Forecasting Electricity Consumption of Advanced Meter Infrastructure (AMI) PT PLN (Persero) Customers using LSTM and ARIMA

Muhammad Zulfi Ashari – NIM. 23522304
Program Studi Magister Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jalan Ganesha 10 Bandung
23522304@std.stei.itb.ac.id

***Abstract*— PT Perusahaan Listrik Negara (Persero) or PLN as one of the State-Owned Enterprises (BUMN) engaged in Public Utilities, especially the electricity sector, has the duty and responsibility to provide reliable electricity for all Indonesian people. In this digitalization era, PLN participates in utilizing the latest technological developments in providing electricity to customers' homes. One of the technologies implemented is Smart Grid. Since 2023, PLN has initiated a Smart Grid implementation program, one of which is the implementation of Smart Meter Advanced Meter Infrastructure (AMI) in several regions of Indonesia, one of which is in the working area of the Unit Pelaksana Pelayanan Pelanggan (UP3) Cikokol. One of the benefits of implementing Smart Meter AMI is that PLN can find out customer electricity usage in real-time with more precise data and more diverse attributes compared to conventional recording methods. The electricity usage data can be processed and utilized for various purposes, one of which is to predict electricity usage in improving electricity distribution management and relatively better planning for electricity infrastructure provision in a region. In this research, we explore the electricity usage data from one of the Data Concentrator Unit (DCU) at the Distribution Substation which contains time series data of electricity usage from several Smart Meters in 1 phase customer houses connected to the DCU and develop a prediction model using the Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA) algorithms. Each model of LSTM and ARIMA was developed with various experiments and compared its performance using Root Mean Square Error (RMSE) and Root Mean Absolute Percentage Error (MAPE) evaluation metrics. At the end, it was found that the ARIMA model had better performance results compared to LSTM in this study with an RMSE value of 13,297 and MAPE of 0,052%.**

Keywords*— *Electricity Consumption Forecasting, Advanced Meter Infrastructure (AMI), Machine Learning, LSTM, ARIMA

I. INTRODUCTION

PT Perusahaan Listrik Negara (Persero) or PLN as one of the State-Owned Enterprises (SOEs) engaged in Public Utilities, especially the electricity sector, has the duty and responsibility to provide reliable electricity for all Indonesian people. In this digitalization era, PLN participates in utilizing technological developments in providing electricity to customers' homes. One of the technologies implemented is Smart Grid including Smart Meter Advanced Meter Infrastructure (AMI) as part of the technology system. The Government of Indonesia has planned Smart Grid implementation activities including Smart Meter AMI as a National Strategic Program [1]

Over time, each region in Indonesia has the potential to experience increased economic growth, the majority of which has a positive impact on increasing electricity consumption in the region. Taking this into account, since 2023 PLN has committed to initiating the AMI Smart Meter implementation program in several regions, one of which is the work area of the Unit Pelaksana Pelayanan Pelanggan (UP3) Cikokol which is a work unit under the Unit Induk Distribusi (UID) Banten. By 2023, PLN UP3 Cikokol has installed 143,000 AMI Smart Meters and 512 Data Concentrator Units (DCU) [2].

One of the benefits of implementing Smart Meter AMI is that PLN can find out customer electricity consumption in real-time with more precise data and diverse attributes compared to conventional methods [3] even though the electricity consumption behavior of each customer may differ according to time, location and the environment around the customer's location [4]. The electricity consumption data can be processed and utilized for various purposes, one of which is to predict electricity consumption in an effort to improve electricity distribution management and relatively better planning for electricity infrastructure provision in a region [5].

With the start of the Smart Meter AMI implementation program, it is expected that PLN UP3 Cikokol can utilize the data obtained from Smart Meter AMI to determine the pattern of customer electricity consumption and predict customer electricity consumption in its area so that it is expected that the demand and supply of electricity distribution can be relatively more balanced and of course can improve customer service, especially related to electricity distribution management.

