used to predict the number of positive COVID-19 patients in Indonesia. This dataset was taken from 02/03/2020 to 31/03/2021. It can be concluded that the selection of parameters in each model is very influential in determining the level of accuracy, MAE, RMSE, and R_2. From the implementation results of each parameter, the best prediction is found in Polynomial Regression with degree 3 because the curve formed is not too far from the actual situation; it has an accuracy level of 98%, whereas with FBProphet, the accuracy level obtained is 95%.

The other researchers, like Lestari (2019), used similar algorithms: Regression Linear and Simple Additive Weighting (SAW). The advantages of using the SAW method include students with a high average report card score, namely with an average report card score of 79.121 in 2011 and 79.057 in 2012, with a difference in score of 0.426, which is greater than the average report card score using Regression- SAW. The number of schools accepted is also a difference of 10 to 20 more than the number of schools accepted compared to Regression-SAW. The school accreditation score received was also higher, namely a difference of 1.721. Execution time is slightly faster, namely a difference of 5.77.

On the other hand, Rodríguez del Águila & Benítez-Parejo (2011), in their research, their describe simple and multiple linear regression models, how they are calculated, and how their

applicability assumptions are checked. An illustrative example is provided based on freely accessible use. In biomedical research, it is common to encounter problems where we want to relate a response variable to one or more variables that can describe the behaviour of the previous variable through a mathematical model.

This paper proposes three regression methods, Simple Linear Regression, Multiple Linear Regression, and Polynomial Regression, like Rahmawati et al. (2021) and Putra et al. (2020), and the results will be compared based on Mean Absolute Error (MEA), Mean Square Error (MSE) and Root Mean Square Error (RMSE) measurements (Chicco, Warrens and Jurman, 2021). Based on the literature that has been studied, these three methods can give good results (Kim & Oh, 2021). For this reason, research needs to be conducted to determine and understand the problem better, which method will provide prediction results with the highest accuracy value. It is suitable as a prediction model for waste that GPT can capture based on the influence of rainfall data and the number of populations at the GPT location.

Methodology

Gross Pollutant Traps

Gross pollutant traps (GPT) remove trash, debris, and coarse sediment from stormwater. Some designs also provide oil separation. These substances are collectively referred to as Gross Pollutants.

Gross Pollutant Traps can be used as a pre-treatment to channel water into ponds or wetlands to limit areas of coarse sediment deposition. It facilitates the removal of finer sediments. Traps may also be used to keep coarse sediment out of the pond, protecting vegetation from damage from the effects of the residue. Traps can also remove coarse sediment before the flow enters the infiltration or filtration device, which would otherwise clog prematurely. GPTs can also serve the purpose of capturing floatable oil, provided they are designed appropriately. The trap offers little if any, flow attenuation.

Most GPTs will also provide some reduction in other pollutants. For example, a coarse sediment trap might also provide:

- Removal of particulate nutrients.
- Removal of metal traces.
- Removal of oil and grease.
- Reduction of bacteria.
- Reduction of substances that require dissolved oxygen.

All the above senses can be partially bound to the sediment and will be excreted along with the trapped sediment.

GPT is a complex tool and consists of very complete and interrelated components. GPT also has various types, depending on their respective purposes. One kind of GPT can be explained, as shown in Figure 1 (Racks, 2004).

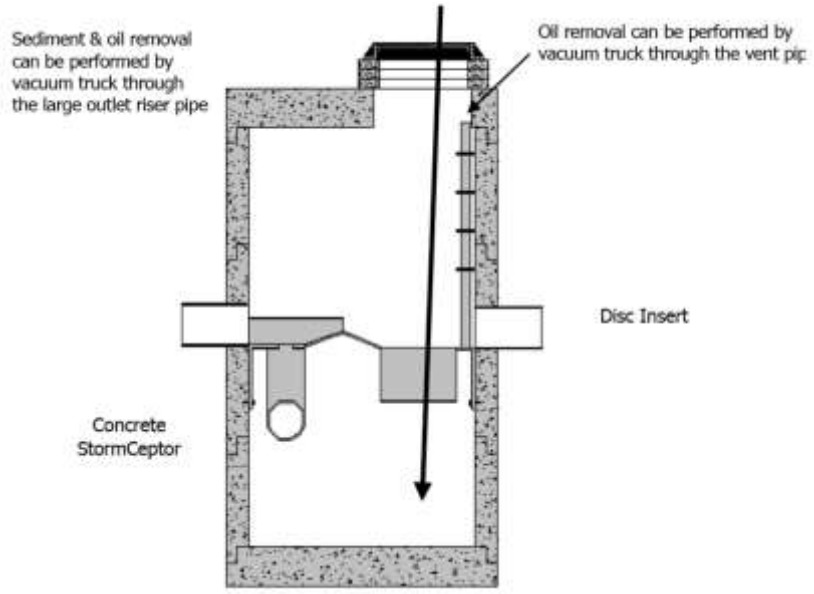


Figure 1. How GPT tools work in general (Racks, 2004).

In this research, there are 2 types of GPT used, namely Continuous Deflective System (CDS) and Cleans All (CA). For CDS, dirty pollutants and sediment are separated from rainwater by centrifugal force, preventing the screen from getting clogged with debris and other items, such as plastic bags. Meanwhile, for the CA type, underground dirty pollutant and sediment traps utilize filtered baskets to separate dirty pollutants, followed by a more bottomless tank for mud retention. Figure 2 and Figure 3 show how each GPT works (Zahari, Said, Sidek, & Basri, 2016).

Figure 2. *Continuous Deflective System* (CDS) (Zahari, Said, Sidek, & Basri, 2016).

