Forecasting Customer Growth at PT PLN (Persero) ULP Lakawan Using Linear Regression

Ramanda Asyari*, Ayi Oktra Maulana, Ikhlas, Adha Rizki, and Muhammad Ridho

Electrical Engineering, Padang State Polytechnic, West Sumatera, Indonesia

* Correspondence e-mail: ramandaasyari@gmail.com

Citation: To be added by editorial staff during production.

Abstract: Electrical energy is a fundamental necessity for the population. PT PLN (Persero) ULP Lakawan is a state-owned enterprise mandated by the government to manage the national power supply. The rising public demand for electricity has led to a significant increase in the potential number of customers. Consequently, forecasting the number of customers for the year 2026 is essential as a basis for strategic planning and decision making to ensure an adequate and reliable electricity supply. This study utilizes historical customer data from 2016 to 2025 to capture trend of customer growth st PT PLN (Persero) ULP Lakawan. By employing a Simple Linear Regression analysis, conducted through both manual Excel calculations and data analytics tools, the study predicts that the number of customers in 2026 will reach 56,632. The model demonstrated high precision with a P-Value of 0. Furthermore, the analysis revealed a substantial impact with an R-Square value of 98.73% and an exceptionally strong correlation (Multiple R) of 0.99. This research indicates that the Simple Linear Regression method is highly accurate for predicting customer growth.

Keywords: Electrical Energy, PT PLN (Persero), Simple Linear Regression Analysis, ULP Lakawan

1. Introduction

Electrical energy is a vital need for the population. Nearly everything we do requires electricity, both in household and industrial activities (Ile et al, 2024). Economic growth, infrastructure development, and population growth will lead to increased public demand for electricity connections from PT PLN (Persero).

PT PLN (Persero) is a government-mandated business entity tasked with providing electricity for the public interest [2]. This includes PT PLN (Persero) ULP Lakawan, a custumer service unit located in South Sulawesi. PT PLN (Persero), as the primary electricity provider in Indonesia, has a significant responsibility to ensure a stable, continuous electicity supply that meets the needs of communities in various regions (Oktavia et al, 2025).

The increasing public demand for electricity creates a significant opportunity for an increase in the number of customers at PT PLN (Persero). Therefore, PT PLN (Persero) must have an electricity system plan, including material inventory, network, and power supply availability. To develop this plan, a forecast of future customer numbers is required.

Predictions of future customer numbers will serve as the basis for PT PLN (Persero) in making decisions. Inaccuracies in predicting future customer numbers can risk overloading distribution equipment or, conversely, inefficient infrastructure investments.

Therefore, researchers use a scientific and measurable forecasting method. Simple Linear Regression is an effective statistical technique for modeling the relationship between time as the independent variable and the number of customers as the dependent variable. By utilizing historical customer data from previous years, this method can accurately project growth trends for the coming year. Simply put, accuracy indicates how close the predicted results are to the actual value (Damayanti et al, 2025).

2. Methods

This study uses data on the number of PT (PLN Persero) customers from 2016-2025 sourced from TUL III-09 on PLN's AP2T website. As for the research, it will predict the number of customers in 2026 using the Simple Linear Regression method using Microsoft Excel which is calculated manually and with data analyst tools, by ensuring the correctness of the data processing results.

Prediction is basically a conjecture or prediction about the occurrence of an event or event in the future. Prediction serves to make a demand plan that must be made which is expressed in quantity (quantity) as a function of time (Syafruddin et al, 2014).

Regression is an important measurement tool in statistics to evaluate the relationship between variables (Nurhaswinda et al, 2025). A research variable is a value/trait/attribute of an object, activity, or person with a variation that is determined to be studied and concluded by the researcher (Sugiyono, 2017).

Simple linear regression (Simple Linear Regression) is an equation model that describes the relationship of one causal variable (X) to a consequence variable (Y) (Rahmadani, 2023). Simple linear regression involves one independent variable and one dependent variable (Nurhaswinda et al, 2025). The influencing variables are called independent or independent variables, while the affected variables are called related or dependent variables (Rudi et al, 2023).

The following formula for simple linear regression is
$$Y = a + bx, \tag{1}$$

with
$$a = \frac{(\sum Y)\,(\sum X^2) - (\sum X)\,(\sum XY)}{n \sum X^2 - (\sum X)^2},$$

$$b = \frac{n\,(\sum XY) - (\sum X)\,(\sum Y)}{n \sum X^2 - (\sum X)^2}.$$

Description :

Y = actual data value

Y' = predictive value

n = amount of data

a = constant

b = regression coefficient (positive or negative)

In a simple linear regression test, there are several parameters that need to be considered, namely Analysis of Variance, Coefficients, and Model Summary (Yusuf, 2021).

