Prediction System on Electricity Consumption using Web-Based LSTM Algorithm

Fathoni Waseso Jati[1], Komang Jaya Bhaskara Mahatya[1], Faisal Candrasyah H[1], Budhi Irawan[1]
[1]Department of Computer Engineering, School of Electrical Engineering, Telkom University, Indonesia

Article Info

Article history:

Received July 31, 2022
Revised August 13, 2022
Accepted August 14, 2022

Keywords:

LSTM
Web Application
Prediction Electricity Usage
Mean Square Error
Time Series

ABSTRACT

Over the year, technology development is growing rapidly, especially electronics devices such as notebooks and smartphones. Rapid technology development, affect lifestyle habits which lead to an increased electrical energy usage. Monitoring the electricity meter in real time encourage user to use their electricity more responsibly, thus improving the efficiency of energy management. This study aims to create a web-based electric usage prediction system which helps user to monitor their electricity usage. Development begins by collecting and cleaning the electrical energy consumption data. The clean data is used in Long-Short Term Memory (LSTM) model development which designed to be able to predict electrical energy usage. The development continued by developing web application as an interface for the user to interact with the monitoring system. Test results shows that the LSTM model can predict electric usage with a Loss Mean Square Error (MSE) value of 0.0071. In addition, the alpha and beta testing of the website shows an accuracy of 100% and 82.64% respectively.

Corresponding Author:

Fathoni Waseso Jati
Department of Computer Engineering
School of Electrical Engineering, Telkom University
Bandung, Indonesia
Email: fathoniwasesojati@student.telkomuniversity.ac.id

1. INTRODUCTION

In the last few years, technological developments are increasing, especially in electronic devices such as smartphones and notebooks. The impact of rapid development of technology, can lead to changes in lifestyle of each. With changes in lifestyle, it will also have an impact on increasing the use of electrical energy. In addition, the negligence of electricity meter users in managing the use of electrical energy that is spent every day can also affect the increase in electrical energy. Several factors influence the increase in electrical energy, namely human factors (80%) and other technical factors (20%) [1]. Quoted from the official website of the Central Statistics Agency (BPS), it is known that electricity consumption in 2017 rose by around 1.02 mWh/capita, then in 2018 it rose by around 1.06 mWh/capita, in 2019 it rose again to reach 1.08 mWh/capita[2]. If the increase in electrical energy continues to be ignored, it can cause a global shortage of electrical energy in a short time.

Technological developments today, allows several approaches that can solve problems such as predictive cases and take the best decisions, namely Machine Learning[3]. In addition, there is another approach that is more advance and more reliable than Machine Learning, namely Deep Learning. In the Deep Learning approach, several algorithms have their respective goals like prediction, classification, and clustering. There are lots of deep learning-based algorithms like CNN, RNN, LSTM, GRU, and DBNs have good performance to solve problems in prediction, clustering, and classification[4]. In research [5]–[8] in the case of predicting the use of electrical energy in the household environment, it is known that the LSTM model is better than other models such as Auto Regressive Integrated Moving Average (ARIMA). Therefore, this research uses the LSTM model to predict the use of electrical energy.

On paper [5] it is known that the results of electricity consumption get a loss value of 5.8267. In paper [6] it is known that the prediction results get a loss value of around 0.03 to 0.2. While paper [7] is known that the prediction results get a loss value of 0.35. Then on paper [8] it is known that the prediction results get a loss value of 3.99. of all the papers that have been mentioned, paper [6] is the one that produces the lowest loss value. However, the paper predicts electricity consumption in one day basis. This research will be improved by predicting electricity consumption in an hourly basis so that the predicted data is more accurate.

In addition to predicting the use of electrical energy, developing a calculation feature for how long the remaining electricity will run out and a website application as an interface for the calculation results. By designing a website application, it is easier for users to monitor and manage electrical energy compared to observing an electricity meter continuously. This study focuses on predicting the electric usage of a single electric meter in one household at Pesona Bali, Dayeuhkolot sub-district, Bandung Regency.

2. METHOD

The system design process consists of Electric Usage Prediction and Web Application processes connected to the database. The following is an overview of the electrical energy-based website usage system. The following is an overview of the web-based electrical energy consumption system.