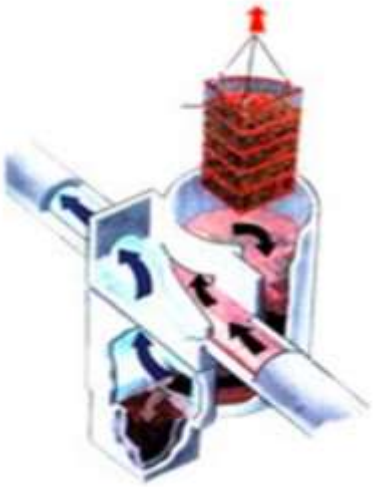


Figure 3. *Cleans All* (CA) (Zahari, Said, Sidek, & Basri, 2016).

Research Methodology

Research methodology stages are the process taken to research so that the research process can be structured well and systematically to achieve the expected goals. The following will explain the stages of the research methodology in Figure 4 below:

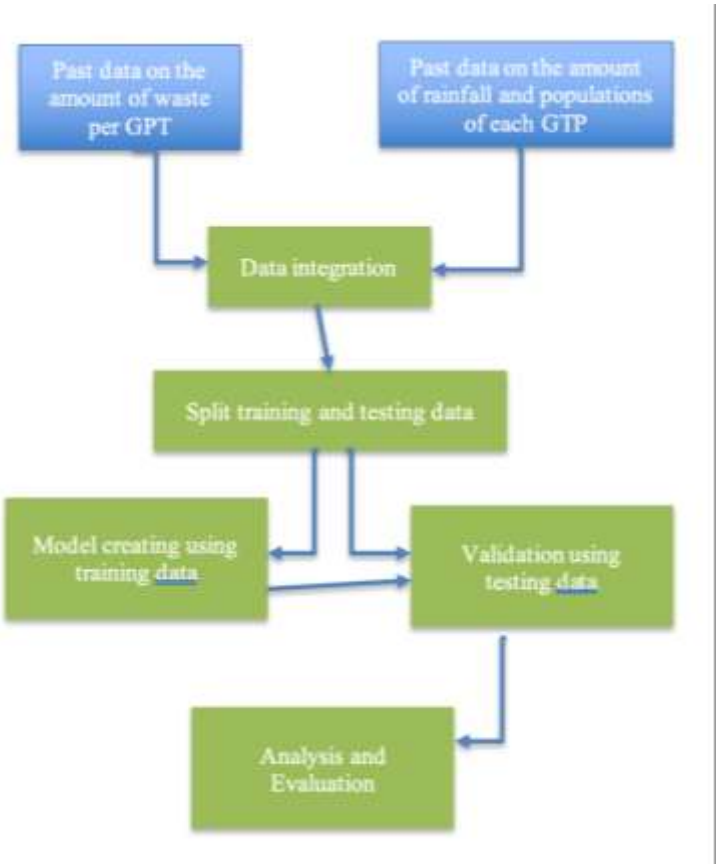


Figure 4. Research Stage

Figure 4 shows the process begins by combining two data sources into one, called an integration process. In this process, it is checked and ensured that the data in one section is available in others. For example, data on the amount of waste transported on one date must be guaranteed that data on the amount of rainfall and number of populations on that date is available.

Data Collections

The data collection stage in this research was not carried out indirectly but using data provided by Jabatan Saliran and Air, Putrajaya, Malaysia. The data consists of two types of data. The first is data on the amount of waste successfully trapped by each GPT. There are around 21 GPTs spread across several points in the Hulu Langat River, Sengalor. The current numbers consist of 4 years of data, from 2019 until 2022. Next, the data that will be used is data on the amount of rainfall and number of populations at the GPT location for 4 years on the same date as the data on the amount of GPT waste, as shown in Figure 5.

	GPT	Lat	Lng	2019-01-01 00:00:00	2019-01-02 00:00:00	2019-01-03 00:00:00	2019-01-04 00:00:00	2
2	GPT_001	3.187266	101.856873	102.0	107.0	112.0	127.0	
3	GPT_002	3.181436	101.85788000000000	79.0	96.0	79.0	65.0	
4	GPT_003	3.176741	101.85629400000000	166.0	179.0	195.0	212.0	
5	GPT_004	3.1732670000000000	101.85427	137.0	129.0	119.0	103.0	
6	GPT_005	3.1692550000000000	101.851955	177.0	197.0	205.0	191.0	
7	GPT_006	3.16578	101.84859200000000	38.0	20.0	11.0	30.0	
8	GPT_007	3.164966	101.853498	99.0	85.0	93.0	81.0	

Figure 5. Snapping of raw data of the amount of waste for each GPT.

Figure 5 shows each GPT data has its geolocation. Using this information, we can draw the correct location of the GPT on the map, as shown in Figure 6.

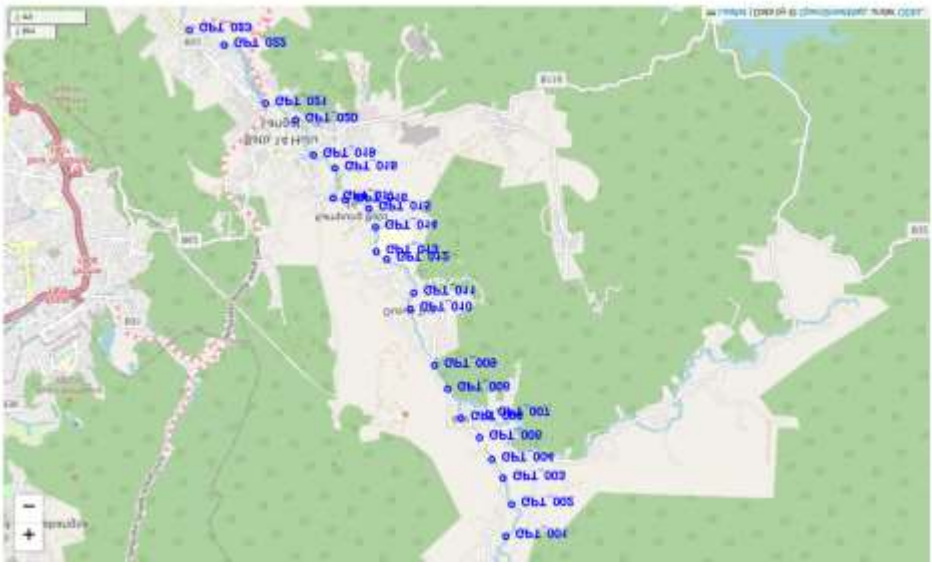


Figure 6. The location of GPT is on the map.

