Electricity Consumption Prediction in Oil and Gas Equipment Service and Maintenance Workshops Using RNN LSTM

Benedict Rafael¹, Asy’ari Muhammad Zacky¹*, and Kurniawan Irwan²

¹Automotive and Robotics Program, Computer Science Department, BINUS ASO School of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480
²IT Onsite Analyst Department, Schlumberger, Cikarang, Indonesia

Abstract. This research offers a Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) model for forecasting power usage in a facility that provides oil and gas equipment service and maintenance. The model was used using hourly electricity consumption data. The LSTM model was chosen because of its compatibility with time-series data and its capacity to capture temporal dependencies and patterns in sequential data, which may be utilized to predict future consumption. Experiments were undertaken in this study to determine the ideal model parameters and evaluate the model's accuracy using the root mean squared error (RMSE), mean absolute error (MAE), and mean absolute percentage error (MAPE) metrics. The findings demonstrated that the suggested model accurately predicted electricity usage with a MAPE of 3%. The quality and quantity of available data for the training dataset may, however, affect the accuracy of the model. Overall, our results indicate that the suggested RNN LSTM model can properly estimate factory power use.

1 Introduction

The worldwide energy consumption has increased significantly from 8,589.9 million tonnes (Mtoe) in 1995 to 13,147.3 Mtoe in 2015 [1]. The power consumption of Asian nations increased in 2016-2017, while the overall domestic energy consumption increased by 6.5% from 2000 to 2017 [2].

In addition, Asia’s total energy consumption is expected to increase dramatically from 6,318 Mtoe in 2020 to 8,770 Mtoe in 2040, where total energy consumption measures the energy demand requirements of end users, excluding losses and inputs related to energy sector changes [2]. The increase in energy demand is reason for concern because the majority of this energy is generated from nonrenewable fossil fuels [3].

In September 2015, the United Nations created Sustainable Development Goals (SDGs) to reduce air pollution and mitigate the health effects of hazardous substances. Ninety-three nations, both developed and developing, have signed the Sustainable Development Goals (SDGs) [4], so that more attention is paid to air pollution, which is the primary objective of the SDGs, especially SDG13 [5].

Indonesia is one of the non-OECD nations where the energy industry is crucial to economic development [6]. In 2021, the industrial sector account for 36% of Indonesia's total final electricity consumption [7]. Indonesia's power usage increased annually [8], with fossil fuels being the majority of the resources utilized to produce electrical energy in Indonesia [9].

Consequently, the industrial sector is responsible for the majority of the Indonesian government's carbon dioxide (CO₂) emissions. Indonesia's Industry sector provided 619.28 Mt of CO₂ emissions in 2021 alone. With the rising demand for energy and industrial CO₂ emissions [10].

Climate change and air pollution are largely caused by the combustion of fossil fuels (releasing carbon dioxide, black carbon, and ozone precursors) and agricultural production (emitting methane) [11]. The CO₂ emissions have a considerable impact on both of these phenomena. Green gas emissions impair the ecosystem by generating air pollution, depletion of the ozone layer, climate change, global warming, and depletion of fossil fuels [12].

Outdoor air pollution is a significant environmental and health risk factor on a global scale, and ambient (outdoor) and household (indoor) air pollution are responsible for approximately seven million fatalities annually. Through hypercapnia, elevated CO₂ concentrations can have numerous adverse health effects (defined as increased blood flow CO₂). According to regional WHO estimates, there are more than 2 million annual deaths related by air pollution in Southeast Asia, of which Indonesia is a member [13].

It emphasizes the threat to human health posed by the increase in CO₂ and green gas emissions. Numerous studies on energy consumption prediction have been conducted in recent years, and the signing of the SDGs 2030 by ninety-three nations has led to a rise in demand for energy efficiency in both the residential and industrial sectors to fulfill the SDGs 2030. Several research has been
undertaken in recent years to estimate building and industrial energy use using machine learning models. Kim et al. [14] employ a CNN-LSTM neural network to forecast residential energy consumption with an RMSE of 0.3085, MAE of 0.2382, and MAPE of 31.84 on a weekly timescale. He et al. predicted energy consumption with an RMSE of 0.6178, an MAE of 0.4544, and a MAPE of 721 using a unique double-layer bidirectional LSTM network with an enhanced attention mechanism [15]. The authors examined the performance of the LSTM, Multi-GRU Layer, and Drop-GRU in predicting energy usage. Sameh et al. [16] argue that the Drop-GRU models are superior to the LSTM and GRU models based on experiments undertaken. Wang et al. [17] performed more work to forecast long-term energy consumption using LSTM with periodicity, where LSTM has an RMSE of 0.3175 for the training set in April and an average RMSE of 1.5448 for the test set from March to July. Ashutosh et al. examine the performance of the SARIMA and LSTM models for forecasting time series data, where the SARIMA model has an RMSE of 0.5506 and a MAPE of 5.2369. While the LSTM has an RMSE of 0.2319 and a MAPE of 3.222 [18].

After analyzing the existing literature on machine learning for energy consumption prediction, we propose employing Recurrent Neural Networks (RNNs) with Long Short-Term Memory (LSTM). Because the RNN LSTM model has various benefits, including the ability to handle long-term dependencies in data, which is crucial for predicting energy consumption patterns that exhibit cyclical activity.

In addition, LSTM-based RNN models may also learn from historical and current data, allowing the model to make accurate predictions even when the data is complex or contains missing values. This research aims to identify the most effective and precise parameter for LSTM-based time series data prediction on the particular implementation of energy consumption in maintenance workshops in industrial areas.