II. LITERATURE REVIEW

A. *Advanced Meter Infrastructure* (AMI)

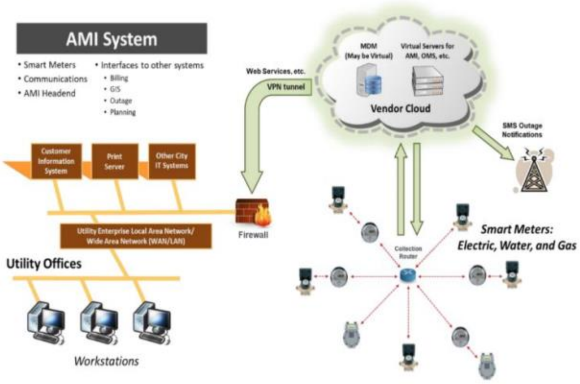

Fig. 1 AMI System Overview

AMI systems are designed to manage energy distribution and utilization by measuring, collecting, analyzing, and controlling data through automated devices. These devices can communicate with various media as needed or at certain time intervals. With the AMI smart meter system, PLN and customers can monitor and analyze electricity consumption based on recorded data. Therefore, AMI is a modern technology that uses smart meters through meter reader media to automatically monitor and analyze customer electricity consumption data [6].

AMI is a system that combines the latest measurement technology, communication networks, and data management systems to measure, collect, and analyze electricity consumption in real-time. In AMI, there are several technologies and applications that are integrated with each other. The following are the components of AMI:

TABLE I AMI Supporting Components

Component	Description
Smart Meter	Digital devices known as smart meters can measure energy consumption at frequent intervals, such as every 15 minutes. In addition, smart meters offers detailed data on usage patterns, voltage quality, and power outage notifications/ Unlike traditional mechanical meters, smart meters are also capable of supporting bidirectional communication.
Communication Networks	AMI depends on communication networks to transmit data from smart meters to utility. The type of network used depends on the infrastructure and coverage requirements and can be either wired such as power line communication (PLC), Ethernet, or wireless such as cellular or radio frequency (RF)
Meter Data Management System (MDMS)	Data management systems are utilized by AMI systems to gather, store, handle, and analyze the amounts of data produced by smart meters. Meters communicate with the utilities to manage energy consumption, recognize patterns, and provide precise billing details.

B. *Machine Learning for Predictive Data Analysis*

Modern companies or organizations collect large amounts of data [7]. To derive value from the data, analysis must be done to extract knowledge that can be used for better decision making. The process of data processing from raw data to knowledge is illustrated in Figure 2

Fig. 2 Illustration of Data Processing into Knowledge

Predictive Data Analytics is an art of developing and using models to predict a set of values based on patterns drawn from historical data [7].
Machine Learning (ML) is defined as an automated process that extracts patterns from data [7]. To develop a model used to predict data, researchers will use Supervised ML. This Supervised ML technique will learn the relationship between descriptive features and target features based on historical data.

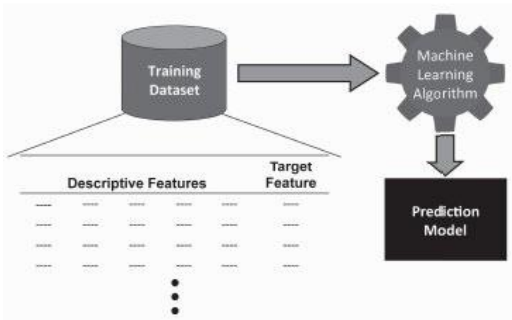

Fig. 3 Model Learning from Historical Data

C. *Long Short-Term Memory (LSTM)*

Long Short-Term Memory (LSTM) LSTM is an extension of the Recurrent Neural Network (RNN) architecture that has strong forecasting capabilities over time series data. The main difference between RNN and LSTM is that LSTM can store long-term time-dependent information and can map reliably

between input and output data [8] The LSTM structure consists of forget gate, input gate, and output gate according to Figure II.4. Forget gate is responsible for determining the amount from the previous memory to be conveyed to the current LSTM unit. The input gate will update the state of the memory cell based on information from the current input and the previous hidden state. Output gate selectively controls the output
The operation of the three gates and memory cells at a given time-step p is as follows:

$$f_p = \sigma(W_p X_p + V_f h_{p-1} + b_f) \quad (1)$$
$$C'_p = tanh(W_c c + V_c h_{p-1} + b_f) \quad (2)$$
$$i_p = \sigma(W_i x_p + + V_c h_{p-1} + b_f) \quad (3)$$
$$C_p = C_{p-1} . f_p + i_p . C'_p \quad (4)$$
$$O_p = \sigma\left(W_o x_p + + V_o h_{p-1} + b_o\right) \quad (5)$$
$$h_p = O_p . \tanh(C_p) \quad (6)$$