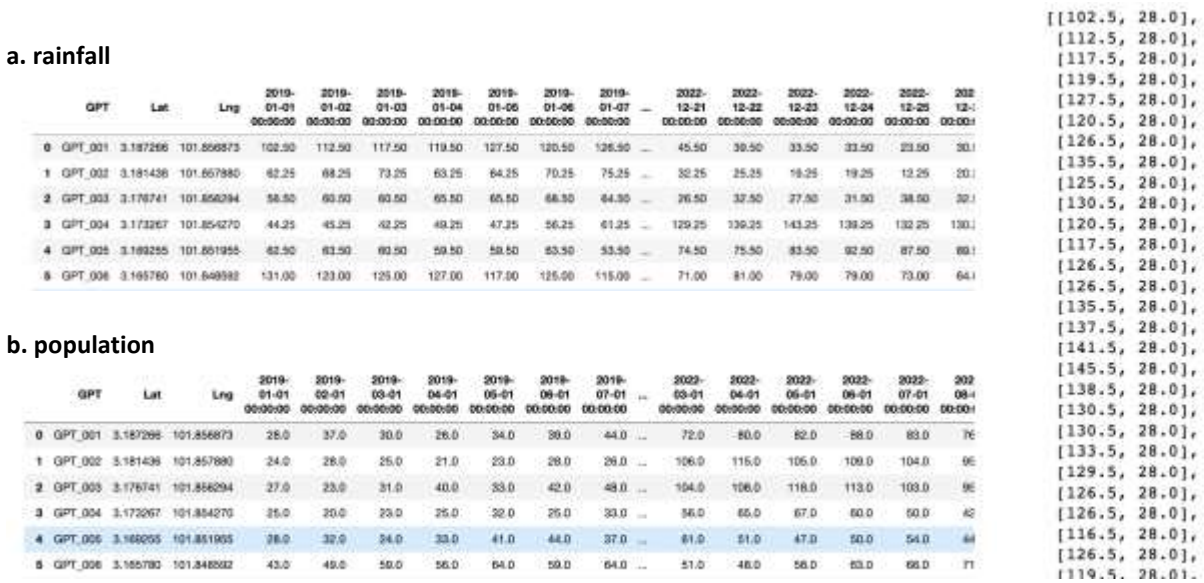
Pre-processing data

At the data pre-processing stage, there are several stages, namely:

1. Carry out data integration.
   This process is carried out to combine data into one data so that it can be appropriately processed.
2. Split the data into training data and testing data.
   This process is carried out to form a model from the learning results on training data and then validate and measure the level of accuracy of the model using testing data.
3. Forming a Model and Validating the Model
   The process of building a model using training data. After the model is formed, a model validation process is carried out by comparing the prediction results from the model to the testing data.
4. Analysis.
   This process is carried out to assess and analyze whether the model formation process has produced a model that is good enough to carry out the prediction process with a high accuracy value.

Data Integration

When dealing with Multiple Linear Regression and Polynomial Regression, the *x* variable consists of the rainfall and population data. On the other hand, the *y* variable is our amount of waste in kg. So, the *x* variable is formed from 2-dimensional data. We should combine carefully since our data have different headers. In rainfall data, we have daily data, but for population data, we have monthly data only. For this situation, we need to map each daily data from rainfall that will have the same value for one month. Figure 7 shows the original data and the result after the combined data.



a. rainfall

	GPT	Lat	Lng	2019-01-01 00:00:00	2019-01-02 00:00:00	2019-01-03 00:00:00	2019-01-04 00:00:00	2019-01-05 00:00:00	2019-01-06 00:00:00	2019-01-07 00:00:00	...	2022-12-21 00:00:00	2022-12-22 00:00:00	2022-12-23 00:00:00	2022-12-24 00:00:00	2022-12-25 00:00:00	202 12-
0	GPT_001	3.187266	101.856873	102.50	112.50	117.50	119.50	127.50	120.50	126.50	...	45.50	39.50	33.50	33.50	23.50	30.
1	GPT_002	3.181436	101.857880	62.25	68.25	73.25	63.25	64.25	70.25	75.25	...	32.25	25.25	19.25	19.25	12.25	20.
2	GPT_003	3.176741	101.856294	58.50	60.50	60.50	65.50	65.50	68.50	64.50	...	26.50	32.50	27.50	31.50	38.50	32.
3	GPT_004	3.172267	101.854270	44.25	45.25	42.25	49.25	47.25	56.25	61.25	...	129.25	139.25	143.25	139.25	132.25	130.
4	GPT_005	3.169255	101.851955	62.50	63.50	60.50	59.50	58.50	63.50	53.50	...	74.50	75.50	83.50	92.50	87.50	89.
5	GPT_006	3.165780	101.848592	131.00	123.00	125.00	127.00	117.00	125.00	115.00	...	71.00	81.00	79.00	79.00	73.00	64.

