energies

Article

Polynomial Regression Model Utilization to Determine Potential Refuse-Derived Fuel (RDF) Calories in Indonesia

Luqman Luqman [1,*], Sarifuddin Madenda [2] and Prihandoko Prihandoko [2]

[1] Faculty of Energy Telematics, Institut Teknologi PLN, Jakarta 11750, Indonesia
[2] Department of Information Technology, Gunadarma University, Depok 16424, Indonesia; sarif@staff.gunadarma.ac.id (S.M.); pri@staff.gunadarma.ac.id (P.P.)
* Correspondence: luqman@itpln.ac.id

Abstract: Waste-to-energy (WTE) is one of the Indonesian government's programs aiming to meet the target of achieving a new and renewable energy (NRE) mix, as well as one of the solutions proposed to overcome the problem of waste. One of the products of WTE is energy derived from raw material waste (refuse-derived fuel/RDF). Using the formula $y = 0.00003\ x^5 - 0.0069\ x^4 + 0.6298\ x^3 - 24.3245\ x^2 + 432.8401\ x + 55.7448$ with $R^2 = 0.9963$, which was obtained by comparing a scatter plot diagram from the RDF calorie test dataset produced through a bio-drying process, the potential RDF calories produced using the waste composition dataset taken from each region in Indonesia can be calculated. The results of the calculations using the determined equations produce a list of provinces with RDF calorie potential, ordered from the largest to the smallest, using which the government can determine which areas are the main priority for processing waste into energy. Thus, through this method, the target of 5.1% renewable energy sourced from waste can be achieved by 2025.

Keywords: polynomial regression; prediction model; renewable energy; refuse-derived fuel; waste to energy

check for updates

Citation: Luqman, L.; Madenda, S.; Prihandoko, P. Polynomial Regression Model Utilization to Determine Potential Refuse-Derived Fuel (RDF) Calories in Indonesia. *Energies* **2023**, *16*, 7200. https://doi.org/10.3390/en16207200

Academic Editors: Cheol-Hong Hwang and Albert Ratner

Received: 9 June 2023
Revised: 12 October 2023
Accepted: 17 October 2023
Published: 22 October 2023

1. Introduction

The utilization of waste as energy is one of the Indonesian government's current programs for obtaining alternative energy, in addition to other forms of energy such as geothermal, hydro, mini/micro-hydro, solar, wind, and ocean waves. Based on the 2021–2030 National Electricity Supply Business Plan (NESBP) [1], the addition of power plants in Indonesia over a 10-year period is planned in order to add a total of 56,024 gigawatts (GW) of capacity. The addition of new power plants comprises an energy mix including coal (54.4%), new and renewable energy (NRE, 23%), gas (22.2%), and fuel (0.4%) to be achieved by the end of 2025. To meet the NRE target, which has a share of 23%, it is necessary to make comprehensive efforts starting from the energy sources, supply chains, and an efficient model, so that NRE can meet the targets outlined in the NESBP. Information technology has a very important role to play in achieving the NRE targets set by the government, especially in processing the data generated during the energy supply chain process so that an efficient and optimal model can be obtained. NRE is a mix of various types of renewable energy, including geothermal, hydro, mini- and micro-hydro, bioenergy/waste-to-energy (WTE) [2], solar, wind, and ocean waves, all with potential in Indonesia [1], as shown in Table 1 below.

The whole world faces the problem of waste, be it in the form of urban waste or various other types of waste [3]. Even agricultural and plantation residues often cause disturbance, not only for the environment but also socially, economically, and politically. Therefore, researchers are starting to look at the use of waste for many needs, especially energy and materials [4–6]. Hence, renewable energy sources related to converting organic waste materials to energy are becoming important attractive substitutes for the near future [7]. The raw materials of waste processed into energy (refuse-derived fuel/RDF) will produce different amounts of calories according to the composition of the waste being processed. Determination of the RDF calorie yield was carried out in laboratory calorie tests with various compositions to determine the optimal composition for producing the most calories from waste. Based on the data from the calorie test results, several forecasting model equations were obtained, and this model was used for calculating RDF calorie potency throughout Indonesia.

Table 1. NRE mix potentials.

No	Type of Energy	Potentials	Installed Capacity	Utilization
1	Geothermal	29,544 MW	1438.5 MW	4.9%
2	Hydro	75,091 MW	4826.7 MW	6.4%
3	Mini- and Micro-Hydro	19,385 MW	197.4 MW	1.0%
4	Bioenergy/Waste-to-Energy (WTE)	32,654 MW	1671.0 MW	5.1%
5	Solar	207,898 MW (4.80 kWh/m^2/day)	78.5 MW	0.04%
6	Wind	60,647 MW ($\geq$4 m/s)	3.1 MW	0.01%
7	Ocean Waves	17,989 MW	0.3 MW	0.002%

To manage waste into energy, properly processed data on the composition of waste in every city or province in Indonesia are needed. Waste composition data obtained from all locations in Indonesia were tested against a casting model from the calorie test data (reference data). The results of the validation test for errors from the forecasting model obtained were used to describe the calorie potential of waste throughout Indonesia, grouped based on the calorie potential of waste in each province.

The results of the depiction of the calorie potential of waste to become energy raw materials (refuse-derived fuel/RDF) can be used by the government to create policies determining which locations have the most potential to utilize RDF in waste management. The resulting RDF can be utilized for the co-firing process [8], i.e., mixing RDF with coal for combustion in a steam power plant boiler. This will reduce the use of coal, supporting the net zero waste by 2050 program.

2. Materials and Methods

2.1. Proposed Methods

The proposed methods are illustrated by the research scheme in Figure 1, which in general consists of two parts. The first part (the top of Figure 1) is the development of a mathematical model based on the ground truth dataset. This section begins with the process of creating a ground truth RDF dataset, and then a scatter plot diagram is generated to analyze its characteristics. Referring to these characteristics, the mathematical approach that can be used to build the model was determined. Furthermore, the developed model was applied in the second part (the bottom of Figure 1) to calculate the potential RDF calories from waste in 34 provinces in Indonesia. The first part of the proposed method in Figure 1 will be discussed more thoroughly in Section 2.2, and the second part will be elaborated upon further in Section 2.3.

2.2. RDF Calorie Potential Forecasting Model Development

The main objective of this section is to build a mathematical equation model that can be used to calculate the potential calories in waste as a renewable energy source. To realize this goal, a ground truth RDF dataset must be created, which will be used as a reference target in the model creation process. This model will become a new standard in waste management and calculating the potential calories that can be produced from every city in Indonesia.