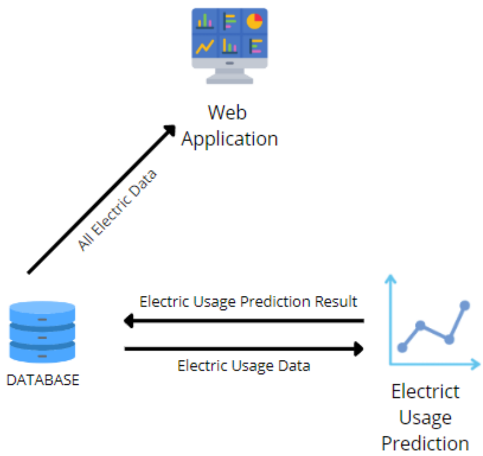

Figure 1. Design System.

The system design starts from the Electric Usage Prediction process by making predictions on electrical energy consumption data using the LSTM model. Electrical energy consumption data is retrieved from the database through the Application Programming Interface (API). The data have been obtained from the image recognition process of the remaining electricity that have converted into electrical energy consumption. After that, the predicted data is sent back to the database. Predicted data in the database, then displayed on a web application page. The process of designing Electric Usage Prediction and Web Application has its stages. These stages can be explained as follows.

2.1 Design Model LSTM

The process of designing the LSTM model is carried out in several stages the data collection process, the data preprocessing process, and the LSTM model-making process. This process is designed to create a model that can predict data on electrical energy consumption. The following are the stages of designing the LSTM model.

2.1.1 Data Collection

The dataset collection process begins with taking pictures of the remaining electricity from the kWh meter type ITRONACE900 IBS DS using a Raspberry PI 3 microcontroller and a Camera Module. Data collection was carried out in one of the Pesona Bali, Dayeuhkolot sub-district, Bandung Regency. The results of the process of taking pictures of the remaining electricity are in the form of an image file in .jpg format with the file name, namely the date when the picture was taken. From the total image dataset that have collected, there are 522 data. The following is the result of taking pictures every 1 hour for 22 days.

Figure 2. Data collection process.

After completing the shooting process, the next step is to calculate the electricity consumption using Excel by subtracting the current electricity balance with the electricity balance in the last one hours on the entire data, with a time of every one hour sequentially. The following is a formula for calculating the use of electrical energy.

$$y(t) = x(t) - x(t-1) \quad (1)$$

where:

$y(t)$ = electrical energy consumption in kWh

$x(t)$ = current electricity balance

$x(t-1)$ = electricity balance in the last one hour

The results of the calculation of overall electrical energy consumption and dataset labeling can be seen in the following figure.

	tanggal	pulsa_listrik	pemakaian_listrik
2	2022-01-29 20:02:34	163,04	
3	2022-01-29 21:02:43	162,37	0,67
4	2022-01-29 22:02:52	161,89	0,48
5	2022-01-29 23:03:01	161,42	0,47
6	2022-01-30 00:03:10	160,96	0,46
7	2022-01-30 01:03:18	160,53	0,43
8	2022-01-30 02:03:27	160,19	0,34
9	2022-01-30 03:03:36	159,87	0,32
10	2022-01-30 04:03:45	159,53	0,34
11	2022-01-30 05:03:53	159,23	0,3
12	2022-01-30 06:04:02	158,94	0,29
13	2022-01-30 07:04:11	158,65	0,29
14	2022-01-30 08:04:20	158,37	0,28
15	2022-01-30 09:04:28	157,92	0,45
16	2022-01-30 10:04:37	157,57	0,35
17	2022-01-30 11:04:46	157,14	0,43
18	2022-01-30 12:04:55	156,66	0,48
19	2022-01-30 13:05:04	156,22	0,44
20	2022-01-30 14:05:12	155,76	0,46
21	2022-01-30 15:05:21	155,28	0,48
22	2022-01-30 16:05:30	154,79	0,49
23	2022-01-30 17:05:39	154,3	0,49
24	2022-01-30 18:05:48	153,69	0,61
25	2022-01-30 19:05:56	152,9	0,79
26	2022-01-30 20:06:05	152,27	0,63
27	2022-01-30 21:06:14	151,78	0,49
28	2022-01-30 22:06:22	151,33	0,45

Figure 3. Electricity usage calculation result.