1) Analysis of Variance serves to test whether the regression model is in accordance with the existing data sample provided that if the p-value is smaller than the α value of 0.05, it can be concluded that the resulting regression model meets the existing data.

2) Coeffiencients serves to test the presence of a significant influence between independent variables on dependent variables. If the p-value of the independent variable is less than the α value of 0.05, then it can be concluded that the variable has a significant effect on the dependent variable.

3) The Summary Model functions to calculate the percentage of the influence of the Independent variable on the dependent variable based on the value of the determination coefficient or also called with. R-square Adjusted (R-sq (adj)).

P_Value

When making a resume about P-value to people outside the field of statistics, it is often necessary to use descriptive language to demonstrate the strength of the evidence (Abadi, 2013).

Tabel 2. Kriteria P-Value

No	Criteria	Remarks
1	$P > 0.10$	There is no evidence against hypothesis 0. Data tends to be consistent with the zero hypothesis
2	$0.05 < P < 0.10$	Weak evidence against hypothesis 0 versus alternative hypotheses
3	$0.01 < P < 0.05$	Evidence for supporting hypotheses versus alternative hypotheses
4	$0.001 < P < 0.01$	Strong evidence of hypothesis 0 versus alternative hypotheses
5	$P < 0.001$	Very strong evidence against hypothesis 0 versus alternative hypotheses

Correlation Coefficient (R) and Coefficient of Determination (R–Square)

The coefficient of determination (R-Square) is a measure that shows how strongly an independent variable has an influence on a dependent variable (Nufaily & Siregar, 2025).

Data Correlation Analysis The Correlation Coefficient (R) and Coefficient of Determination (R–Square) is a form of data analysis in research that aims to determine the strength or shape of the direction of the relationship between two variables and the magnitude of the influence caused by one variable (independent variable/independent variable/variable X) on another variable (bound variable/dependent variable/variable Y) (Hutapea, 2025).

The correlation coefficient test is a number that expresses the strength of the relationship between two or more variables, it can also determine the direction of the relationship of the two variables. The correlation value (R2) ranges from (+1) to (-1), where (+) indicates that there is a positive relationship between variable X and variable Y and vice versa. In the correlation efficiency test, there is the following level of relationship.

Table 3. Correlation Rate Table

Interval Cowphysin	Relationship Level
1,00	Perfect
0,80 – 0,999	Very Powerful
0,60 – 0,799	Strong
0,40 – 0,599	Quite Powerful
0,20 – 0,399	Weak
0,00 – 0,199	Very weak

3. Results and Discussion

Based on data from TUL III-09, the following is data on the number of customers of PT PLN (Persero) ULP Lakawan over the last 10 years along with the graph.

Table 4. Total Customers of PT PLN Persero ULP Lakawan

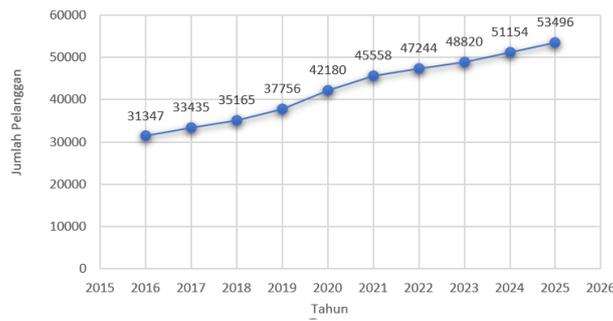| Year	Number of Customers
2016	31347
2017	33435
2018	35165
2019	37756
2020	42180
2021	45558
2022	47244
2023	48820
2024	51154
2025	53496

Figure 1. Graph of the Number of PLN ULP Lakawan Customers

Based on the graph above, it can be seen that the number of PLN ULP Lakawan customers from 2016-2025 tends to increase from year to year.

Predict the number of customers in 2026 using:

1. The Linear Regression Method is simple using manual calculations in Microsoft excel

   a. Values X^2, Y^2, XY

   First, find the values of X^2, Y^2, XY by squaring and multiplying by the variables x (years), and the variable y (number of customers) from each existing data.

   Example calculation:

   Year 2016 :

   X^2 = (2016) * (2016) = 4064256

   Y^2 = (31347) * (31347) = 982634409

   XY = (2016) * (31347) = 63195552

   Here are the results and overall totals of the X^2 Y^2 XY Value from 2016 to 2025 shown in the table.