b. population

	GPT	Lat	Lng	2019-01-01 00:00:00	2019-02-01 00:00:00	2019-03-01 00:00:00	2019-04-01 00:00:00	2019-05-01 00:00:00	2019-06-01 00:00:00	2019-07-01 00:00:00	...	2022-03-01 00:00:00	2022-04-01 00:00:00	2022-05-01 00:00:00	2022-06-01 00:00:00	2022-07-01 00:00:00	202 08-
0	GPT_001	3.187266	101.856873	28.0	37.0	30.0	26.0	34.0	39.0	44.0	...	72.0	80.0	82.0	88.0	83.0	76
1	GPT_002	3.181436	101.857880	24.0	28.0	25.0	21.0	23.0	28.0	26.0	...	106.0	115.0	105.0	109.0	104.0	96
2	GPT_003	3.176741	101.856294	27.0	23.0	31.0	40.0	33.0	42.0	48.0	...	104.0	106.0	118.0	113.0	103.0	96
3	GPT_004	3.172267	101.854270	25.0	20.0	23.0	25.0	32.0	25.0	33.0	...	56.0	65.0	67.0	60.0	50.0	42
4	GPT_005	3.169255	101.851955	28.0	32.0	24.0	33.0	41.0	44.0	37.0	...	61.0	51.0	47.0	50.0	54.0	64
5	GPT_006	3.165780	101.848592	43.0	49.0	59.0	56.0	64.0	59.0	64.0	...	51.0	46.0	56.0	63.0	66.0	71

```
[[102.5, 28.0],
 [112.5, 28.0],
 [117.5, 28.0],
 [119.5, 28.0],
 [127.5, 28.0],
 [120.5, 28.0],
 [126.5, 28.0],
 [135.5, 28.0],
 [125.5, 28.0],
 [130.5, 28.0],
 [120.5, 28.0],
 [117.5, 28.0],
 [126.5, 28.0],
 [126.5, 28.0],
 [135.5, 28.0],
 [137.5, 28.0],
 [141.5, 28.0],
 [145.5, 28.0],
 [138.5, 28.0],
 [130.5, 28.0],
 [130.5, 28.0],
 [133.5, 28.0],
 [129.5, 28.0],
 [126.5, 28.0],
 [126.5, 28.0],
 [116.5, 28.0],
 [126.5, 28.0],
 [119.5, 28.0],
```

Figure 7. The original data and the result after the combined data

Figure 7 shows that each data GPT will pair between rainfall and population to become a 2-dimensional data *x* variable (right side of the picture).

Splitting Data

The dataset is split into 2 parts, first for training data and second for testing data. From 4 years of data, 3 years were used as training data and the last year as testing data, as shown in Figure 8.

```
_columns = sampah.shape[1] -3
_year = 365
_limit = _columns - (_year * 1)

# x = curah hujan
x_train = np.array(curah_hujan.loc[_row][3:_limit].to_list()).reshape((-1, 1))
x_test = np.array(curah_hujan.loc[_row][_limit:].to_list()).reshape((-1, 1))
# y = sampah
y = np.array(sampah.loc[_row][3:].to_list())
y_train = np.array(sampah.loc[_row][3:_limit].to_list())
y_test = np.array(sampah.loc[_row][_limit:].to_list())
```

Figure 8. The splitting data

Figure 8 shows that using the _limit variable, we calculate a one-year data split from the original data as testing data for the *x* and *y* variables become *x_train* and *x_test* also *y_train* and *y_test* variables.

The following process is forming the model and validation. The last technique is to do the analysis. Two final approaches will be presented in the next chapter.

Results and Discussion

The model is created based on three algorithms: Simple Linear Regression, Multiple Linear Regression, and Polynomial Regression. In the following the experiment details will be shown as follows.

In single linear regression, the *x* variable only uses data from rainfall, and the *y* variable uses the amount of waste in kg. For multiple linear and polynomial regression, the *x* variable consists of rainfall and population data, and then the pairing data was created. After that, based on a splitting scenario, as previously explained, the model was created. Based on the developed model, we calculated some parameters to measure the performance of our model.

For each GPT data, we created the model based on training data, and the measurement results for testing data were collected, as shown in Table 1 and Figure 9.

Table 1. The measurement results for each GPT data.

		Single Linear Regression			Multiple Linear Regression			Polynomial Regression		
No	GPT	MEA	MSE	RMSE	MEA	MSE	RMSE	MEA	MSE	RMSE
1	GPT_001	61.90	5,582.31	7.87	34.01	1,884.15	5.83	40.17	2,555.78	6.34
2	GPT_002	51.91	3,928.63	7.21	47.69	2,968.61	6.91	48.29	3,131.14	6.95
3	GPT_003	35.41	1,849.91	5.95	38.34	1,981.18	6.19	25.12	984.11	5.01
4	GPT_004	48.77	3,225.62	6.98	39.86	2,252.70	6.31	69.93	8,674.21	8.36
5	GPT_005	37.26	1,920.30	6.10	34.19	1,674.40	5.85	39.61	2,255.53	6.29

6	GPT_006	30.54	1,407.43	5.53	30.74	1,325.13	5.54	39.31	2,172.42	6.27
7	GPT_007	24.64	901.65	4.96	23.42	797.35	4.84	49.71	3,755.57	7.05
8	GPT_008	52.25	3,782.16	7.23	41.04	2,496.96	6.41	45.61	3,596.09	6.75
9	GPT_009	39.25	2,134.98	6.27	33.99	1,697.85	5.83	37.82	2,088.57	6.15
10	GPT_010	35.99	1,926.55	6.00	39.43	2,272.79	6.28	45.62	3,091.94	6.75
11	GPT_011	40.53	2,287.73	6.37	37.70	1,932.79	6.14	40.28	2,209.80	6.35
12	GPT_012	37.01	1,820.85	6.08	34.34	1,562.89	5.86	34.35	1,655.01	5.86
13	GPT_013	54.71	5,265.03	7.40	39.14	2,347.69	6.26	100.89	15,044.64	10.04
14	GPT_014	48.79	3,530.69	6.99	42.45	2,737.96	6.52	49.16	3,706.95	7.01
15	GPT_015	40.28	2,255.48	6.35	39.35	2,265.07	6.27	45.18	2,981.21	6.72
16	GPT_016	47.66	2,848.22	6.90	44.11	2,552.70	6.64	45.82	2,967.97	6.77
17	GPT_017	38.50	2,190.97	6.20	38.55	2,202.82	6.21	46.38	3,408.15	6.81
18	GPT_018	47.73	3,162.52	6.91	47.04	3,017.53	6.86	53.62	3,949.74	7.32
19	GPT_019	46.73	3,275.58	6.84	42.27	2,613.74	6.50	69.30	8,085.17	8.32
20	GPT_020	29.99	1,450.96	5.48	32.98	1,604.06	5.74	104.52	19,472.94	10.22
21	GPT_021	34.03	1,743.17	5.83	34.21	1,877.77	5.85	46.29	3,804.05	6.80
22	GPT_022	42.60	2,443.72	6.53	40.76	2,263.87	6.38	53.61	3,820.78	7.32
23	GPT_023	45.03	2,968.54	6.71	44.17	2,589.56	6.65	42.68	2,490.25	6.53
	Average	**42.24**	**2,691.43**	**6.46**	**38.25**	**2,126.94**	**6.17**	**51.01**	**4,604.44**	**7.04**