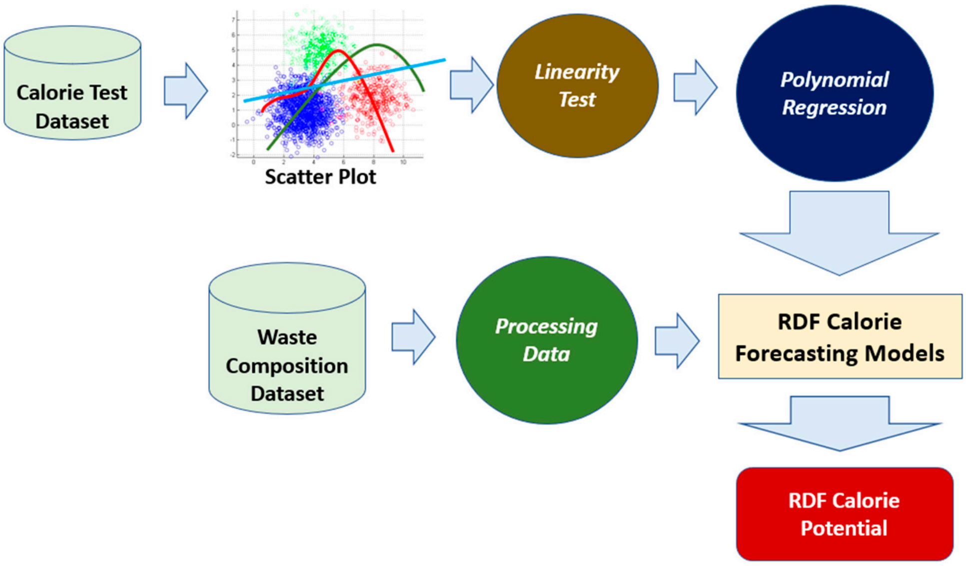

Figure 1. Research scheme.

2.2.1. Ground Truth Dataset (Calorie Test Dataset)

The aim of creating a ground truth RDF dataset is to find out how much caloric value is in one kilo gram of RDF, based on the percentage of organic and non-organic material contained in it (excluding glass and iron). The process of forming this dataset takes the longest time in the research stage. Starting from waste collection, a "peuyemization" or bio-drying process is carried out to remove the water content, followed by the process of determining variations in organic and non-organic composition in every one kg of the RDF sample. A total of 90 RDF samples were made with varying percentages of organic and non-organic composition ranging from 100%:0%, 99%:1%, and 98%:2% to 10%:90%. The final process involves calculating the caloric value contained in the sample through laboratory tests using a bomb calorimeter. Laboratory tests were carried out with a multidisciplinary research team at the ITPLN Waste-to-Energy Research Center, and the results were validated by the Indonesia Power Suralaya Laboratory, which is certified by the National Accreditation Committee for Testing Laboratories.

The organic and non-organic percentage compositions of all RDF samples are shown in Table 2 column "X". Meanwhile, in column "Y", the laboratory test results of the caloric value of each RDF sample are presented. For example, in the first row, the RDF sample with a composition of 100% organic and 0% non-organic has 3008 kcal/kg RDF. Furthermore, the RDF sample in the second row with a composition of 99% organic and 1% non-organic contains calories of 3114 kcal/kg, and this continues until the last row, which is the RDF sample, which has a 10% organic and 90% non-organic composition, and can produce calories of 2631 kcal/kg.

Table 2. Ground truth RDF dataset from the laboratory test results.

Organic (%)	Non Organic (%)	Calories (kcal/kg)	Organic (%)	Non Organic (%)	Calories (kcal/kg)	Organic (%)	Non Organic (%)	Calories (kcal/kg)
X		Y	X		Y	X		Y
100	0	3008	70	30	6138	40	60	3653
99	1	3114	69	31	6096	39	61	3619
98	2	3221	68	32	6054	38	62	3585
97	3	3327	67	33	6012	37	63	3551
96	4	3434	66	34	5970	36	64	3517

Table 2. *Cont.*

Organic (%)	Non Organic (%)	Calories (kcal/kg)	Organic (%)	Non Organic (%)	Calories (kcal/kg)	Organic (%)	Non Organic (%)	Calories (kcal/kg)
X		**Y**	**X**		**Y**	**X**		**Y**
95	5	3540	65	35	5929	35	65	3483
94	6	3646	64	36	5887	34	66	3449
93	7	3753	63	37	5845	33	67	3415
92	8	3859	62	38	5803	32	68	3380
91	9	3966	61	39	5761	31	69	3346
90	10	4072	60	40	5719	30	70	3312
89	11	4193	59	41	5633	29	71	3278
88	12	4314	58	42	5547	28	72	3244
87	13	4434	57	43	5462	27	73	3210
86	14	4555	56	44	5376	26	74	3176
85	15	4676	55	45	5290	25	75	3142
84	16	4797	54	46	5204	24	76	3108
83	17	4918	53	47	5118	23	77	3074
82	18	5038	52	48	5033	22	78	3040
81	19	5159	51	49	4947	21	79	3006
80	20	5280	50	50	4861	20	80	2972
79	21	5366	49	51	4740	19	81	2938
78	22	5452	48	52	4619	18	82	2904
77	23	5537	47	53	4499	17	83	2869
76	24	5623	46	54	4378	16	84	2835
75	25	5709	45	55	4257	15	85	2801
74	26	5795	44	56	4136	14	86	2767
73	27	5881	43	57	4015	13	87	2733
72	28	5966	42	58	3895	12	88	2699
71	29	6052	41	59	3774	11	89	2665

In the table, it can be seen that organic waste contains much more carbon as a source of calories than non-organic waste. Thus, in this research, the variable used to create a waste calorie prediction model is the organic percentage value. Looking at the data in Table 2, the percentage of organic vs. calories in one kg of RDF can be visualized using the bar diagram in Figure 2 and the scatter plot diagram in Figure 3. In both figures, the x-axis is the percentage of organics in RDF, and the calories produced per kg RDF are on the y-axis.