2.1.2 Preprocessing

Pre-processing is a technique to improve the quality of data from previously raw data into ready-to-use data[9]. This preprocessing process is carried out after the data collection process is complete. The dataset of electricity consumption is still raw or not ready for use. The preprocessing process is carried out in several stages so that the data have used for making LSTM models. The stages of the pre-processing process are the Data Transformation process and the Data Cleaning process. The following is a picture of the pre-processing process on the electrical energy consumption dataset.

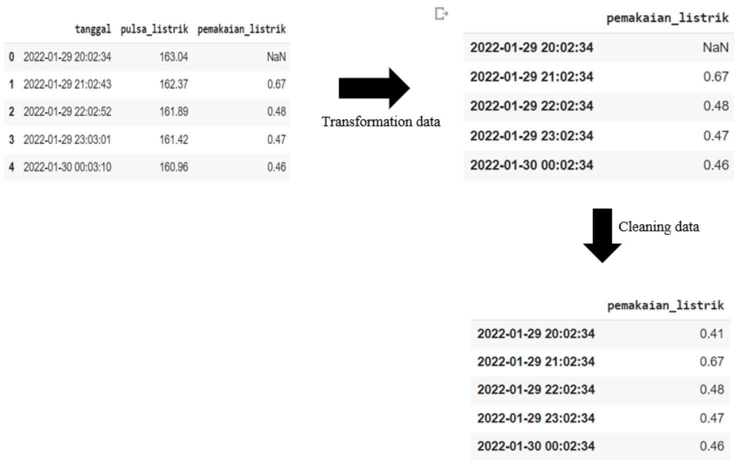

Figure 4. Process pre-processing

The pre-processing process in Figure 4 starts with the Data Transformation process, which is the process of changing data into a Time Series form, where the replacement of the value in the first column is in the appropriate date sequence. Then the Data Cleaning process is carried out, namely looking for the Missing Value and Outlier values in the electricity consumption column to be changed to the average value of the overall data and selecting the required column.

2.1.3 Modelling LSTM

The results of the dataset from the pre-processing process are then used in the process of making the LSTM model. There are several stages in the process of making the LSTM model. The initial stage is the Splitting Data process by splitting the data into two, namely data training and data validation. The data train is used to train the model, while data validation is used for the process of model improvement and model validation. Next is the Window Size process by divides the data into several groups. The Window Size process is used as input value and target value in LSTM modeling. The Window Size illustration process can be seen below.

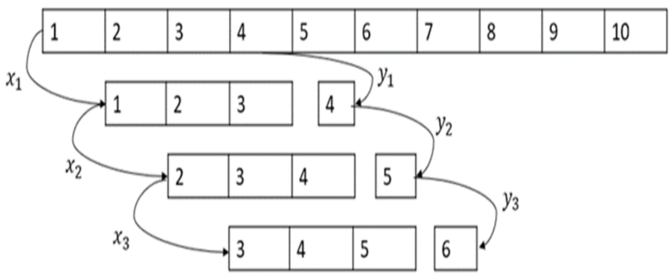

Figure 5. Ilustration process Window Size

Seen Figure 5 is an illustrative example of the Window Size process by dividing data such as x1, x2, and x3 by 3 window sizes to predict the target value of y1, y2, and y3. Furthermore, the process of initializing the parameters that affect the process of making the LSTM model is carried out. Initialization parameters used, namely:

1. Windows Size
2. Epoch
3. Batch Size
4. Units
5. Optimizer with Learning Rate
6. Activation Function.

In addition, there is an initialization of the dataset size that can affect the performance of the LSTM model. Furthermore, the LSTM model training process is used to form a pattern according to the train data and can predict the validation data well. The final stage, the model that has been carried out in the training process is then stored.

2.2 Design Web Application

The process of designing a website that is used as an interface for prediction results. This process is assisted by using a Laravel Framework and TailwindCSS. The Laravel framework is used to develop website applications with the concept of models, views, and controllers which can speed up and easier data integration on the backend[10]. While TailwindCSS is used to speed up the process of making a responsive UI website[11]. web application design process diagram can be seen as follows.