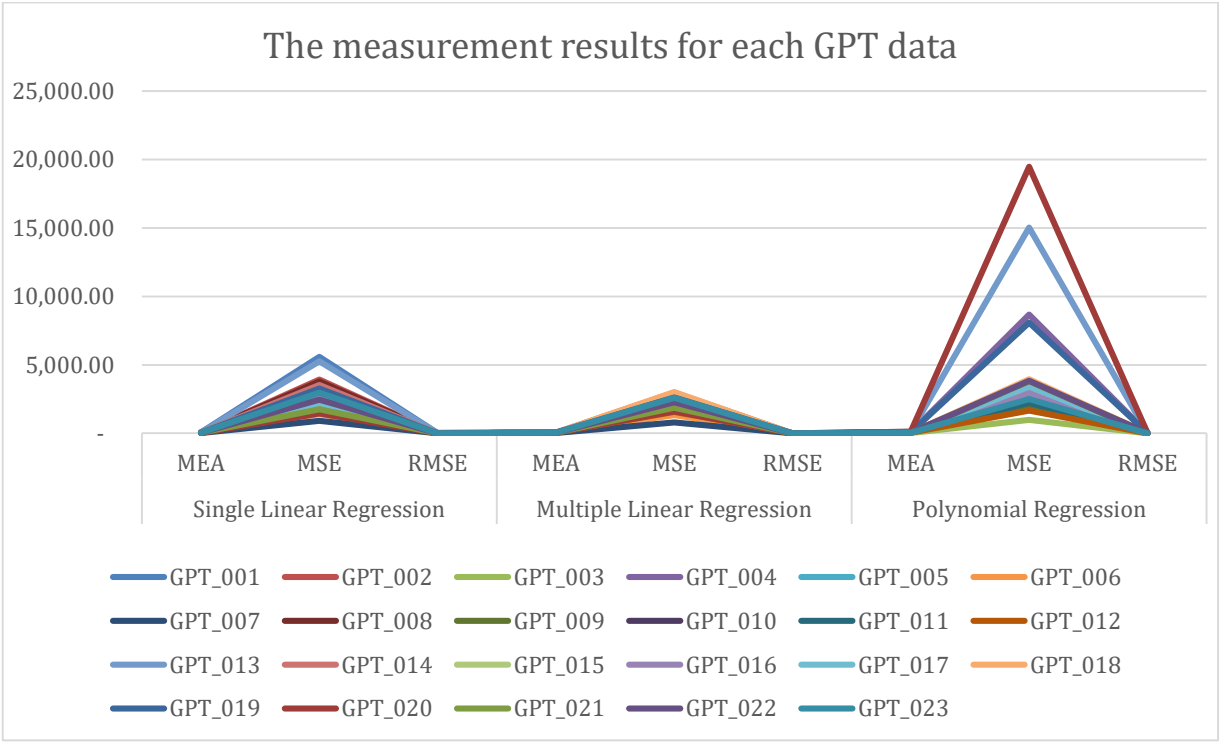
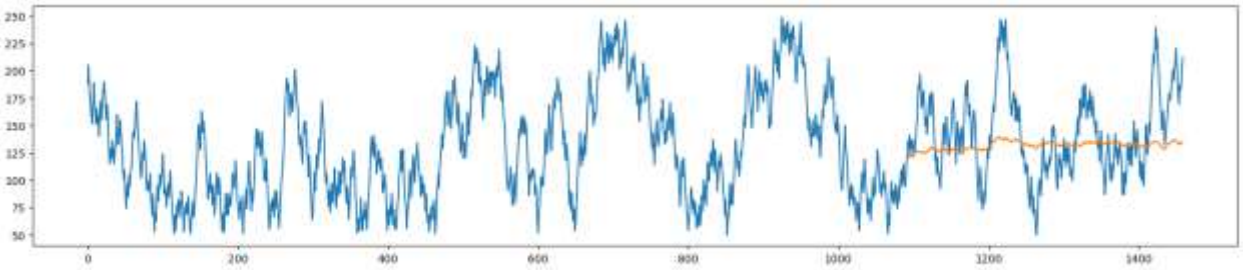


Figure 9. The measurement results for each GPT data

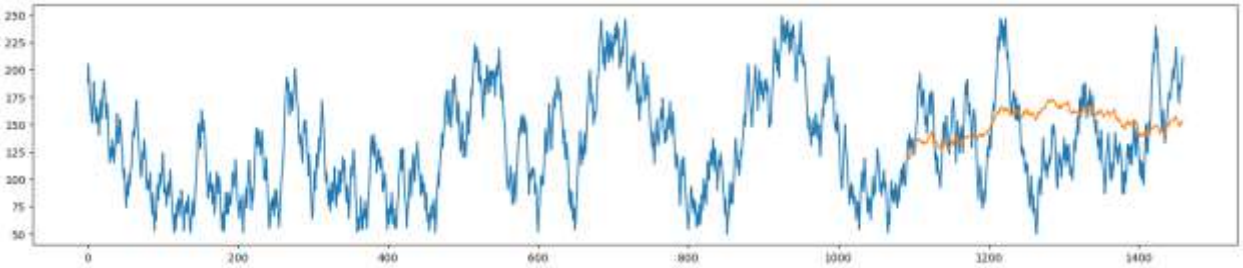
Based on the result shown in Table 1 and Figure 9, the MSE from Polynomial Regression is too high compared with Single and Multiple Linear Regression. For MEA and RMSE measurement, the three algorithms give the results quite similar. The Multiple Linear Regression gives more minor results for all GPTs than the three algorithms. It means that Multiple Linear Regression created the best model to predict the amount of waste. Based on the average values in Table 1, for all measurements on each algorithm, the best model was created by Multiple Linear Regression, followed by Single Linear Regression, and the last one is Polynomial Regression.