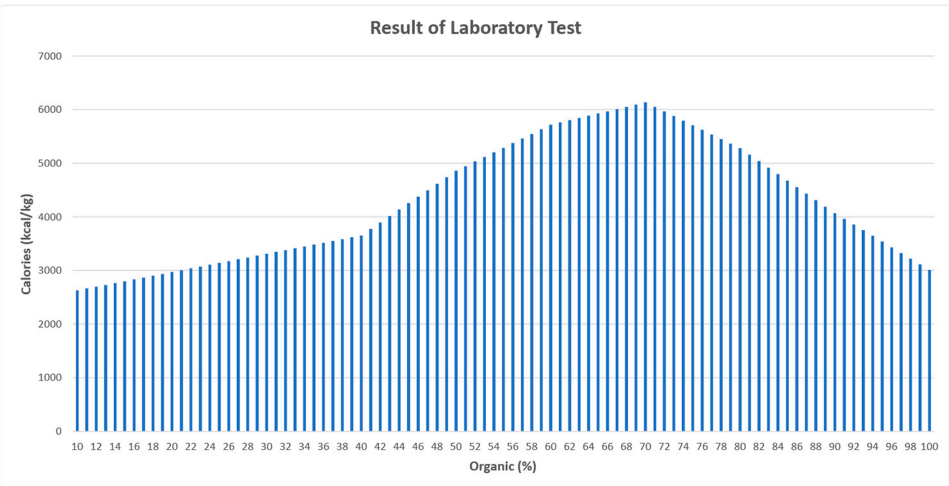

Figure 2. Results of the laboratory test.

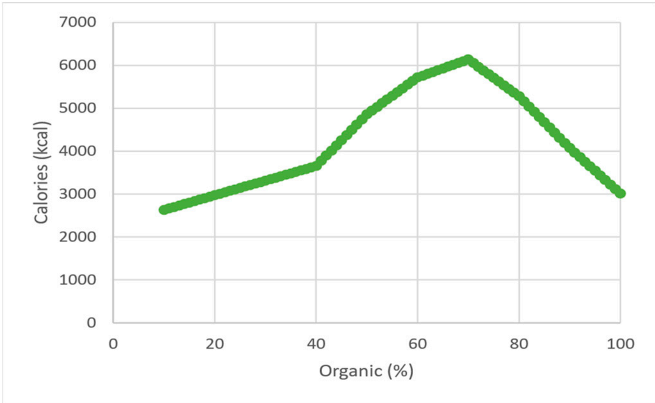

Figure 3. Scatter plot diagram.

Before determining which mathematical model can be applied, it is necessary to analyze the curve shape of the dataset used. Referring to the scatter plot diagram in Figure 3, in general, the shape of the curve can be divided into three parts. The first curve is in the range of 10% to 40%; the calories produced increase linearly as the organic percentage value in the RDF increases. The second curve is between 41% to 70%; the caloric value increases quadratically, following the increase in the organic percentage in the RDF. The last curve is in the range of 71–100%; the caloric value decreases linearly as the organic percentage in the RDF increases. The peak of the curve with the highest caloric value is RDF, which has an organic percentage of 70%, or RDF with a combination of organic and non-organic content of 70%:30%.

2.2.2. Regression Analysis and Modeling

Regression analysis is an analysis of the dependence of one or more independent variable on one dependent variable, with the aim of estimating or predicting the average value of the population based on the values of the independent variables [9]. A regression analysis used to predict one dependent variable based on one independent variable is called a simple regression analysis, while a regression analysis used to predict one dependent variable based on one or more independent variables is called a multiple regression analysis. In addition, regression can also be used to measure the strength of the relationship between two or more variables. Regression analysis is also used to show the direction of the relationship between the independent variable and the dependent variable [10]. Regression analysis is essentially divided into two forms, namely linear regression analysis and non-linear regression analysis.

Non-linear regression is a method of regression analysis used to obtain a non-linear model that is used to determine the relationship between the dependent variable and the independent variable. Non-linear models (i.e., non-linear in the parameters to be estimated) can be divided into two parts, namely intrinsic linear models and intrinsic non-linear models. If a model is intrinsically linear, then this model can be expressed through appropriate transformations of the variables into standard linear forms, such as exponential regression. Yet, if a model is intrinsically non-linear, then this model cannot be converted into a standard form. If the relationship between the dependent variable Y and the independent variable X is non-linear, this means that if the original data X_i and Y_i are made into a scatterplot, it does not follow a straight line, but follows a specifically shaped curve, such as an exponential curve. Thus, regression analysis, which is suitable for explaining the relationship between X and Y, is a simple form of non-linear regression analysis [11]. If the linear form is accepted, then followed by the fact that the regression is a unit, it is ensured that the regression coefficient obtained cannot be ignored; then, conclusions can be made based on the regression.

According to the curve shape of the ground truth RDF dataset in Figure 3 and the regression analysis, the better mathematical approach to use is the *k* order polynomial model in one variable [10,12] as given by Equation (1). To form a mathematical equation model that fits the curve of the ground truth RDF dataset (in Table 2), mathematical modeling and simulations were carried out using polynomial regression equations: first order, second order, third order, fourth order, and fifth order. The resulting regression models are shown in Equations (2)–(6):

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \ldots + \beta_k x^k + \varepsilon \tag{1}$$

$$\text{First order: } y = 21.974x + 3089 \tag{2}$$

$$\text{Second order: } y = -1.28x^2 + 162.78x + 100.18 \tag{3}$$

$$\text{Third order: } y = -0.0342x^3 + 4.3626x^2 - 105.1x + 3454.2 \tag{4}$$

$$\text{Fourth order: } y = 0.0003x^4 - 0.0983x^3 + 9.1343x^2 - 242.17x + 4664.1 \tag{5}$$

$$\text{Fifth order: } y = 0.00003x^5 - 0.0069x^4 + 0.6298x^3 - 24.325x^2 + 432.84x + 55.745 \tag{6}$$

Furthermore, each of these equations is reapplied using the dataset in Table 2, and the calorie calculation results are displayed by orange, cyan, purple, black, and red curves in Figure 4. The orange curve represents a first-order polynomial equation with an R2 value of 0.2676. The cyan curve reflects a second-order polynomial equation with an R^2 value of 0.7687. The third-order polynomial equation curve is colored purple with an R^2 value of 0.9588, and the blue one is the curve of the fourth-order polynomial equation with an R^2 value of 0.9661. The last curve in red represents a fifth-order polynomial equation with an R^2 value of 0.9963 (close to 1). This last curve almost exactly follows the path of the ground truth curve. From the point of view of the R^2 value, which describes how much the independent variable "x" influences the dependent variable "y" with the minimum number of variables "x" used [13–17], the higher the R^2 value, the more accurate the fit between the model and the dataset [18]. Thus, it can be assumed that the proposed fifth-order polynomial model can be used to predict the caloric value of waste, especially in all cities/provinces in Indonesia.