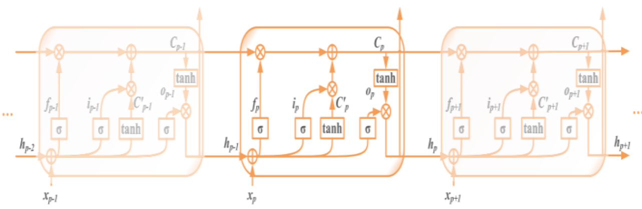

Fig. 4 Structure of LSTM

Where:

- V_f, V_c, V_i, V_o refers to the recurrent weights of the LSTM
- W_c, W_i, W_o denotes the weight matrix of forget gate, memory cell, input gate, output gate in order
- b_f, b_c, b_i denotes the weight matrix of forget gate, memory cell, input gate, output gate in order
- σ is a sigmoid activation function whose output ranges from 0 to 1, where 0 represents the information that is completely forgotten and 1 represents the information that is completely forgotten.

D. *Autoregressive Integrated Movinf Average (ARIMA)*

Autoregressive Integrated Moving Average (ARIMA) is a statistical method used to analyze and forecast time series data, generalizing the Autoregressive Moving Average (ARMA) model [9]. ARIMA is applied to non-stationary data with a focus on removing trends and seasonal components to better understand or forecast a time series. The ARIMA model consists of 3 main parts, including Autoregressive (AR), Integrated (I), and Moving Average (MA), each part handling a specific aspect of the data to obtain an optimal fit.
The ARIMA model is denoted by the parameters (p,d,q) where:

- p represents the lag order, i.e. the number of lag observations incorporated into the model
- d denotes the degree of divergence, indicating the number of times the observations are diverged to make the time series stationary
- q represents the order of the moving average window, which means the window size of the moving average.

The equation used in the ARIMA model is as follows:

$$y'_t = I + \alpha_1 y'_{t-1} + \alpha_2 y'_{t-2} + \cdots + \alpha_p y'_{t-p} + e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} \quad (7)$$

Where:

- y'_t is the time series value at time t
- I is the average of the time series
- α is the autoregressive parameter
- e is the error at time tt
- θ is the moving average parameter

ARIMA models are used to predict future values of a time series based on past observations and the relationship between them. It is particularly useful for forecasting when the underlying process that generates the observations is an ARIMA process [10].

E. *Evaluation Metrics*

In measuring the performance of LSTM and ARIMA models, there are several evaluation metrics that can be used, as follows:

1) *Root Mean Square Error (RMSE)*

RMSE is used to measure the difference between predicted and actual values Lower RMSE values indicate better performance. The following is the equation form of RMSE:

$$\sqrt{\frac{1}{2}\sum_{i=1}^{n}(EC_{predicted} - EC_{actual})^2} \quad (8)$$

Where:

EC is energy consumption

2) *Mean Absolute Percentage Error (MAPE)*

MAPE measures the average percentage of difference between predicted and actual values. A lower MAPE value indicates better performance. Here is the equation form of MAPE:

$$\frac{1}{2}\sum_{t=1}^{n}\left|\frac{y_t - y*_t}{y_t}\right| x\ 100\% \quad (9)$$

Where:
y_t = actual value in period t
$y*_t$ = predicted value in period t

F. *Relevant Research*

In expanding knowledge related to the topic and research methods being studied, researchers reviewed several previous studies that studied the use of Machine Learning algorithms in predicting electricity consumption in a region.