Further, we focus on the higher result for Polynomial Regression and compare it with other algorithms for the same GPT data. Based on Table 1, the higher RME result is GPT_020. So, for analysis, the visualization of training data and testing data, including the prediction data, was visualized in Figure 10 and Figure 11. On the other hand, the lower RME result is GPT_007. The visualization of training and testing data, including the prediction data, was visualized in Figure 12 and Figure 13.

a. single_linear_regression



b. multiple_linear_regression



c. polynominal_regression

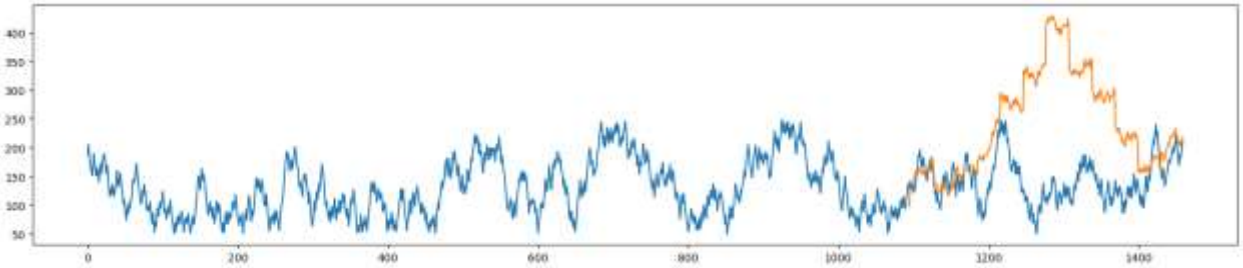
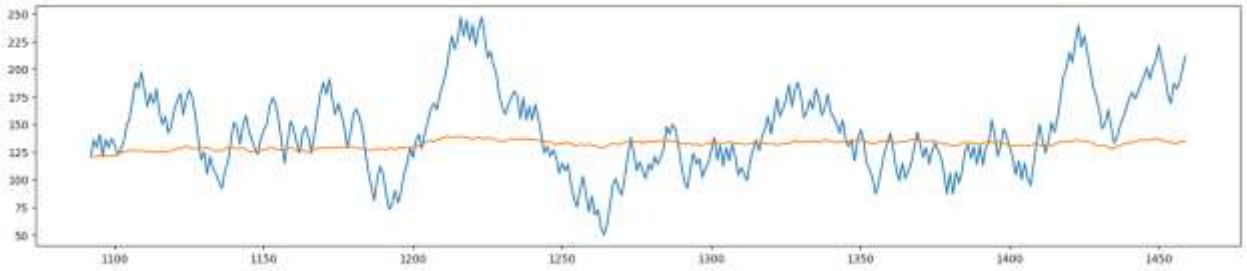
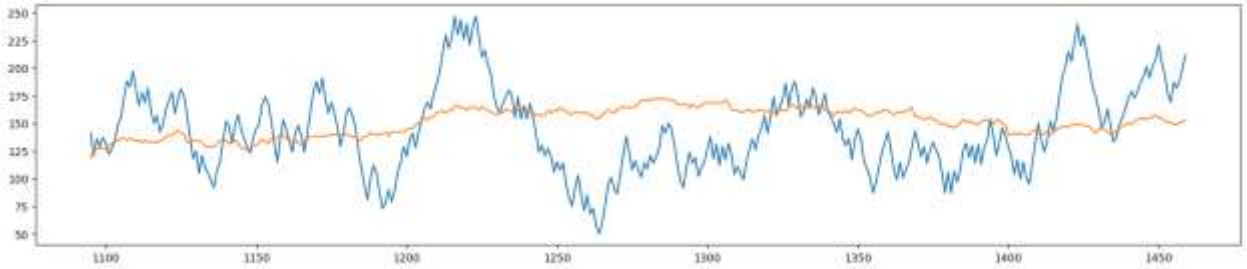


Figure 10. Visualize the training and testing data, including the prediction results for GTP_020.

a. single_linear_regression



b. multiple_linear_regression



c. polynominal_regression

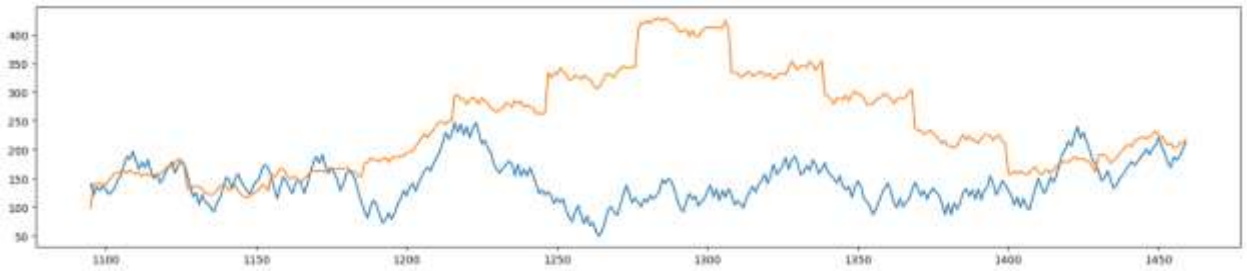
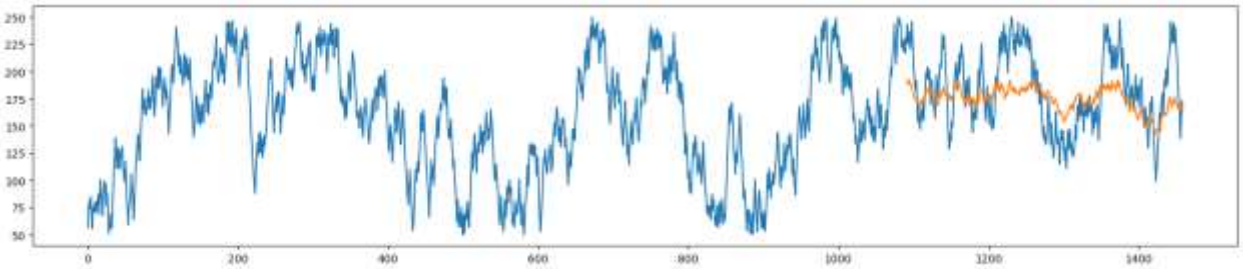