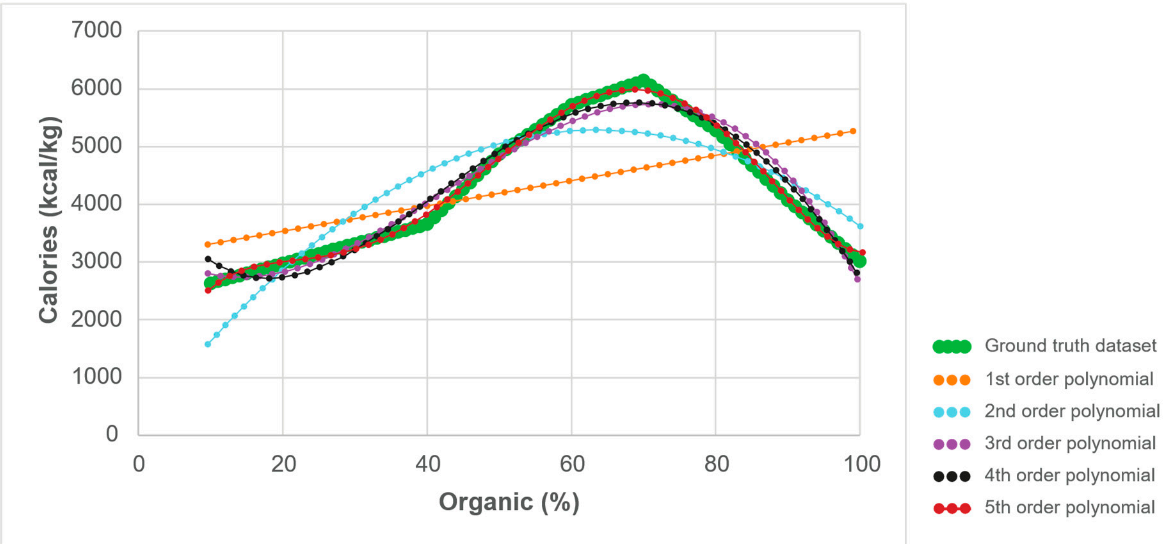

Figure 4. Polynomial regression.

2.3. Model Validation

The validation process for the polynomial regression model that has been developed to predict the calorie content in waste was carried out in accordance with the process sequence shown in Figure 1. This started with collecting the waste composition data, followed by the data preprocessing process, which consists of grouping and determining the organic and non-organic content in waste, and then calculating the normalization of the percentage of organic and non-organic content; finally, the RDF calorie forecasting model is produced, and the result is the RDF calorie potential.

2.3.1. Waste Composition Dataset

A dataset is a set of data that come from past information and are managed into information. In general, the dataset obtained still has noise in each of its attributes, so it is necessary to pre-process the data so that the dataset can be used for the clustering process. The organic waste composition dataset was obtained from 2019, 2020, and 2021, containing the total city data of 618 districts/cities from 34 provinces. The next dataset obtained was the RDF calorie testing dataset based on the composition of organic and non-organic waste. From the laboratory results dataset, an equation function with the highest R^2 result is sought with the optimal number of variables, and this equation function is the one that will be used to calculate the RDF calorie potential from the waste composition dataset throughout Indonesia. Figure 5 shows the percentage of organic components contained in waste in each province in Indonesia for 3 years (2019, 2020, and 2021). Meanwhile, Figure 6 shows the average amount of RDF that can be produced per day in the period of 2019 to 2021.

Next, after the organic waste composition dataset is determined for each province, a dataset of RDF production for each province is collected, which is taken from the average amount of RDF production per day for 3 years (2019, 2020 and 2021). The graph of total RDF production is shown in Figure 6. The graph shows the provinces that produce RDF in large quantities, such as East Java, Central Java, Jakarta, West Java, North Sumatra and Banten. This amount of RDF is obtained from 30% of the amount of waste produced by each province after processing the waste into raw energy materials.

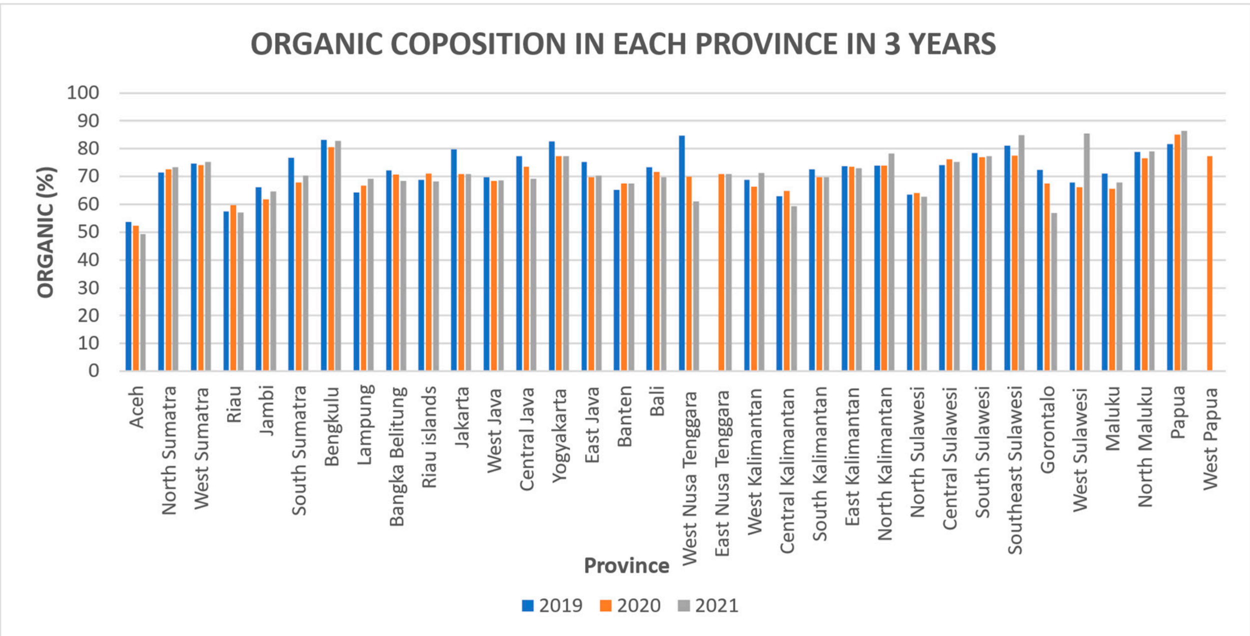

Figure 5. Characteristics of organic waste.