Table 5. The Value of X^2, Y^2 and XY

NO	X	Y	X^2	Y^2	XY
1	2016	31347	4064256	982634409	63195552
2	2017	33435	4068289	1117899225	67438395
3	2018	35165	4072324	1236577225	70962970
4	2019	37756	4076361	1425515536	76229364
5	2020	42180	4080400	1779152400	85203600
6	2021	45558	4084441	2075531364	92072718
7	2022	47244	4088484	2231995536	95527368
8	2023	48820	4092529	2383392400	98762860
9	2024	51154	4096576	2616731716	103535696
10	2025	53496	4100625	2861822016	108329400
Sum		426155	40824285	18711251827	861257923

b. Values of constants a and b

$$a = \frac{(\sum Y)\,(\sum X^2) - (\sum X)\,(\sum XY)}{n\,\sum X^2 - (\sum X)^2}$$

$$a = \frac{426155 \; x \; 40824285 - 20205 \; x \; 861257923}{10 \; x \; 40824285 - (20205)^{\wedge}2}$$

a = -5143224,291

Thus,

$$b = \frac{n\,(\sum XY) - (\sum X)\,(\sum Y)}{n\,\sum X^2 - (\sum X)^2}$$

$$b = \frac{10 \; x \; (861257923) - 20205 \; x \; 426155}{10 \; x \; 40824285 - (2025)^{\wedge}2}$$

b = 2566,6121.

c. Simple Linear Regression Test

Y = a + bx

Y = -5143224.291 + 2566.6121 (2026)

= 56731,8236

2. The Linear Regression method is simple using manual calculations of data analyst tools.
The research data that has been input into Microsoft excel with the linear regression analysis method using the Data Analysis tool then select regression then ok as listed in the image below:
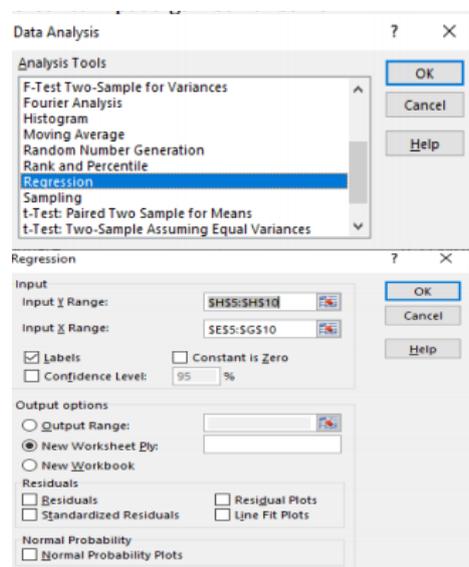

Figure 3. Tools Data Analyst dan Regression

a. Constant a and Regression Coefficient (b).

Based on the data analyst tools in Microsoft Excel, the P-Value value of 0, less than 0.05, can be concluded that variable x significantly affects variable Y.

Table 6. Simple Linear Regression

	Coefficients	Standard Error	t Stat	P-value
Intercept	-5143224,2909	207708,8485	-24,7617	0,0000
X Variable 1	2566,6121	102,8006	24,9669	0,0000

The value of constant (a) is shown at the intercept value of a, which is -5143224.2909, the value of the regression coefficient (b) is 2566.6121 with a t-calculated value (to) of 24.9669 and a significance value of p-value of 0. The positive value regression coefficient shows that the direction of the influence is positive, meaning that the higher the year, the number of PLN customers will also increase.

So,

$$Y = a + bx$$
$$Y = -5143224.2909 + 2566.6121x$$

b. Correlation Coefficient (R) and Coefficient of Determination (R^2)

X is the year and Y is the customer's predicted amount. Multiple R = 0.99 means that the correlation value between variable x and y is 0.99 so the correlation between x and y is very strong.

R Square = 0.9873 means that the value of the determination coefficient (R-Square) is 98.73%, this explains that X can explain Y by 98.73%, while the rest is influenced by other factors. Adjusted R Square = 0.9875 This means that all independent variables have a strong relationship with the dependent variables used. Adjust R_square is an important property of R2 to compare two R2 of two models, if the variable is free of more than two, it is best to use adjusted R square. This value is not used because the number of variables is only one.

Table 7. Regression Statistic Results

Regression Statistics	
Multiple R	0,9936
R Square	0,9873
Adjusted R Square	0,9857
Standard Error	933,7329
Observations	10,0000

From the table above, we can conclude that the regression model we obtained is suitable for predicting future events.

3.   Prediction for 2026

To predict the number of customers in 2026, with the values of a and b already obtained, the simple linear regression equation is :

$$Y = -5143224.2909 + 2566.6121x$$
$$= -5143224.2909 + 2566.6121\ (2026)$$
$$Y = 56.732$$

So, the number of customers in 2926 is 56.732 customers.

To find out the predicted number of customers each year, after knowing the values of a, b, we can determine the predicted value of the number of customers each year by entering x as the year to be predicted.