Figure 11. Visualize comparison only the testing data and prediction results for GPT_020.

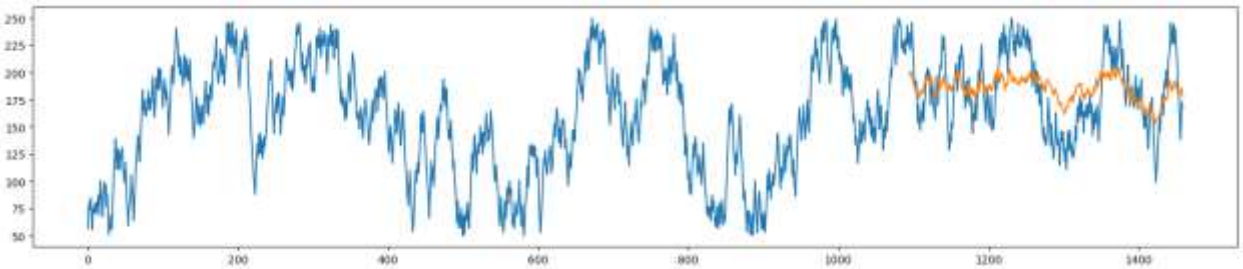
Based on Figure 10 and Figure 11, the pattern of the *y* variable for testing compared with the prediction result is quite different, especially for the Polynomial Regression result (c, part). The model cannot predict well. Returning to Table 1, the outcome for GTP_020 differs significantly compared to other GPT results. The MSE result from the Polynomial algorithm gives the worst result, 19,472.94, compared with Multiple Linear Regression and Single Linear Regression at 1,604.06 and 1,450.96. The higher MSE means the prediction result is too different from the original values. Our model fails to detect the pattern from the training dataset.

From Figure 11, for parts a and b, Single Linear Regression and Multi Linear Regression, the graph shows the prediction value lines are too flat. Compared with the actual values, the line is up and down significantly. The created models are generalizing the pattern, so the prediction result line is flat.

a. single_linear_regression



b. multiple_linear_regression



c. polynominal_regression

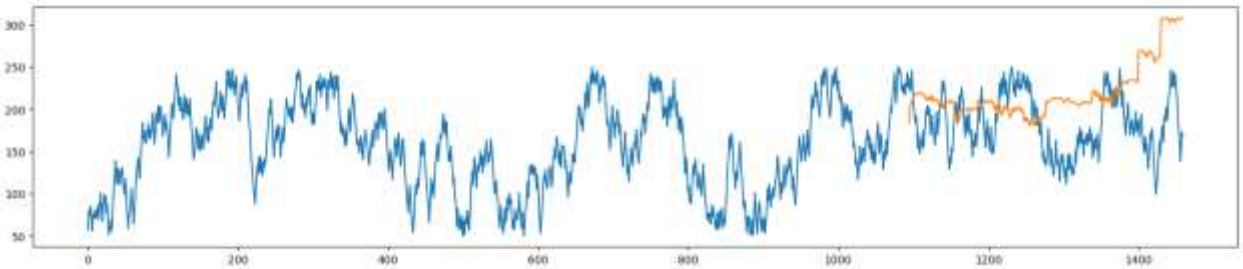
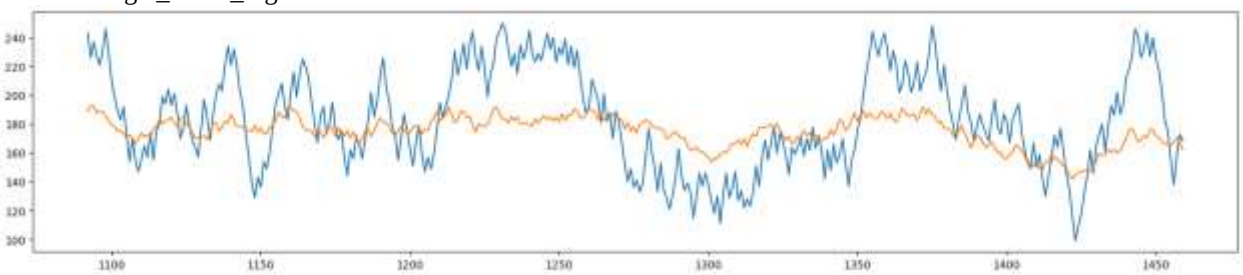
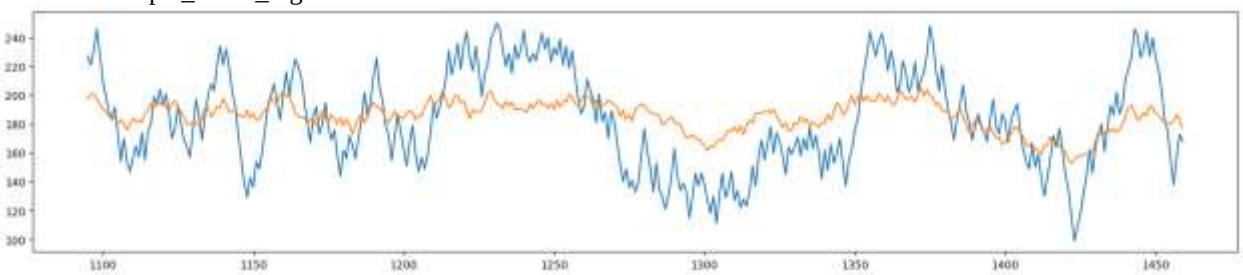