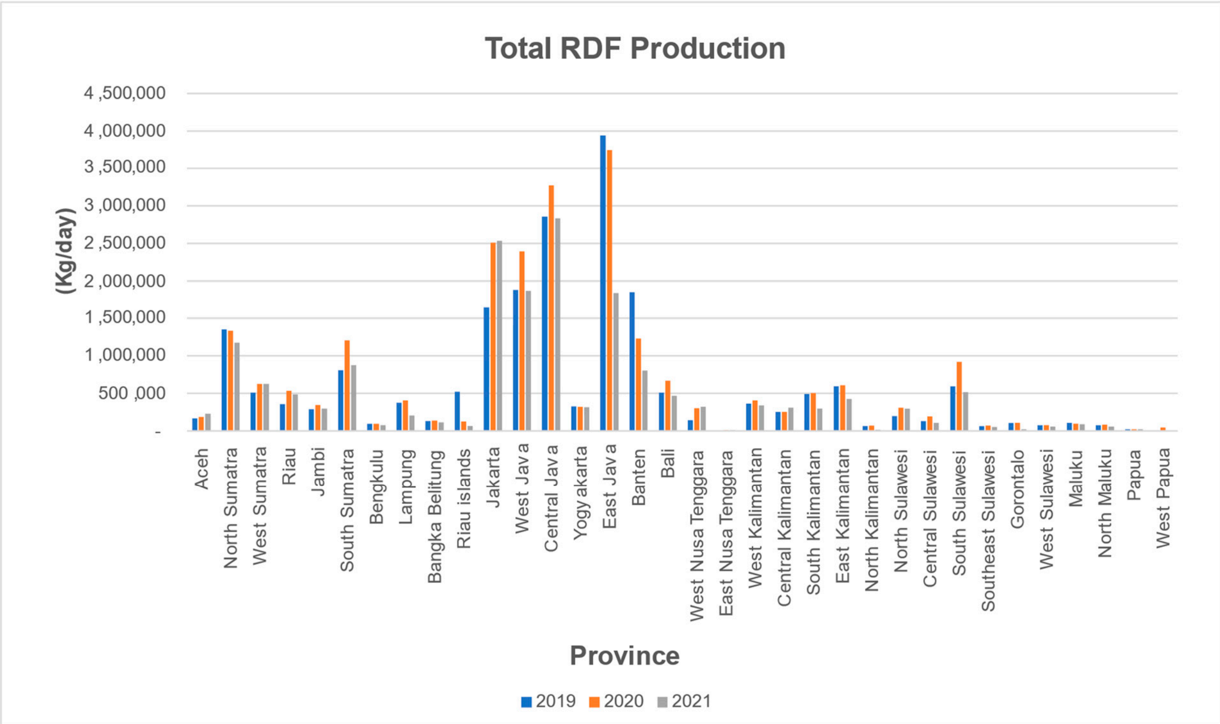

Figure 6. Total RDF production.

The data preprocessing process is a technique used in initial data mining to improve raw data being collected from various data sources, transforming them into cleaner information so that they can be used in further processing. This process is also known as the first step in obtaining all available information by cleaning, filtering, and combining the data. Three problems are generally solved in the preprocessing phase: handling missing values, noise-affected data, and inconsistent data. Missing values are considered inaccurate data, because the missing values in the data make the information in them become irrelevant. This problem often occurs when there are problems in the data collection process, such as data entry errors. Noisy data include wrong data and outliers (data points that are far from other data points) that might be found in other datasets. Several incidents of data noise are caused by human error in the form of labeling errors, as well as other data collection problems. Data inconsistency will occur when files containing the same data are stored in different file formats. This inconsistency includes duplication in different formats, such as wrong code names, etc. To deal with the problem of missing values, noise, and inconsistent data, the stages of data cleaning, data reduction, data transformation, and data integration are carried out. There are some data that experience noise or missing values, so mean calculations are performed on the existing data. Mean calculations can be performed on annual mean data and provincial mean data if a city's data have noise or missing values.

2.3.2. Processing Data

The proposed fifth-order polynomial regression model produced in Equation (6) is then used to calculate the calorie potential of waste produced by cities in Indonesia for 3 consecutive years: 2019, 2020, and 2021. The results of the calorie potential calculation are shown in Table 3. The first column shows the name of the city, the second column is the total number of tons of gross waste estimated on average per day, and the fifth column is the weight of the RDF net waste (i.e., 30% of the gross waste weight), which can be processed to produce calories. The third column states the percentage of organic content in the RDF. This organic percentage becomes the x variable in Equation (6). The fourth column is the calculation result from Equation (6), as the potential calories that can be produced in every 1 kg of waste according to the percentage of organic content in the third column.

The last column states the average total calories that can be produced each day by each city. This result is obtained from multiplying the fourth and fifth columns.

Table 3. Data processing results in 2021.

Province	Average		Total	
	Organic (%)	Calories Potential (kcal/kg)	RDF Production (kg/day)	Calories Potential (Gcal/day)
Central Java	69.14	5381.91	2,832,456	15,226
Jakarta	70.92	4994.74	2,534,331	13,600
West Java	68.69	5892.39	1,864,335	11,011
East Java	70.39	5346.11	1,838,235	10,383
North Sumatra	73.36	5468.74	1,172,376	6573
South Sumatra	70.22	5490.26	876,996	4901
Banten	67.51	5849.91	803,883	4688
West Sumatra	75.21	5519.28	623,151	3485
Bali	69.78	5578.19	466,527	2670
South Sulawesi	77.26	5275.90	515,061	2621
East Kalimantan	73.01	5635.16	424,764	2400
Riau	57.05	4927.46	486,057	2328
Yogyakarta	77.24	5549.15	313,302	1781
West Nusa Tenggara	61.11	5488.39	317,232	1728
North Sulawesi	62.69	5534.00	295,089	1689
Central Kalimantan	59.30	5435.41	304,038	1688
West Kalimantan	71.25	5289.66	334,758	1660
South Kalimantan	69.76	5497.36	294,612	1635
Jambi	64.56	5418.51	294,942	1521
Lampung	69.17	5711.55	205,104	1178
Aceh	49.33	4794.52	229,037	1112
Bangka Belitung Island	68.41	5799.70	111,504	648
Central Sulawesi	75.23	5697.15	104,169	586
Maluku	67.76	5986.27	83,310	499
Riau islands	68.27	5986.40	62,016	371
Bengkulu	82.82	5028.35	71,265	358
North Maluku	79.08	5434.43	56,646	303
West Sulawesi	85.46	4686.79	57,423	269
Southeast Sulawesi	84.95	4755.04	48,045	228
Gorontalo	56.89	5458.33	19,380	106
Papua	86.35	4565.76	17,526	80
North Kalimantan	78.23	5527.33	12,543	69
East Nusa Tenggara	70.83	5959.08	7800	46