4. Conclusion

Based on the results of the study predicting the number of customers of PT PLN (Persero) ULP Lakawan in 2026 using Microsoft Excel, both through manual and data analysis tools, consistent results were obtained, indicating a projected total of 56,732 customers. The regression analysis confirms that the independent variable (time) has a statistically significant influence on customer growth, as indicated by a p-value < 0.05, an R-squared value of 98.73%, and a very strong correlation coefficient (Multiple R = 0.99). These findings have important implications for electricity planning and policy, as accurate customer forecasts enable PT PLN (Persero) to anticipate future electricity demand, optimize capacity expansion, and plan distribution network development more effectively. Reliable predictions also support informed decision-making related to investment allocation, infrastructure reinforcement, and operational efficiency, thereby reducing the risks of supply shortages or overinvestment. Consequently, the results of this study provide a quantitative basis for short- to medium-term strategic planning and policy formulation in the electricity sector at the local service unit level.

References

Abadi, S. (2013). Aplikasi Paket Statistik Untuk Metode Regresi Linier Dengan Menggunakan Microsoft Excel. Syntax: Jurnal Informatika, 2(02).

Damayanti, A., Marleny, F. D., & Ningrum, A. A. (2025). Implementasi Regresi Linear Berganda Untuk Prediksi Penjualan Pada Pt Trimandiri Sarana Propetindo Banjarmasin. Jurnal Informatika dan Teknik Elektro Terapan, 13(3).

Hutapea, T. W. S. (2025). Dampak Digitalisasi Terhadap Peningkatan Produktivitas Usaha Mikro, Kecil, dan Menengah (UMKM) pada Era Ekonomi Digital di Kota Medan (Doctoral dissertation, Universitas Sumatera Utara).

Ile, K. M., Mangopo, D., Reza, M. A., Liga, M., Sinaga, A. S., & Suparno, S. (2024). Studi Perkiraan Kebutuhan Energi Listrik Pelanggan: Di Pt. Pln (Persero) Unit Layanan Pelanggan Abepura. Jurnal Teletronic, 1(3), 11-16.

Kemenhu. (n.d.). JDIHN.GO.ID. Retrieved from https://jdih.kemenkeu.go.id/kamus hukum/pt-perusahaan-listrik-negara-(persero),?id=22959008acdcf67c8f6a05fa93a08f1a.

Muriyatmoko, D. (2018). Analisa volume terhadap sitasi menggunakan regresi linier pada jurnal bereputasi di Indonesia. Jurnal Simantec, 6(3).

Nufaily, F. R., & Siregar, M. U. (2025). Analisis Ketertarikan Pengguna Microsoft Excel Online untuk Pengolahan Data Silsilah Keluarga Menggunakan TAM dan TPB. JISKA (Jurnal Informatika Sunan Kalijaga), 10(3), 279-293.

Nurhaswinda, N., Egistin, D. P., Rauza, M. Y., Rahma, R., Ramadhan, R. H., Ramadani, S., & Wahyuni, W. (2025). Analisis regresi linier sederhana dan penerapannya. Jurnal Cahaya Nusantara, 1(2), 69-78.

Oktavia, A., Sri Wahyu, Fitroh Resmi. (2025). Analisis Single Exponential Smoothing Dan Doubl Exponential Smoothing Dalam Peramalan Jumlah Pelanggan Listrik Di Pt. Pln Ulp Padang. E-Jurnal.Uinib.Ac.Id, 28-34.

PT PLN (Persero). (n.d.). Aplikasi Pelayanan Pelanggan Terpadu (AP2T). https://ap2t.pln.co.id

Rahmadani, N. (2023). Analisis Prakiraan Kebutuhan Dan Ketersediaan Energi Listrik Di Kabupaten Kolaka Timur Tahun 2022-2031 Dengan Menggunakan Metode Dkl 3.2, Regresi Linear, Dan Software Leap.

Rudi, W. S., Pranoto, Y. A., & Ariwibisono, F. X. (2023). Penerapan Metode Regresi Linier Dalam Peramalan Penjualan Kue Di Toko Karya Bahari Samarinda Berbasis Website. JATI (Jurnal Mahasiswa Teknik Informatika), 7(4), 2451-2457.

Sipayung, S. P., Siagian, N. A., & Ginting, A. K. B. (2024, June). Analisis Prediksi Jumlah Penerimaan Mahasiswa Baru Dengan Metode Regresi Linier Sederhana. In Seminar Nasional Inovasi Sains Teknologi Informasi Komputer (pp. 376-382).

Syafruddin, M. L. H. D. D., Hakim, L., & Despa, D. (2014). Metode Regresi Linier Untuk Prediksi Kebutuhan Energi Listrik Jangka Panjang (Studi Kasus Provinsi Lampung). Jurnal Informatika dan Teknik Elektro Terapan, 2(2).

Yusuf, F. A. (2021). Analisa Statistik Pengaruh Derajat Putih Precipitated Calcium Carbonate Terhadap Kualitas Derajat Putih Kertas (Doctoral dissertation, Institut Teknologi Sains Bandung).