Electricity Load on substations using ARIMA and LSTM models – A comparisan

Case study : Trisakti substation in South Kalimantan

Satria Dina Astari - 23522309
Program Studi Magister Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
E-mail (gmail): satriadina100@gmail.com

***Abstract*— Electrical energy is a very vital energy in everyday life. This fact triggers an increase in the substation feeder load from year to year. To ensure that the electricity supply remains safe, economical and reliable, electricity load forecasting is needed. Electrical loading at substations is influenced by various factors and often contains linear and non-linear patterns. This makes forecasting using statistical or traditional methods currently used inadequate because they do not consider these factors.**

This research uses ARIMA and LSTM methods to overcome the problems mentioned. The Arima model is used to predict components that have linear characteristics. Testing was carried out at the Trisakti substation. The test results show that the proposed ARIMA model can significantly improve forecasting accuracy with an RMSE of 0.5464978840437401.

Keywords: load forecasting, ARIMA, LSTM, substation

I. INTRODUCTION (*HEADING 1*)

Electricity improves peoples prosperity in various fields (economy, education, health, industry etc), PLN manages electricity from generation, transmission to distribution. Substation (GI) are important for continuity of electricity supply. PLN UP2D Kalselteng has a task to provide electricity for public interest to provide a supply of electricity that can meet consumer needs. Substation capacity planning is needed to overcome load imbalances that can affect electricity distribution. PLN UP2D Kalselteng is trying to keep the substation operating normally by minimizing disruptions due to overloading. As a basis for planning, both operational planning and electricity development system planning, one of the important things is accurate forecasting to determine the need for electricity within a certain period of time. By collecting data historical, an analysis of the feeder loading patterns at the substation can be carried out as a strategy to avoid load imbalances on the feeder equipment at the substation.

II. RESEARCH MATERIALS AND METHODOLOGY

II.1Methods

II.1.1 ARIMA

Autoregressive Integrated Moving Average (ARIMA) is often also called the Box-Jenkins time series method. ARIMA has very good accuracy for short-term forecasting, while for long-term forecasting its accuracy is not good.

Bentuk umum model ARIMA adalah

$$Yt = b_0 + b_1 Y_{t-1} + \cdots + b_n Y_{t-n} - a_1 e_{t-1} - \cdots - a_n e_{t-n} + e_t \quad (1)$$

Yt is stationary value series, Y_{t-1}, Y_{t-n} is past value of the series, e_{t-1}, e_{t-n} is independent variable which is a lag of residual, e_t is residual, b_0 is constanta, b_1, b_n, a_1, a_n is model coeffient.

II.1.2 Auto ARIMA

Auto Arima works by performing differencing operations and root tests such as Augmented Dickey-Fuller (ADF), Kwiatkowski-Phillips-Schmidt-Shin (KPSS) and Phillips-Perron to determine order d. After getting order d, the model fitting process is then carried out at the given interval of p and q values. To determine the best model, this research uses the Akaike Information Criterion (AIC). The p and q values with the smallest AIC are selected as the best model parameters.

II.1.3 LSTM

The basic structure of LSTM has two types of vectors, namely h and c. The basic idea of this network is that it can learn what to keep for a long time, what to discard, and what to read from it. Every time input is received in a cell, memory is added and memory is removed.

II.2 Related Works

II.2.1 IHSG Forecasting Analysis using ARIMA time series modeling (Susanti et al, 2020).

From 3 models developed, namely model 1or ARIMA (3,0,1), model 2 or ARIMA (0,3,1), and model 3 ARIMA (7,3,1). Model 1 AIC is 9.6837 and SC is 9.7257. In model 2 the AIC value is 9.6803 and SC is 9.7223. In model 3 the AIC is 9.6627 and the SC value is 9.7187. From these results it can be seen that model 3 has the smallest AIC and SC, therefore model 3 is the most suitable model. After forecasting with model 3, the Root Mean Squared Error (RMSE) value was equal to 30.33293, MAE 22.99950 and MAPE 0.002615.

II.2.2 Cholesterol and Glucose Prediction using ARIMA and LSTM models (Krishnamoorthy et al., 2024).

The results showed that the ARIMA model had better performance in predicting glucose and cholesterol levels compared to the LSTM model. The ARIMA RMSE result for the predicted value of glucose is 31.24, which is lower than the LSTM which has an RMSE value of 109.43. The predicted cholesterol value shows that the RMSE results using ARIMA are 54.62, which is lower than LSTM which has an RMSE value of 110.

II.2.3 ARIMA vs LSTM on NASDAQ stock exchange data. (Kobiela et al., 2022).

The research results show that the ARIMA model has better performance than LSTM. For 30 day predictions, ARIMA's RMSE value is around 3 times lower than LSTM. For 3-month average predictions, ARIMA's RMSE value is around 1.8 times better than LSTM. For an average prediction of 9 months, ARIMA's RMSE value is around 2.1 times better than LSTM.