2.3.3. RDF Calorie Forecasting Model Validation

The model validation process is a test of the accuracy of the calorie prediction model equation, carried out by comparing the results using the prediction model equation with the largest difference from the reference data (the calorie test dataset). The attributes of the calorie potential of each city produced will be calculated using the difference between the upper and lower limits of the calculations obtained based on the calorie test dataset. The largest difference will be used as the deviation value (error) of the resulting calorie prediction model equation.

The stages of carrying out model validation are as follows:

1. A lower limit (round down) and an upper limit (round up) are set on the organic attributes of each city.
2. The lower and upper limit calories are determined by referring to the calorie value according to the calorie test dataset.
3. The deviation difference between the lower and upper limit caloric values and the potential calories for each city is calculated, with the result being the absolute value (positive number) of the deviation.

4. The deviation value of each city is the maximum result of the deviation of the upper or lower limits.
5. The deviation value of each determined city (maximum deviation) is then compared to the calorie potential of each city produced in percent (%)
6. The deviation value/calorie prediction model's validation is thereby obtained from the average deviation (%) of all cities.

The following are the results of calculating the calorie difference from the calorie test dataset compared with the calorie potential dataset obtained from the calculation results using all the calorie prediction model equations, in stages according to Figure 7.

					1	1	2	2	3	3	4	5
					Lab Test Calories (Reference)				Deviation			
No	Province	City	Organic (%)	Calorie Potential (kcal)	Lower Limit (%)	Upper Limit (%)	Lower Calories (kcal)	Upper Calories (kcal)	Lower Limit (kcal)	Upper Limit (kcal)	Max (kcal)	Error (%)
1	Aceh	City 1	51.09	4913.73	51	52	4946.80	5032.60	33.07	118.87	118.87	2.42%
2		City 2	28.05	3156.29	28	29	3244.20	3278.27	87.91	121.98	121.98	3.86%
3		City 3	37.66	3647.39	37	38	3550.80	3584.87	96.59	62.52	96.59	2.65%
4	North Sumatra	City 4	55.56	5342.63	55	56	5290.00	5375.80	52.63	33.17	52.63	0.99%
5		City 5	61.25	5770.53	61	62	5760.90	5802.80	9.63	32.27	32.27	0.56%
6		City 6	54.43	5239.98	54	55	5204.20	5290.00	35.78	50.02	50.02	0.95%
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....	.....
615	Papua	City 615	80.89	5255.97	80	81	5280.00	5159.20	24.03	96.77	96.77	1.84%
616	Papua	City 616	78.11	5538.60	78	79	5451.60	5365.80	87.00	172.80	172.80	3.12%
617	Papua	City 617	77.26	5612.89	77	78	5537.40	5451.60	75.49	161.29	161.29	2.87%
618	West Papua	City 618	81.75	5157.91	81	82	5159.20	5038.40	1.29	119.51	119.51	[illegible]
									Average		**102.69**	6 **1.95%**

Figure 7. The stages of calculating model deviations from the dataset of the laboratory test results.

The stages of the process for calculating model validation based on Figure 7 in the order according to the yellow highlighted numbers are as follows:

1. The lower limit (round down) and upper limit (round up) are determined for the organic attributes of each city; for example, in City 1, the organic value is 51.09%, and the lower limit = 51%, and the upper limit = 52%.
2. The lower limit and upper limit calories are determined by referring to the calorie value according to the calorie test dataset (Table 2) with a potential calorie value of 4946.80 kcal/kg (for the organic value lower limit of 51%) and a potential calorie value of 5032.60 kcal/kg (for the organic value upper of limit of 52%).
3. The difference between the lower limit calorie value and the upper limit calorie value is determined with the calorie potential of each city, with the result being the absolute value (positive number) of the deviation. For example, in City 1, the lower limit deviation value is 4946.80–4917.73, resulting in 33.07, while the upper limit deviation value is 5032.60–4917.73, resulting in 118.87.
4. The deviation value for each city is the maximum value of the difference between the upper and lower limits. From example point 3, the deviation values obtained are 33.07 and 118.87, so the largest deviation value is 118.87 for City 1.
5. The specified deviation value for each city (maximum deviation) is then compared with the calorie potential of each city produced in percent (%). For example, in City 1, when the resulting deviation value of 118.87 is divided by the potential calories of 4917.73, a value of 2.42% will be obtained.
6. The deviation value/calorie prediction model validation is obtained from the average deviation (%) of all cities, with a deviation value of = 1.95%.

Based on the stages in Figure 7, the average deviation of all equations can be determined using the results in Table 4. Thus, based on the validation tests on the deviations of the waste composition dataset for all cities in Indonesia with the reference data (lab test calories), it can be concluded that the fifth polynomial regression $y = 0.00003\,x^5 - 0.0069\,x^4 + 0.6298\,x^3 - 24.3245\,x^2 + 432.8401\,x + 55.7448$ is the best equation.

Table 4. Deviation testing results.

No	Regression Type	R^2	Validity Test
1	Linear Regression	0.2676	20.93%
2	Second-order Polynomial Regression	0.7687	10.65%
3	Third-order Polynomial Regression	0.9588	4.73%
4	Fourth-order Polynomial Regression	0.9661	5.64%
5	Fifth-order Polynomial Regression	0.9963	1.95%

It can be seen in Table 4 that the R^2 validation value has a correlation with the results of the validation test; in linear regression with $R^2 = 0.2676$, a deviation (error) of 20.93% is obtained when used to calculate the potential calories in the waste composition dataset, whereas in the fifth-order polynomial regression with a value of $R^2 = 0.9963$, a deviation (error) of 1.95% is obtained when calculating the potential calories in the waste composition dataset.