Samir M. Shariff (2022) conducted research related to the use of ARIMA and LSTM algorithms in predicting electricity consumption in Colombia based on electricity consumption data from a Regional Transmission Organization in Colombia called PJM Interconnection LLC. The available data is in the form of hourly electricity consumption data with units of Megawatts (MW). Based on the exploration results, it is concluded that the performance of the LSTM model is better than ARIMA as measured using Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Normalized Mean Squared Error (NMSE), and Normalized-Root Mean Squared Error (NRMSE) [11]

Alhussein et al. (2020) conducted research on time-series data on electricity consumption of Household customers (RT) as many as 10,000 customers from Smart Grid Smart City (SGSC) funded by the Australian Government obtained information that time-series data on electricity consumption of Household customers (RT) can be done with a hybrid method approach, namely by combining Convulational Neural Network (CNN) and LSTM. CNN has the ability to learn the representation of the time series data and learn the attributes that can have a great influence in predicting customer electricity consumption. At the end of the study, it was stated that the use of the CNN and LSTM combination model resulted in relatively better accuracy performance based on the Mean Absolute Percent Error (MAPE) measurement compared to using only the LSTM algorithm [12].

Chinnaraji et al (2022) conducted research on electricity consumption/load time series data from the Supervisory Control and Data Acquisition (SCADA) system obtained from the University of California Irvine (UCI) Machine Learning Repository. Chinnaraji et al. explored the use of the LSTM algorithm by optimizing the hyper-parameters model and the number of hidden layers to improve the performance of the LSTM model. One of the drawbacks of implementing the LSTM model (default) is that there are still some local optimum problems, such as sparse training data. Therefore, it is necessary to perform hyper-parameter tuning in the data training process. At the end of the study, the exploration results of the improved LSTM model produced better performance compared to the LSTM model without adjusting hyper-parameters and hidden layers based on MSE, MAE, and RMSE measurements [13]

Torres et al. (2022) conducted research on electricity consumption time series data in Spain from 2007 to 2016 with a data time interval of 10 minutes. Torres et al. explored the use of LSTM algorithm including hyper-parameters optimization using random-search strategy developed using Keras-Tuners framework. At the end of the study, it was found that the error rate of the proposed model was less than 1.5% and had the lowest error rate compared to Linear Regression, Decision Tree, Ensemble Tree, and Deep Neural Networks [14]).

Elsaraiti et al. (2021) conducted research on electricity consumption time series data in South Tripoli from the National Electricity Corporation in Libya with a time span of electricity consumption data from January 1, 2016 to March 24, 2016 (12 weeks). Elsaraiti et al. explored the use of seasonal ARIMA algorithms to predict electricity consumption in the future. At the end of the study, information was obtained that the ARIMA (6,1,0)(1,1,6) model was the best model that could predict the value of electricity consumption for the next two weeks with the lowest MAPE error measurement, namely 4.332% [15].

Based on the literature review conducted, the researcher seeks to make an additional contribution to the use of the LSTM and ARIMA algorithms on electricity consumption data from one of the substations at PLN UP3 Cikokol by optimizing the existing hyper-parameters, then comparing the performance of the models that have been developed to get the best model.

III. Methodology

A. Dataset

In this study, the research object used is electricity consumption data from 47 PLN UP3 Cikokol customers who have used AMI Smart Meters. The electricity consumption data is obtained from the Data Concentrator Unit (DCU) installed at each Distribution Substation through the MDMS application. The data is in the form of time series data with a time interval of 15 minutes in the period January 1 to April 30, 2024 with a total of 110,087 rows of data and a total of 45 columns of data. Figure 5 is an example of data from the measurement of electricity consumption by Smart Meter obtained from MDMS application.