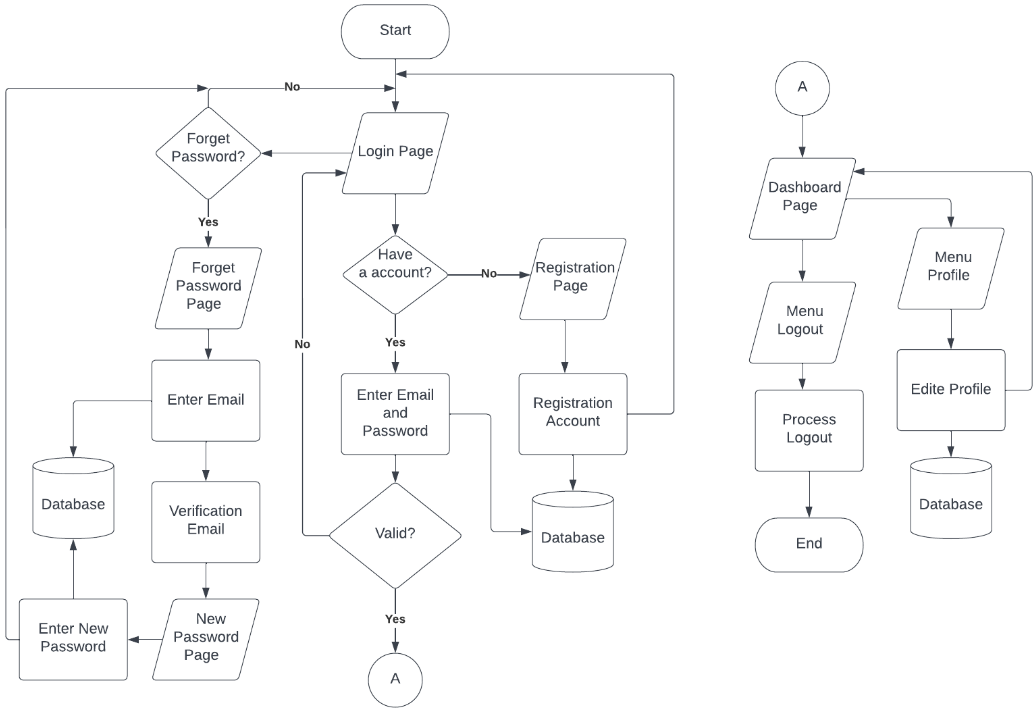

Figure 6. Flowchart design web application

The process of designing a web application starts by going to the login page. If the user does not have an account, they must register an account first. After registration, users can return to the login page to enter the dashboard page by entering their email and password. If the login process is valid, the user will be directed to the dashboard page. On the dashboard page, the user can see the prediction results for the next 1 day and how long the electricity will run out. The result of how long the electricity will run out is obtained from the division between the electricity usage with a total of 24 predicted data every 1 hour. Besides being able to monitor electricity data, users can also edit profiles such as changing passwords. To exit the dashboard page, the user can perform the logout process.

3. RESULTS AND DISCUSSION

After completing the LSTM model and website design process, the LSTM model testing process with Hyperparameter Tunning is carried out, and website testing with Alpha and Beta tests.

3.1 LSTM Hyperparameter Testing

The process of testing the hyperparameter tunning of the LSTM model is carried out to find the best model with the minimum loss value. Loss value calculation used an Error Measure, namely Mean Square Error (MSE). The following is the Mean Square Error formula for evaluating the LSTM model.

$$Mean\ Square\ Error = \frac{1}{n}\sum_{i=1}^{n}(y_j - \hat{y}_j)^2 \quad (1)$$

details:

n = total data observations

y_j = actual value

$\hat{y}_j$ = prediction value.

Before starting the test, doing initialization of the LSTM model parameters. Those parameters consist of Adam optimizer with Learning Rate 0.001, Units with value 128, Batch Size with size 1, activation function using ReLu, Epoch with value 50, and Window Size with size 1. The test scenarios used to get the best LSTM model are test dataset, Window Size, Epoch, and Learning Rate. The test starts from the dataset test, where 5 test scenarios are carry out with different dataset sizes and five times the same test in each dataset size. The dataset size values tested are 90:10, 80:20, 70:30, 60:40, and 50:50. The following are the best results for each of the size test data sets.