Figure 12. Visualize the training and testing data, including the prediction results for GTP_007.

a. single_linear_regression



b. multiple_linear_regression



c. polynominal_regression

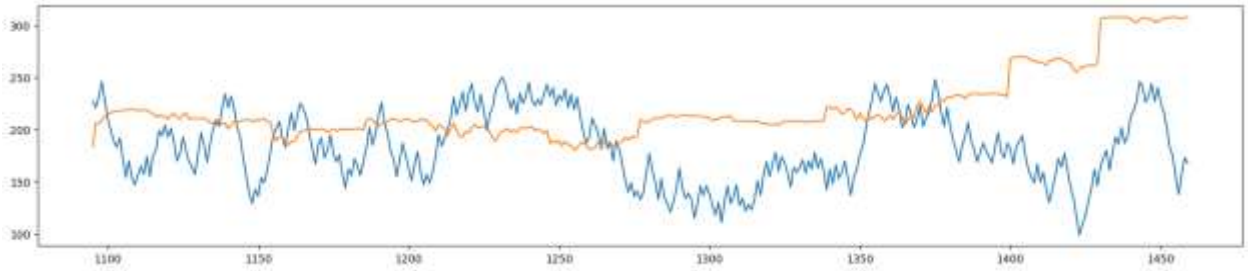


Figure 13. Visualize comparison only the testing data and prediction results for GPT_007.

From Figure 12 and Figure 13, the prediction results show the line slightly follows the actual values, but the patterns are still quite different. In Table 1, the Multiple Linear Regression gives the best outcome for MSE measurement, at 797.35, compared with Single Linear Regression and Polynomial Regression at 901.65 and 3,755.57. The smallest value for Multiple Linear Regression means the predictions are pretty similar to actual values. It means our model successfully captures the training dataset's pattern and predicts the testing dataset.

Conclusion

In this paper, the waste prediction from the GPT dataset is collected, and then the pre-processing dataset is done. The needed procedure combines and integrates data before splitting data and creating modeling. The data was divided between 75% and 25% for training and testing. The experiments were conducted, and the results were collected and reported. The results show that the Multiple Linear Regression gives the best accuracy, with 797.35 MSE values for GPT_020. The worst outcome is Polynomial Regression, in which the MSE value is 19,472.94 for GPT_007. It means the model created for GPT_020 successfully captures the training data pattern. Conversely, the Polynomial Regression fails to capture the pattern carefully, and the prediction result is too generalized, so the prediction shows a flat line.

References

Allison, R., Chiew, F., & McMahon, T. (1997). *Stormwater Gross Pollutants by*. 26. Retrieved from http://www.clearwater.asn.au/sites/clearwater.asn.au/files/resources/CRC

Chicco, D., Warrens, M.J., and Jurman, G., (2021). *The coefficient of determination R-squared is more informative than SMEPA, MEA, MAPE, MSE and RMSE in regression analysis evaluation,* PeerJ Computer Science. Vol.7. No. 3.

DID. (2012). Urban Stormwater Management Manual for Malaysia (MSMA 2nd Edition). In *CME 2007 Conference - Construction Management and Economics: "Past, Present and Future"* (pp. 373–383). Department of Irrigation and Drainage (DID), Malaysia.

Fitzgerald, B., & Bird, W. S. (2011). Literature Review: Gross Pollutant Traps as a Stormwater

Management Practice. In *Auckland Council Technical Report* (Vol. 4525).

Kim, Y., & Oh, H. (2021). Comparison between multiple regression analysis, polynomial regression analysis, and an artificial neural network for tensile strength prediction of BFRP and GFRP. *Materials*, *14*(17), 1–13.

Lestari, S. A. (2019). Lestari, S. A. (2019). *Analisis perbandingan metode regresi linier - simple additive weighting (saw) dengan metode simple additive weighting (saw) untuk seleksi mahasiswa baru jalur undangan.* Degree Thesis, Universitas Sebelas Maret, Solo, Indonesia.

Mohd Sidek, L., Basri, H., Md Said, N. F., Mohd Jayothisa, W., Mohd. Sabri, A. F., & Md Noh, M. N. (2014). a Study on Effectiveness and Performance of Gross Pollutant Traps for Stormwater Quality Control for River of Life (Rol) Project. *13th International Conference on Urban Drainage,* (September 2014), 7–12.

Madhani, J., & Brown, R. (2015). The capture and retention evaluation of a stormwater gross pollutant trap design. *Ecological Engineering*, *74*.

Putra, B.J., Kurniawan, T.B., Antoni, D., Mirza, A.H., (2020), *Prediksi Kebutuhan Alat Kesehatan Rumah Sakit Menggunakan Metode Algoritma Regression Linear dan Naïve Bayes.* Jurnal Informatika Global, Vol. 11(2).

Racks, T. (2004). Urban stormwater management – III. *Environmental Hydraulics and Sustainable Water Management, Two Volume Set*, 1501–1501. https://doi.org/10.1201/b16814-245

Rahmawati, T. A., Marthasari, G. T., & Hayatin, N. (2021). *Perbandingan Model Polynomial Regression dan Facebook Prophet untuk Prediksi Jumlah Pasien Positive COVID-19 di Indonesia*. *3*(5), 471–482.

Rodríguez del Águila, M. M., & Benítez-Parejo, N. (2011). Simple linear and multivariate regression models. *Allergologia et Immunopathologia*, *39*(3), 159–173.

Zahari, N. M., Said, F., Sidek, L. M., & Basri, H. (2016). *Gross Pollutant Traps : Wet Load Assessment at Sungai Kerayong , Malaysia*.