3. Results

Based on the calculation of the caloric potential of energy raw materials for cities in Indonesia using a model from the fifth-order polynomial regression equation with waste composition data obtained from https://sipsn.menlhk.go.id/sipsn/public/data/komposisi (accessed on 1 September 2022) (Table 3), the resulting potential calories of energy raw materials (RDF) from all cities in Indonesia are obtained. Next, the cities are grouped by province by calculating the average calorie potential of energy raw materials for each province, so that the calculation of this potential becomes an actual picture of the calorie potential of energy raw materials for each province, as shown in Table 5.

Table 5. RDF caloric potential.

No	Province	RDF Caloric Potential (Gcal/day)			
		2019	2020	2021	Average
1	East Java	21,366.32	20,492.80	10,383.05	17,414.06
2	Central Java	15,285.58	17,948.61	15,225.67	16,153.29
3	Jakarta	8839.53	13,474.05	13,600.24	11,971.27
4	West Java	10,895.98	13,942.48	11,010.50	11,949.65
5	Banten	10,846.63	7099.22	4687.51	7544.45
6	North Sumatra	7675.29	7318.71	6573.38	7189.13
7	South Sumatra	4470.25	6378.12	4901.50	5249.95
8	South Sulawesi	2923.75	4897.75	2620.55	3480.68
9	West Sumatra	2759.52	3417.76	3484.62	3220.63
10	Bali	2852.60	3728.89	2670.33	3083.94
11	East Kalimantan	3315.45	3387.21	2399.59	3034.08
12	South Kalimantan	2640.59	2777.10	1634.97	2350.88
13	Riau	1907.11	2739.27	2327.69	2324.69
14	West Kalimantan	1731.28	1925.22	1660.12	1772.20
15	Yogyakarta	1580.87	1802.62	1780.91	1721.47
16	Lampung	1751.87	1991.08	1178.19	1640.38
17	Jambi	1492.66	1715.99	1520.89	1576.51
18	Central Kalimantan	1417.70	1425.18	1688.39	1510.42
19	North Sulawesi	1093.42	1740.22	1688.80	1507.48
20	West Nusa Tenggara	691.38	1790.14	1727.97	1403.16
21	Riau Islands	3104.68	726.90	371.24	1400.94

Table 5. *Cont.*

No	Province	RDF Caloric Potential (Gcal/day)			
		2019	2020	2021	Average
22	Aceh	855.66	933.84	1112.47	967.32
23	Central Sulawesi	746.02	1054.88	585.79	795.56
24	Bangka Belitung Island	762.17	790.46	647.93	733.52
25	Maluku	597.27	539.94	498.72	545.31
26	Bengkulu	449.18	467.85	358.48	425.17
27	North Maluku	413.26	426.76	302.67	380.90
28	Gorontalo	472.68	485.53	105.78	354.66
29	West Sulawesi	300.48	331.59	269.13	300.40
30	Southeast Sulawesi	314.07	354.21	228.46	298.91
31	North Kalimantan	349.98	390.45	69.33	269.92
32	West Papua		253.58		253.58
33	Papua	90.01	82.96	80.02	84.33
34	East Nusa Tenggara		45.84	46.48	46.16

According to the research scheme in Figure 2, the calorie yield of energy raw materials (RDF) produced from the WTE process is determined. From this process, a calorie test dataset is obtained and used as a standard for testing the results of other energy raw material products. The calorie test dataset is then analyzed by looking at the results of the scatter plot. A scatter plot is a graphical representation that signifies the linear relationship between pairs of independent variables [13]. At this stage, the selection of linear regression or polynomial regression is carried out. Polynomial regression is one of the most widely used curve fitting methods [12]. The equation model that has been selected with the highest R^2 with the optimal number of independent variables will be the RDF calorie determination model equation. Using the RDF calorie determination model equation, the RDF caloric potential for all cities will be obtained, and can be used for decision making by the government to determine further policies in the management of WTE in an area.

By implementing the fifth-order polynomial regression equation, it is possible to calculate the RDF caloric potential throughout Indonesia, in each province or city, with the waste composition dataset for 2019, 2020, and 2021. The results of calculating the RDF caloric potential are displayed by year and province in Figures 8–10.

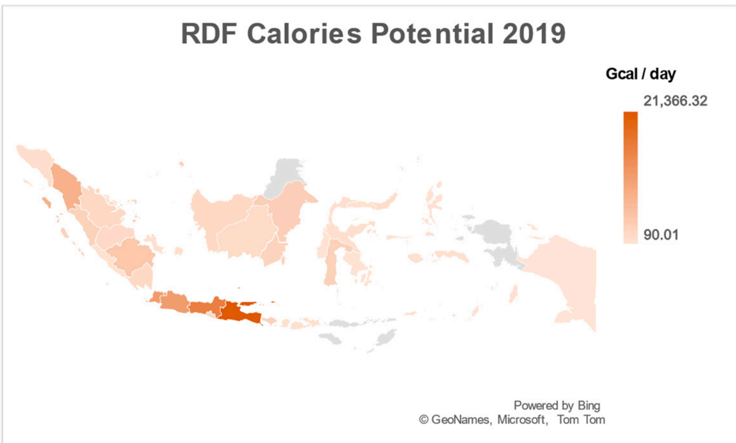

Figure 8. RDF calorie potential in 2019.

Figure 9. RDF calorie potential in 2020.

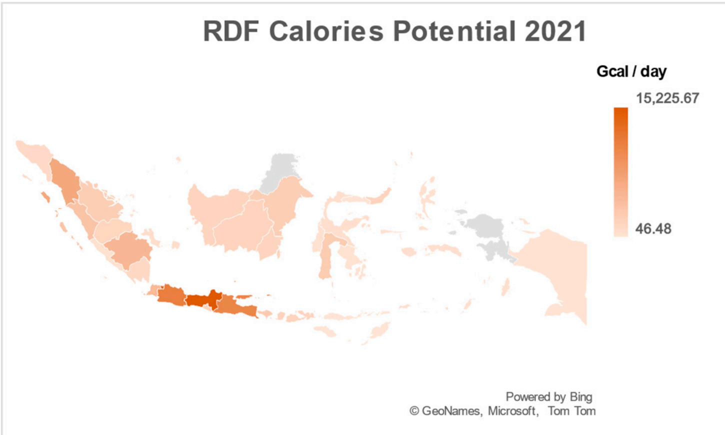

Figure 10. RDF calorie potential in 2021.