U_ID	UP3	UP3	COMM_ADDR	DATA_TIME	II_ACTI	IE_ACT	II_REA(	IE_REA	VOL_R	VOL_S	VOL_T	CUR_R	CUR_S	CUR_T	PF	IMP_AC_P	METER_ID
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	64,00	0,00	0,00	18,00	231,10			1,60			0,95	258,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	262,00	0,00	0,00	34,00	231,80			4,67			0,99	1047,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	604,00	0,00	59,00	0,00	220,10			11,52			0,99	2415,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	130,00	0,00	30,00	0,00	205,70			2,62			0,97	521,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	307,00	0,00	89,00	0,00	211,50			6,07			0,96	1226,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	220,00	0,00	62,00	0,00	225,80			4,07			0,96	879,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	221,00	0,00	11,00	7,00	196,00			4,66			0,99	885,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	32,00	0,00	49,00	0,00	223,70			1,06			0,55	130,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	80,00	0,00	0,00	18,00	234,30			1,75			0,96	319,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	28,00	0,00	25,00	0,00	221,60			0,71			0,75	113,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	70,00	0,00	7,00	0,00	225,00			1,39			0,99	280,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	0,00	0,00	0,00	0,00	214,60			0,00			1,00	0,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	18,00	0,00	8,00	0,00	221,70			0,38			0,93	72,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	371,00	0,00	105,00	0,00	221,60			7,04			0,96	1486,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	182,00	0,00	0,00	3,00	226,20			3,39			1,00	730,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	286,00	0,00	47,00	0,00	213,40			5,48			0,99	1143,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	113,00	0,00	11,00	0,00	205,00			2,26			1,00	450,00	246XXXXXXXX
BANTEN	CIKOKOL	CIKOKOL	246XXXXXXXX	4/29/2024 11:45:00 PM	30,00	0,00	21,00	0,00	217,70			0,69			0,82	123,00	246XXXXXXXX

Fig. 5 Example of Electricity Consumption Measurement Data from MDMS

B. Soution Design

Figure 6 describes the process of developing ML models using the LSTM and ARIMA algorithms to obtain the best performing models and shows a comparison of actual electricity consumption data and predicted results in the form of visualizations.

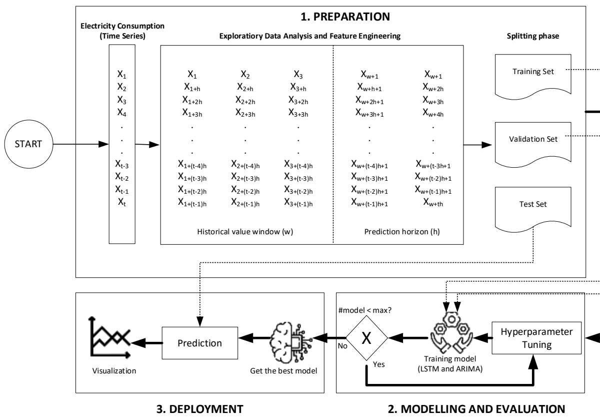

Fig. 6 Proposed Solution Design

IV. Analysis and Result System Testing

A. Testing Scenario

This test scenario is used to evaluate the performance of the training results model based on the LSTM and ARIMA algorithms in predicting electricity consumption based on the average electricity consumption of 47 1-phase customer homes that have installed Smart Meter AMI at PLN UP3 Cikokol.

The data used in this test is electricity consumption data for the period January 1 to April 30, 2024 with a time interval of 15 minutes. The data is divided into 2 division groups with different composition combinations for training data, validation data, and test data with details, as follows:

1) Division group I:

a. Training data (80%): 2024-01-01 00:00:00 to 2024-04-06 19:00:00 or 9.290 rows of data

b. Validation data (10%): 2024-04-06 19:15:00 to 2024-04-18 21:15:00 or 1.161 rows of data

c. Test data (10%): 2024-04-18 21:30:00 to 2024-04-30 23:45:00 or 1.162 rows of data

2) Division group II:

a. Training data (70%): 2024-01-01 00:00:00 to 2024-03-25 16:45:00 or 8.129 rows of data

b. Validation data (15%): 2024-03-25 17:00:00 to 2024-04-12 20:00:00 or 1.741 rows of data

c. Test data (15%): 2024-04-12 20:15 to 2024-04-29 23:45:00 or 1.743 rows of data

B. Model Experiment

1) LSTM

In developing a model with the LSTM algorithm, researcher determine 5 parameters that will potentially affect the performance of the model, including input, epochs, batch size, window size, and timestep. Researcher use several parameter values that will be used in developing the model. The list of parameter values can be seen in table 2 below:

TABLE II List of parameter values for LSTM modelling

Exp.	***Input***	***Epochs***	***Batch Size***	***Window Size***	***Timestep***
1	10	20	10	10	1
2	12	30	12	12	1
3	15	50	32	32	1
4	50	50	32	32	5
5	50	150	128	128	5

In addition, researchers also applied dropout with a value of 0.25, Relu optimizer, and Adam activation to the model setting with this LSTM algorithm.