Table 1. Result of dataset testing

Testing	Data Train	Data Test	Mean Square Error
1st Test	90%	10%	0.0086
2st Test	80%	20%	0.0082
3st Test	70%	30%	0.0088
4st Test	60%	40%	0.0087
5st Test	50%	50%	0.0088

As seen in Table 1, the dataset test scenario with five different dataset size test processes got the best results on the dataset size of 80% train data and 20% test data. The best test results for the dataset get a minimum Loss Mean Square Error value of 0.0082. after that, the size value in the best dataset is used for the Window Size test. Window Size testing was doing in 5 test scenarios with different Window Sizes and five times the same test in each Window Size. For the tested Window Sizes, namely 1, 3, 5, 7, and 9. Here are the best results for each Window Size test.

Table 2. Result of Window Size testing

Testing	Dataset	Window Size	Mean Square Error
1st Test	80:20	1	0.0083
2st Test	80:20	3	0.0080
3st Test	80:20	5	0.0071
4st Test	80:20	7	0.0072
5st Test	80:20	9	0.0073

As seen in Table 2, the Window Size test scenario with five different Window Size tests got the best results at Window Size 5 with a minimum Loss Mean Square Error value of 0.0071. after that, use the size value on the best Window Size for the Epoch test. The Epoch test was carried out with the same test scenario as the Dataset and Windows Size testing. The Epoch values tested are 10, 50, 100, 150, and 200. Here are the best results for each Epoch test.

Table 3. Result of Epoch testing

Testing	Dataset	Window Size	Epoch	Mean Square Error
1st Test	80:20	5	10	0.0079
2st Test	80:20	5	50	0.0073
3st Test	80:20	5	100	0.0072
4st Test	80:20	5	150	0.0071
5st Test	80:20	5	200	0.0074

As seen in Table 3, the Epoch test scenario with five different Epoch value test processes got the best results at the Epoch value of 150 with a minimum Loss Mean Square Error value of 0.0071. Next, use the best Epoch value for the Learning Rate test. Learning Rate testing is carried out with the same test scenarios as testing Dataset, Window Size, and Epoch. The Learning Rate values tested are 0.0001, 0.0005, 0.00075, 0.001, and 0.0025. The following are the best results for each Learning Rate test.

Table 4. Result of Learning Rate testing

Testing	Dataset ,Window Size and Epoch	Learning Rate	Mean Square Error
1st Test	Dataset 80:20, Windows Size 5, and Epoch 150	0.0001	0.0078
2st Test		0.0005	0.0071
3st Test		0.00075	0.0073
4st Test		0.001	0.0073
5st Test		0.0025	0.0075

As seen in Table 4, the Learning Rate test scenario with five different Learning Rate value test processes got the best results at the Learning Rate value of 0.0005 with a minimum Loss Mean Square Error value of 0.0071.

From testing the dataset, Window Size, Epoch, and Learning Rate, 100 LSTM models were obtained. Of all these models, there are three best models with different dataset and parameter test scenario values and the same Loss Mean Square Error value, which is 0.0071. In the three best LSTM models, one of the models is taken with parameter a Dataset value of 80:20, Epoch 150, Window Size 5, and Learning Rate 0.0005. The following are the results of training on the best model.

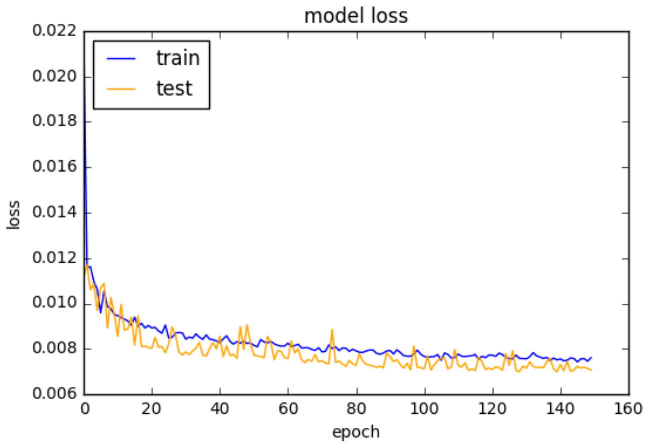

Figure 7. Training result of Loss value

As seen in the picture above, the graph of the Loss value on the test or validation data is below the Loss value graph on the train data, which shows that the best model does not experience Overfitting. Overfitting occurs when the Loss value on the data validation that bad and the Loss value on the data train is good. Figure 8 is the prediction result of the best model on electrical energy consumption data with 80% train data and 20% test data. Figure 8 shows that the pattern of the predictive data has followed the pattern of the test data so that the number of 522 data is sufficient to make a predictive model.