Based on the calculation of the calorie potential of energy raw materials using a fifth-order equation, the potential to transform waste into energy can be determined, as shown in Table 5. The government can use these results to determine policies for transforming waste into energy raw materials so that the renewable energy target stipulating a proportion of energy from waste (bioenergy) can be met. In Table 5, it can be seen that the ten provinces with the largest calorie potential for energy raw materials per day are East Java, Central Java, Jakarta, West Java, Banten, North Sumatra, South Sumatra, South Sulawesi, West Sumatra, and Bali. The mechanism of bio-drying is a variation of the aerobic decomposition used within mechanical–biological treatment (MBT) to stabilize waste, which makes it analogous to composting but achievable in the short term [6,19].

4. Conclusions

This research uses two datasets, namely the RDF calorie test dataset (as reference data) and the waste composition dataset from throughout Indonesia (as test data). This is a new approach used to predict RDF calorie potential more accurately, and is the novelty of this research. The equation was obtained by comparing five equation models, namely through linear regression and polynomial regression. Based on scatter plot observations and evaluation calculations, it was found that the results of the RDF caloric testing dataset produced an optimal function equation model of fifth-order polynomial regression with $R^2 = 0.9963$, and the error validity test = 1.95% with the equation $y = 0.00003\,x^5 - 0.0069\,x^4 + 0.6298\,x^3 - 24.3245\,x^2 + 432.8401\,x + 55.7448$. The results of this equation were used to calculate the calorie potential of the waste in each province in Indonesia, and the results of calculating the potential for RDF are the second novelty of this research.

Author Contributions: All authors contributed equally to this work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The data used in this study can be downloaded at: https://sipsn.menlhk.go.id/sipsn/public/data/komposisi, accessed on 1 September 2022.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. PLN. *Textbook of National Electricity Supply Business Plan (NESBP) PT PLN (Persero) 2021–2030*; PT PLN: Jakarta, Indonesia, 2021; pp. 68–80.
2. Usta, S.; Gencoglan, C.; Gencoglan, S. Estimation of Daily Average Global Solar Radiation with Nonlinear Regression Models Developed Using Some Meteorological and Geographical Parameters. *Dicle Üniversitesi Mühendislik Fakültesi Mühendislik Dergisi* **2022**, *3*, 589–597. [CrossRef]
3. Chen, D.M.C.; Bodirsky, B.L.; Krueger, T.; Mishra, A.; Popp, A. The world's growing municipal solid waste: Trends and impacts. *Environ. Res. Lett.* **2020**, 15. [CrossRef]
4. Brahimi, T.; Kumar, C.R.; Abdulmajid, M.; Alyamani, N. Sustainable Waste Management through Waste to Energy Technologies in Saudi Arabia: Opportunities and Environmental Impacts. In Proceedings of the International Conference on Industrial Engineering and Operations Management, Riyadh, Saudi Arabia, 26–28 November 2019; Volume 9.
5. Albores, P.; Petridis, K.; Dey, P.K. Analysing Efficiency of Waste to Energy Systems: Using Data Envelopment Analysis in Municipal Solid Waste Management. *Procedia Environ. Sci.* **2016**, *35*, 265–278. [CrossRef]
6. Latosińska, J.; Żygadło, M.; Dębicka, M. The Biological Drying of Municipal Waste in an Industrial. *Energies* **2022**, *15*, 1039. [CrossRef]
7. Anssari, O.M.; Alkaldy, E.A.; Almudhaffar, N.; AlTaee, A.N.; Ali, N.S. A feasibility study of electrical energy generation from municipal solid waste in Iraq: Najaf case study. *Int. J. Electr. Comput. Eng.* **2020**, *10*, 3403–3411. [CrossRef]
8. Wang, W. Integrated Assessment of Economic Supply and Environmental Effects of Biomass Co-Firing in Coal Power Plants: A Case Study of Jiangsu, China. *Energies* **2023**, *16*, 2725. [CrossRef]
9. Kraemer, H.C.; Blasey, C. Linear Regression Analysis. In *How Many Subjects?: Statistical Power Analysis in Research*; Sage Publications: Thousand Oaks, CA, USA, 2017; pp. 73–85. [CrossRef]
10. Gujarati, D.N. *Basic Econometrics*, 4th ed.; Gary Burke: New York, NY, USA, 2003; Volume 82. [CrossRef]
11. Fox, J. *Applied Regression Analyis and Generated Linear Models*, 3rd ed.; SAGE: London, UK, 2016.
12. Jose, J.T.; Chattopadhyay, A.B. Modeling of the magnetizing phenomena of doubly fed induction generator using neuro-fuzzy algorithm considering non-linearity. *Int. J. Electr. Comput. Eng.* **2019**, *9*, 23. [CrossRef]
13. Shrestha, N. Detecting Multicollinearity in Regression Analysis. *Am. J. Appl. Math. Stat.* **2020**, *8*, 39–42. [CrossRef]
14. Yang, M.; Xu, D.; Chen, S.; Li, H.; Shi, Z. Evaluation of machine learning approaches to predict soil organic matter and pH using vis-NIR spectra. *Sensors* **2019**, *19*, 263. [CrossRef] [PubMed]
15. Abaszade, R.G.; Mammadov, A.G.; Kotsyubynsky, V.O.; Gur, E.Y.; Bayramov, I.Y.; Khanmamadova, E.A.; Kapush, O.A. Modeling of Voltage-Ampere Characteristic Structures on the Basis of Graphene Oxide/Sulfur Compounds. *Int. J. Tech. Phys. Probl. Eng.* **2022**, *14*, 302–306.
16. Amar, L.A.; Taha, A.A.; Mohamed, M.Y. Prediction of the final size for COVID-19 epidemic using machine learning: A case study of Egypt. *Infect. Dis. Model.* **2020**, *5*, 622–634. [CrossRef] [PubMed]
17. Brown, A.M. A step-by-step guide to non-linear regression analysis of experimental data using a Microsoft Excel spreadsheet. *Comput. Methods Programs Biomed.* **2001**, *65*, 191–200. [CrossRef] [PubMed]
18. Fan, C.; Ding, Y. Energy & Buildings Cooling load prediction and optimal operation of HVAC systems using a multiple nonlinear regression model. *Energy Build.* **2019**, *197*, 7–17. [CrossRef]
19. Bhatsada, A.; Patumsawad, S.; Towprayoon, S.; Chiemchaisri, C.; Phongphiphat, A.; Wangyao, K. Modification of the Aeration-Supplied Configuration in the Biodrying Process for Refuse-Derived Fuel (RDF) Production. *Energies* **2023**, *16*, 3235. [CrossRef]