III. EXPERIMENTAL DESIGN

III.1 Requirement Analysis of System

The distribution of feeder loads for a substation which is initially evenly distributed can become unbalanced due to the growth of an area in terms of population growth, economic sector, industrial sector, offices and so on. This imbalance can cause overloading of some feeders, which can result in equipment disruption, causing power outages.

PLN is required to maintain the continuity of the electricity supply delivered to consumers. Determining the correct substation capacity is very important to ensure adequate electricity supply and avoid load overload. Lack of proper planning can result in the substation's inability to meet increasing electricity demand.

III.2 Scope of System

III.2.1 Time and Place

This research took place at PT. PLN (Persero) Unit Pelaksana Pengatur Distribusi. Data collection was carried out from January 2021 to December 2023.

III.2.2 Types and Sources of Data

The data source used in this research is data regarding the size of the electrical energy feeder load in the form of amperes. This feeder data is manually input at hourly intervals every day at each substation. History data is stored for 12 months of the year. Starting from January 2021 to December 2023.

III.3 Proposed Solution

Solutions applied to the ARIMA and LSTM model implementation process. ARIMA and LSTM are used to make initial predictions from training data and test data. The system design scheme that will be built in this research can be seen in Figure 3.2.

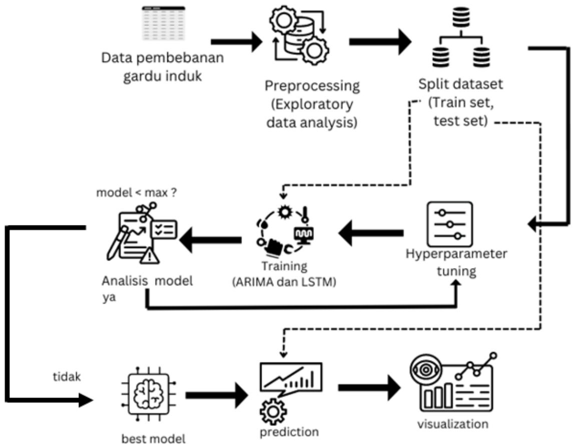

Figure 1. 1 Flowchart of the proposed research process

III.4 Experimental Design

III.4.1 Preprocessing Data

The data preparation stage aims to provide clean data for the next stage. This stage may consist of data selection, cleaning, construction and reorganization of data in preparation for the modeling process. Statistics can help you know what areas need improvement.

At this stage, data preparation greatly influences the quality of data modeling. Data preparation includes handling missing values, filling in values and reducing values.

1. Deal with missing values
   In several attributes in the dataset, there are missing values. However, for correlated attributes, there were no missing values. So that data transformation can be carried out directly.

2. Perform data transformation
   ARIMA and LSTM have different requirements for data preparation. ARIMA requires a complete time series form. Meanwhile, LSTM requires data in matrix form as part of its gate structure.
3. Split the dataset
   To carry out machine learning model creation, testing and evaluation requires several pieces of complete data. This encourages the need to divide the data into several parts, namely training data and test data. Training data is used to form a machine learning model. Then, test data is used to evaluate the performance of the model that has been created.

Before calculations are carried out using the ARIMA method, a series of tests are first carried out such as data stationarity, differentiation process and correlogram testing to determine the autoregression coefficient.

III.4.2 Data Modelling

The modeling phase is a part of data analysis. At this stage, the data will be subjected to various algorithms and modeling techniques to obtain the maximum model. This stage will produce regularities or patterns obtained from the data. This modeling is very dependent on the data preparation stage. Modeling will utilize data through preparation in the previous stages. Modeling in the implementation of this research will utilize ARIMA and LSTM models.

III.4.3 Evaluation

The evaluation stage is used to improve the performance of the modeling stage. This stage aims to assess the suitability of the modeling output with the problem formulation that has been defined. Evaluation is carried out to determine the accuracy and quality of a model. The performance of machine learning will be carried out by utilizing RMSE (root mean squared error) and MAE (Mean Absolute Error). This test was carried out on 20% of the test data which was divided from the complete dataset. This performance value will determine how a model represents data.

Root Mean Square Error (RMSE) is the standard deviation of the residual (prediction error). (Wang & Lu, 2018) Residuals are a measure of how far away from the data points the regression line is. RMSE is a measure of how spread out these residuals are. In other words, it tells how concentrated the data is around the line of best fit. RMSE is usually used in regression analysis to verify experimental results. The smaller the RMSE value, the better the prediction results.

Mean Absolute Error (MAE) is a measure of how accurate a forecasting system is. The absolute error at each time period minus the actual value divided by the actual value. This matrix shows the distance of predicted values.

IV. RESULT AND ANALYSIS

IV.1 Test Scenario

To determine the effectiveness of the model proposed in this research, a comparison of the accuracy of the prediction results with the forecasts from the models used was carried out. The methods used are ARIMA and LSTM. Test evaluation was carried out using the MAE and RMSE evaluation matrices. The best model is the model that has the smallest MAE and RMSE values in the category being forecasted.