2) ARIMA

In developing a model with the ARIMA algorithm, there are 3 parameters that will potentially affect the performance of the model, including p, d, and q where each parameter has been explained previously in Chapter II. Researchers use several parameter values that will be used in developing the model. The list of parameter values can be seen in table 3 below:

TABLE III List of parameter values for ARIMA Modelling

Exp.	***P***	***D***	***Q***
1	0	1	6
2	1	1	6
3	2	1	5
4	5	1	0
5	6	1	0

C. Test Result

The following are the results of the performance comparison of the LSTM and ARIMA models with the best parameters based on the experimental and testing results in section IV.2:

TABLE IV Comparison Results of This Research and Previous Research

Model	**Best Parameter**	**Division Group**	**RMSE**	**MAPE**
LSTM	*input: 15, epochs: 50, batch_size; 32, window_size: 32, timestep: 1*	I	17,12	0,11%
ARIMA	*p: 5, d: 1, q: 0*	I	13,297	0,052%

Figures IV.1 and IV.2 show the visualization of the comparison of actual and predicted electricity consumption values from the LSTM and ARIMA models based on test data in division group I.

	Date	Actual	LSTM_Predicted	ARIMA_Predicted
0	2024-04-19 05:30:00	163.17	186.289993	197.76
1	2024-04-19 05:45:00	132.17	162.970001	163.10
2	2024-04-19 06:00:00	119.48	142.960007	132.58
3	2024-04-19 06:15:00	110.70	133.309998	119.11
4	2024-04-19 06:30:00	99.04	126.180000	111.68
...	...	...	...	...
1125	2024-04-30 22:45:00	210.33	197.029999	203.51
1126	2024-04-30 23:00:00	215.62	199.470001	210.74
1127	2024-04-30 23:15:00	198.56	204.169998	214.66
1128	2024-04-30 23:30:00	192.20	193.250000	199.41
1129	2024-04-30 23:45:00	189.78	185.710007	192.55

1130 rows × 4 columns

Fig. 7 Dataframe view of actual and predicted values of LSTM and ARIMA models

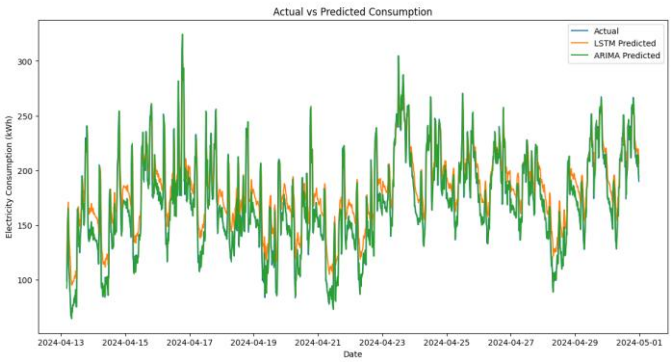

Fig. 8 Visualization of the comparison of actual and predicted values from the LSTM and ARIMA models

Based on several experiments and tests that have been carried out, it can be concluded that the ARIMA model has better performance and is relatively stable compared to the LSTM model despite changing parameters and dividing different data in each experiment, this can be caused because the electricity consumption data used is stationary or tends to remain. Based on the Augmented Dickey-Fuller (ADF) test (Mushtaq, 2011) conducted on the overall electricity consumption data, it is found that the ADF statistical value is below the critical line (1%, 5%, 10%) according to Figure IV.3 and the p value of 3.3357183958899614e-30 is smaller than the significance value of 0.05 so it can be concluded that the data studied is stationary so that it is in line with the performance results of the ARIMA model that have been submitted previously.