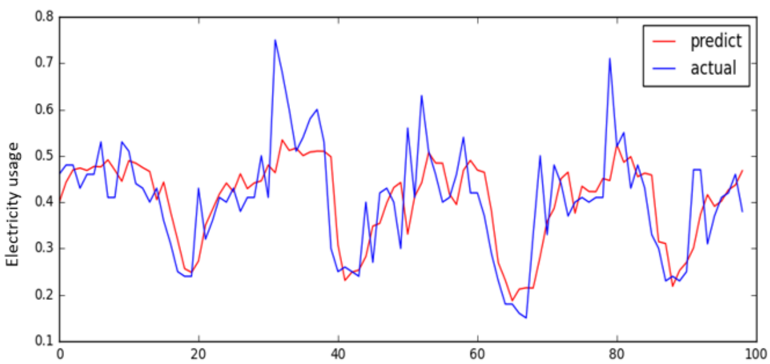

Figure 8. The testing results

3.2 Web Application Testing

Web application testing is done in a way, namely the alpha and beta test. Alpha testing completed by testing the function of the website without knowing the system is made (Black Box). There are eight negligent test scenarios, namely testing the account registration feature, dashboard login, electricity top up, forgetting passwords, logging out, notification of remaining electricity, and regenerating tokens. The 8-feature test process in the alpha test shows that all the features are running well with an accuracy value of 100%.

After completing the alpha testing, beta testing is carried out to determine the quality of the application from the user's side. Beta testing was carried out on July 7, 2022, through filling out a questionnaire on the Google Forms platform. In filling out the questionnaire, there were 35 respondents with an average answer of strongly agreeing (82.64%).

3.3 Deplyoment

After testing the LSTM model and web application, the best LSTM model is deployed. The deployment model is overall running at runtime on the local server. The following is the result of the LSTM model deployment process in predicting the use of electrical energy for the next day and predicting how long the remaining electricity will run out.

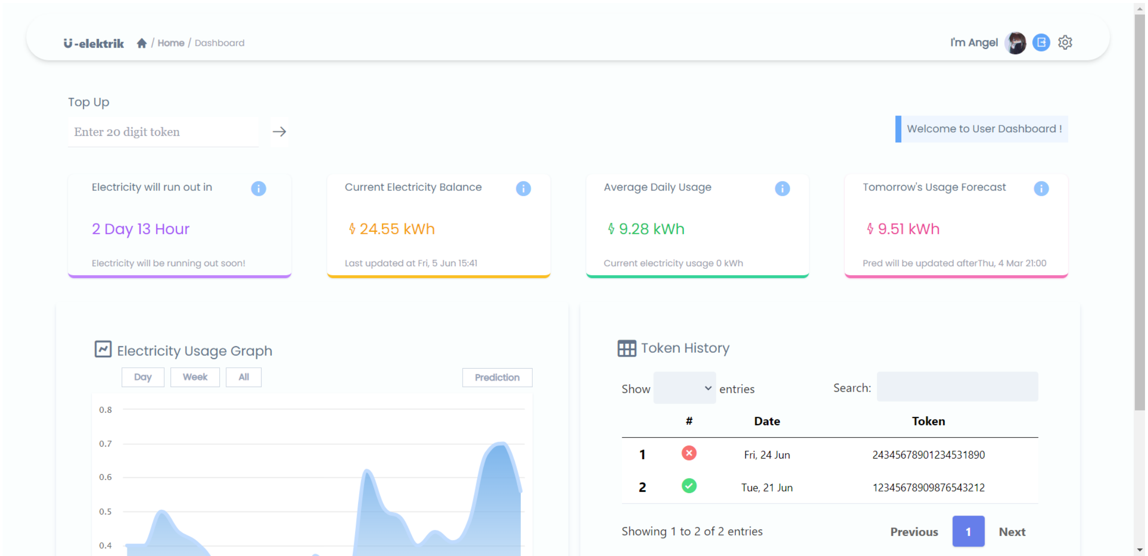

Figure 9. Result of prediction on dashboard page

As seen in Figure 8, the results of the electricity run out in 2 days and 13 hours are obtained from the division of the total remaining electricity now (24.55 kWh) with the predicted results added up for one day (9.51 kWh).