IV.2 Hasil Pengujian

IV.2.1 ARIMA

Table 4.1 Matriks evaluasi data training

Matriks	**Nilai**
RMSE	0.4999
MAE	0.3994

Table 4.2 Matriks evaluasi data test

Matriks	**Nilai**
RMSE	0.5464
MAE	0.4551

Tabel 4. 3 Best Model ARIMA

parameter	Nilai
p	2
d	1
q	5

IV.2.2 LSTM

Table 4.4 Matriks evaluasi data training

Matriks	**Nilai**
RMSE	6.347
MAE	3.594

Table 4.4 Matriks evaluasi data test

Matriks	**Nilai**
RMSE	4.728
MAE	2.024

No	Parameter	Nilai
1	n_input	2
2	n_nodes	100
3	n_epochs	20
4	n_batch	70
5	n_diff	0

IV.3 Analysis and Evaluation of test result

Based on the test and analysis results, the ARIMA model is superior in predicting substation loading compared to the LSTM model. The advantage of ARIMA is that it can still provide quite good results and is simpler in terms of interpretation and implementation.

However, it is important to note that the LSTM model lies in its ability to model non-linear relationships and capture long-term dependencies in the data, which is very important in the context of substation loading which is influenced by many complex factors. Therefore, the choice of the best model may depend on specific needs and resource availability.

IV.4 Comparison with previous research results

In research by Krishnamoorthy et al., 2024, the ARIMA and LSTM model cholesterol and glucose predictions showed that the ARIMA model had better performance in predicting compared to the LSTM model, as seen from the lower RMSE value in the ARIMA model. In other research, namely research on ARIMA vs LSTM on NASDAQ stock exchange data, it shows that the ARIMA model has a lower RMSE value that is three times better than LSTM.

From previous research and research conducted in this case study, it was found that the ARIMA model produces lower

RMSE and MAE values than LSTM, so this shows that ARIMA is considered to have better performance than LSTM.

V. CONCLUSION

Based on a series of experiments carried out in this research, several conclusions were obtained that ARIMA models proposed in this research produces forecast with better accuracy than LSTM.

ACKNOWLEDGMENT *(Heading 5)*

We would like to extend our heartfelt appreciation to PLN Ruky sandra ary murti () for their invaluable support throughout our research on Electricity load Maintenance for their Electrical Substation Components. We are particularly grateful for PT PLN in providing us the datasets, such as transformer's data, feeder's data, and their power outage data, so that we can perform the development of the machine learning model in this research. The success of this machine learning model is the reflection of our partnership with PT PLN. With hope that this machine learning model proves beneficial to PT PLN's endeavors, we express our gratitude for their ongoing support.

REFERENCES

[1] Géron, A. (2019): Hands-on Machine Learning whith Scikit-Learing, Keras and Tensorfow (2nd ed.), (Nicole Tache, Ed.)O'Reilly Media, Inc, O'Reilly Media, Inc., United states of America, 510.

[2] Kobiela, D., Krefta, D., Król, W., and Weichbroth, P. (2022): ARIMA vs LSTM on NASDAQ stock exchange data, Procedia Computer Science, 207(Kes), 3830–3839. https://doi.org/10.1016/j.procs.2022.09.445I.S. Jacobs and C.P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G.T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271-350.

[3] Krishnamoorthy, U., Karthika, V., Mathumitha, M. K., Panchal, H., Jatti, V. K. S., and Kumar, A. (2024): Learned prediction of cholesterol and glucose using ARIMA and LSTM models – A comparison, Results in Control and Optimization, 14(December 2023), 100362. https://doi.org/10.1016/j.rico.2023.100362R. Nicole, "Title of paper with only first word capitalized," J. Name Stand. Abbrev., in press.

[4] PT. PLN (Persero) (2009): Teleinformasi data untuk pemeliharaan instalasi sistem tenaga listrik, (169).Transl. J. Magn. Japan, vol. 2, pp. 740-741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[5] PT. PLN (Persero) (2020): Pemeliharaan Kubikel Tegangan Menengah, Buku Pedoman Pemeliharaan, 1, 7–8.

[6] SADEQ, A. (2008): GABUNGAN DENGAN METODE ARIMA (Studi pada IHSG di Bursa Efek Jakarta), Universitas Stuttgart.

[7] Utama, R. E., Gani, N. A., Jaharuddin, and Priharta, A. (2018): Manajemen Operasi dan Rantai Pemasok Edisi Revisi (1st ed.), UM Jakarta Press, Jakarta, 326.

[8] H. Ponto. (2018): Dasar Tenik Listrik. Deepublish. Yogyakarta

[9] Setiawan, I. N., & Setiawan, W. (2013): Perbandingan Peramalan Beban Listrik Jangka Pendek Menggunakan Support Vector Machine dan Jaringan Syaraf Tiruan Perambatan Balik. Universitas Udayana. Bali.

[10]

PERNYATAAN

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 12 Juni 2024

Satria Dina Astari dan 23522309