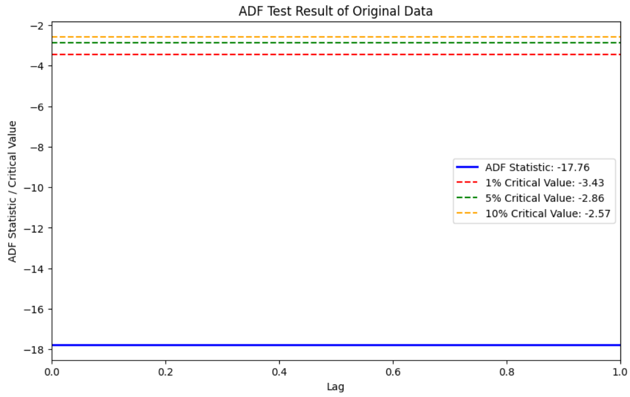

Figure 1. Result of ADF Test to Original Data

On the other hand, the LSTM model has lower performance because it is more suitable for non-stationary data and requires a relatively larger amount of training data. It is evident in the previous experiment that the RMSE and MAPE values in division group I with 80% training data are better than those in division group II with 70% training data.

D. Comparison with Recent / Previous Reseach Reults

researchers conducted 5 experiments for the LSTM and ARIMA models using several parameter values that refer to previous related research [11] [15]. The following are the results of the comparison of model performance in this study and previous related research:

TABLE V Comparison Results of This Research and Previous Research

	This Research		**Shariff, 2022**		**Elsaraiti dkk., 2021**	
	I	II	I	II	I	II
Best Para meter	*(input: 15, epochs: 50,* Parame ter *batch_s ize; 32, window _size: 32, timeste p: 1)*	*(p: 5, d: 1, q: 0)*	*(input unit:* 12, *batch size:* 12, *epoch s:* 30)	*(p: 2, d: 1, q: 5)*	[redacted]	*(p: 6, d: 1, q: 0)*
RMS E	17,12	13,297	38,05	248,67	[redacted]	-
MAP E	0,11%	0,052%	-	-	[redacted]	4,332%

Where:
I is LSTM model
II is ARIMA model

In Shariff's study, the LSTM model (input unit: 12, batch size: 12, epochs: 30) gave better performance results compared to the ARIMA (2,1,5) model in predicting customers. However, in this study, the ARIMA (5,1,0) model had better results than the LSTM model (input: 15, epochs: 50, batch_size; 32, window_size: 32, timestep: 1) which can be seen in the RMSE and MAPE values obtained in section IV.4, although the parameter value (2,1,5) obtained the best RMSE and MAPE performance in the experiment with the II division group. In the study by Elsaraiti et al., the SARIMA model (p: 6, d: 1, q: 0 and p: 1, d: 1, q: 6) gave the best performance results, but in this study these parameters still produced a higher MAPE so that the best parameters for the ARIMA model were obtained, namely p: 5, d: 1, q: 0. Based on the researcher's analysis, this difference is very possible because there are differences in data, model building methods, and the composition of the training-validation-test data division.

V. CONCLUSION

In this study, an ML model was developed to predict electricity consumption in an area at PLN UP3 Cikokol. The model was developed using the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, with the algorithms used being Long Short-Term Memory (LSTM) and Autoregressive Integrated Moving Average (ARIMA). Based on the results of training and testing using average electricity consumption data from 47 1-phase customers who have installed AMI Smart Meters, the ARIMA model was obtained which had the best performance with the lowest RMSE and MAPE values compared to the LSTM model with various parameter configurations that had been carried out. The optimal parameters for applying the LSTM algorithm to electricity consumption data are input: 15, epochs: 50, batch_size; 32, window_size: 32, timestep: 1, while for the ARIMA algorithm, ‘p’ is worth 5, ‘d’ is worth 1, ‘q’ is worth 0. 3. Based on the test results, a larger training data composition or 80% tends to provide better model performance results compared to training data of 70%.

There are some suggestions for future research. Firstly, increase the amount of data that can be used in training to improve the performance of LSTM and ARIMA models. Secondly, adding external data that may affect customer electricity usage, such as weather data or data on other factors that can affect customer electricity usage. Thirdly, Combining the LSTM and ARIMA algorithms or with other algorithms to improve the performance of the resulting models. The last one, add additional evaluation metrics, such as Normalized Mean Square Error (NMSE), Mean Absolute Error (MAE), and R-Squared to get a more comprehensive picture of model performance.