4. CONCLUSION

The test shows that the best LSTM model has an MSE value of 0.0071. Meanwhile, both the alpha test and beta test yield an average accuracy of 100% and 82.64% respectively. These results indicating that the LSTM model has a good performance in predicting the user's electric usage and all the website features are working properly and easy to use. Therefore, providing accurate feedback for the user so they can conveniently manage their electric usage. The prediction can be further improved by increasing the number of datasets. In addition, deploying the model in the cloud enable continuous operation.

RERERENCES

[1] R. Linda, L. Susanti, and H. R. Zadry, "Intervention Selection to the Awareness of Energy-Saving Behavior in the Public Sector," *MATEC Web Conf.*, vol. 248, pp. 0–4, 2018, doi: 10.1051/matecconf/201824803002.

[2] Badan Pusat Statistik (BPS), "Konsumsi Listrik per Kapita." https://www.bps.go.id/indicator/7/1156/1/konsumsi-listrik-per-kapita.html (accessed Oct. 04, 2021).

[3] M. A. Sarwar, N. Kamal, W. Hamid, and M. A. Shah, "Prediction of diabetes using machine learning algorithms in healthcare," *ICAC 2018 - 2018 24th IEEE Int. Conf. Autom. Comput. Improv. Product. through Autom.*

Comput., no. September, pp. 1–6, 2018, doi: 10.23919/IConAC.2018.8748992.
[4] A. Mosavi, S. Ardabili, and A. R. Várkonyi-Kóczy, “List of Deep Learning Models,” *Lect. Notes Networks Syst.*, vol. 101, no. January, pp. 202–214, 2020, doi: 10.1007/978-3-030-36841-8_20.
[5] J. Hyeon, H. Lee, B. Ko, and H. Choi, “Deep learning-based household electric energy consumption forecasting,” *J. Eng.*, vol. 2020, no. 13, pp. 639–642, 2020, doi: 10.1049/joe.2019.1219.
[6] M. M. Sachin, M. P. Baby, and A. S. Ponraj, “Analysis of energy consumption using RNN-LSTM and ARIMA model,” *J. Phys. Conf. Ser.*, vol. 1716, no. 1, 2021, doi: 10.1088/1742-6596/1716/1/012048.
[7] A. M. Pirbazari, M. Farmanbar, A. Chakravorty, and C. Rong, “Short-term load forecasting using smart meter data: A generalization analysis,” *Processes*, vol. 8, no. 4, 2020, doi: 10.3390/PR8040484.
[8] I. Mpawenimana, A. Pegatoquet, V. Roy, L. Rodriguez, and C. Belleudy, “A comparative study of LSTM and ARIMA for energy load prediction with enhanced data preprocessing,” *2020 IEEE Sensors Appl. Symp. SAS 2020 - Proc.*, pp. 1–6, 2020, doi: 10.1109/SAS48726.2020.9220021.
[9] A. P. Joshi and B. V. Patel, “Data Preprocessing: The Techniques for Preparing Clean and Quality Data for Data Analytics Process,” *Orient.J. Comp. Sci. Technol*, vol. 13, pp. 0–3, 2020, [Online]. Available: http://www.computerscijournal.org/vol13no2-3-and-4/data-preprocessing-the-techniques-for-preparing-clean-and-quality-data-for-data-analytics-process/
[10] X. Chen, Z. Ji, Y. Fan, and Y. Zhan, “Restful API Architecture Based on Laravel Framework,” *J. Phys. Conf. Ser.*, vol. 910, no. 1, 2017, doi: 10.1088/1742-6596/910/1/012016.
[11] A. Wathan, “Get started with Tailwind CSS.” https://tailwindcss.com/docs/installation (accessed Dec. 17, 2021).