REFERENCES

[1] Pemerintah Indonesia, “Peraturan Presiden No. 18 Tahun 2020.” 2020. Diakses: 23 April 2024. [Daring]. Tersedia pada: https://peraturan.bpk.go.id/Details/131386/perpres-no-18-tahun-2020

[2] PT PLN (Persero) UID Banten, “Berita Acara Pemadanan Data Implementasi AMI Tahun 2023.” 2023.

[3] T. K. Sai Pandraju, S. Samal, S. R., S. M. Yaseen, R. Nandal, dan D. Dhabliya, “Advanced metering infrastructure for low voltage distribution system in smart grid based monitoring applications,” *Sustainable Computing: Informatics and Systems*, vol. 35, hlm. 100691, Sep 2022, doi: 10.1016/j.suscom.2022.100691.

[4] F. M. Andersen, P. A. Gunkel, H. K. Jacobsen, dan L. Kitzing, “Residential electricity consumption and household characteristics: An econometric analysis of Danish smart-meter data,” *Energy Economics*, vol. 100, hlm. 105341, Agu 2021, doi: 10.1016/j.eneco.2021.105341.

[5] D. Ignatiadis, G. Henri, dan R. Rajagopal, “Forecasting Residential Monthly Electricity Consumption using Smart Meter Data,” dalam *2019 North American Power Symposium (NAPS)*, Wichita, KS, USA: IEEE, Okt 2019, hlm. 1–6. doi: 10.1109/NAPS46351.2019.9000285.

[6] E. Ko dan G. S. Jeong, “Analysis of Standards for Electricity Advanced Metering Infrastructure in the Housing Sector,” dalam *2023 IEEE Green Energy and Smart Systems Conference (IGESSC)*, Long Beach, CA, USA: IEEE, Nov 2023, hlm. 1–6. doi: 10.1109/IGESSC59090.2023.10321754.

[7] J. D. Kelleher, B. M. Namee, dan A. D’Arcy, *Fundamentals of Machine Learning for Predictive Data Analytics*, 3 ed. 2015.

[8] H. Zang *dkk.*, “Residential load forecasting based on LSTM fusing self-attention mechanism with pooling,” *Energy*, vol. 229, hlm. 120682, Agu 2021, doi: 10.1016/j.energy.2021.120682.

[9] “Introduction to ARIMA: nonseasonal models.” Diakses: 9 Maret 2024. [Daring]. Tersedia pada: https://people.duke.edu/~rnau/411arim.htm

[10] R. J. Hyndman dan G. Athanasopoulos, *Forecasting: Principles and Practice*, 2nd ed. 2016. Diakses: 9 Maret 2024. [Daring]. Tersedia pada: https://otexts.com/fpp2/

[11] S. M. Shariff, “Autoregressive Integrated Moving Average (ARIMA) and Long Short-Term Memory (LSTM) Network Models for Forecasting Energy Consumptions,” *EJECE*, vol. 6, no. 3, hlm. 7–10, Mei 2022, doi: 10.24018/ejece.2022.6.3.435.

[12] M. Alhussein, K. Aurangzeb, dan S. I. Haider, “Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting,” *IEEE Access*, vol. 8, hlm. 180544–180557, 2020, doi: 10.1109/ACCESS.2020.3028281.

[13] R. Chinnaraji dan P. Ragupathy, “Accurate electricity consumption prediction using enhanced long short-term memory,” 2022, doi: https://doi.org/10.1049/cmu2.12384.

[14] J. F. Torres, F. Martínez-Álvarez, dan A. Troncoso, “A deep LSTM network for the Spanish electricity consumption forecasting,” *Neural Comput & Applic*, vol. 34, no. 13, hlm. 10533–10545, Jul 2022, doi: 10.1007/s00521-021-06773-2.

[15] M. Elsaraiti, G. Ali, H. Musbah, A. Merabet, dan T. Little, “Time Series Analysis of Electricity Consumption

Forecasting Using ARIMA Model," dalam *2021 IEEE Green Technologies Conference (GreenTech)*, Denver, CO, USA: IEEE, Apr 2021, hlm. 259–262. doi: 10.1109/GreenTech48523.2021.00049.

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, 8 Juni 2024
Ttd
Muhammad Zulfi Ashari – NIM. 23522304