2 Methodology

The proposed methodology uses data on energy consumption per hour from the service and maintenance workshop as a dataset. The workshop building areas are equipped with a smart power meter Siemens 7KT0310 (Fig. 1) to measure electricity consumption in that area. The device has industrial grade with specifications as seen in Table 1.

<table>
<thead>
<tr>
<th>Measurement</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Voltage $V_{L-N}$</td>
<td>+/- 0.5 % Full Scale (300V AC)</td>
</tr>
<tr>
<td>Voltage $V_{L-L}$</td>
<td>+/- 0.5 % Full Scale</td>
</tr>
<tr>
<td>Current</td>
<td>+/- 0.5 % Full Scale (6A)</td>
</tr>
<tr>
<td>Frequency</td>
<td>+/- 0.1 %</td>
</tr>
<tr>
<td>Power factor</td>
<td>+/- 0.01 % Digit</td>
</tr>
</tbody>
</table>

Table 1. Siemens 7KT0310 Technical Specification

The proposed methodology flowchart covers the following steps as illustrated in Fig. 2.

Fig. 2. Flowchart of the proposed methodology

The daily energy consumption per hour values are presented in Table 2 for the first six hours. Due to the confidentiality of the data, the total consumption has been censored. In this dataset, there was minimal data processing due to the technician writing the data hourly based on the display, primarily on miss typos.

<table>
<thead>
<tr>
<th>Date</th>
<th>Time</th>
<th>Hour</th>
<th>Date</th>
<th>Time</th>
<th>Energy Consumption per hour (kWh)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sat Jan 01 2022</td>
<td>00:00:00</td>
<td>00</td>
<td>Jan 01</td>
<td>00:00:00</td>
<td></td>
</tr>
<tr>
<td>Sat Jan 01 2022</td>
<td>01:00:00</td>
<td>01</td>
<td>Jan 01</td>
<td>01:00:00</td>
<td></td>
</tr>
<tr>
<td>Sat Jan 01 2022</td>
<td>02:00:00</td>
<td>02</td>
<td>Jan 01</td>
<td>02:00:00</td>
<td></td>
</tr>
<tr>
<td>Sat Jan 01 2022</td>
<td>03:00:00</td>
<td>03</td>
<td>Jan 01</td>
<td>03:00:00</td>
<td></td>
</tr>
<tr>
<td>Sat Jan 01 2022</td>
<td>04:00:00</td>
<td>04</td>
<td>Jan 01</td>
<td>04:00:00</td>
<td></td>
</tr>
<tr>
<td>Sat Jan 01 2022</td>
<td>05:00:00</td>
<td>05</td>
<td>Jan 01</td>
<td>05:00:00</td>
<td></td>
</tr>
<tr>
<td>Sat Jan 01 2022</td>
<td>06:00:00</td>
<td>06</td>
<td>Jan 01</td>
<td>06:00:00</td>
<td></td>
</tr>
</tbody>
</table>

Step one involves preparing the Secondary data as the dataset, which is a text-based CSV format that will be used for training the model. The dataset consists of
electricity consumption over 9 months from 1 January 2022 until 26 September 2022, with each day containing 13 data samples (00:00 - 06:00 and 18:00 - 23:00) that are captured every hour, for a total sample size of 3,380.

The electricity consumption dataset from 1 January 2022 until 26 September 2022 will go through the data cleaning process. During data cleaning, several instances of incomplete data were identified. For example, on 24 March 2022 where the recorded data only extended until 19:00, whereas the complete day's data should include records until 23:00. Furthermore, from 25 March 2022 until 31 March 2022 there was a span of missing data. Another instance occurred on 12 May 2022, where the recorded data was only extended until 06:00. To ensure data integrity and eliminate abnormal data and outliers, the incomplete data were removed from the dataset during the cleaning process.

Step 2: Selecting features. The first eight months of data (January to August) serve as a training set, while September serves as a test set. For the input data to be processed by the LSTM layer

Fig. 3 is preprocessed and molded into a 3D tensor with shape (n samples, n timesteps, n features).

Fig. 3. LSTM model architecture

The third step is predictive modeling. This paper proposes the Keras Sequential Model, which is composed of 5 layers: 4 LSTM layers and 1 Dense layer.

As input data for the first LSTM layer, a 3D tensor shape was used. This layer employs 13 time steps, with 78 features per time step, and 24,960 trainable parameters.

To prevent overfitting, a dropout layer with a value of 0.1 was utilized in conjunction with a regularization approach. In the second and third LSTM layers, the prior dropout's output serves as an input. There are 48,984 parameters in this layer, and a dropout layer with a dropout value of 0.1 is utilized [19].

In the final LSTM layer, the preceding dropout output is used as an input. In this layer, the output will be a 78-dimensional, two-dimensional form tensor. This layer consists of 48,984 parameters, and a dropout layer with a dropout value of 0.1 is utilized.

In the Dense layer, the output from the dropout layer is mapped to a single output prediction. This layer contains 79 trainable parameters. This model is trained using a dataset of power usage. Based on our experiment, the parameter will be changed to build the most accurate model feasible.

Step 4: Model Evaluation. The RMSE (Root Mean Square Error), MAE (Mean Absolute Error), and MAPE (Mean Absolute Percent Error) are used to evaluate the model's usefulness and precision in predicting electricity usage. RMSE was employed to measure the average magnitude of the mistake in the forecasts, taking both the size and direction of the errors into account. Using the MAE to calculate the average absolute difference between actual and anticipated values, which is less sensitive to outliers and high mistakes than the RMSE.

MAPE was utilized to quantify the relative magnitude of the mistake in the forecasts as a percentage of the actual value. Where \( \hat{y}_i \) indicates the predicted values for the time \( i \) of a regression's dependent variable \( y_i \), and \( n \) represents the observed variables over \( n \) times. By exploiting the advantages of each statistic, the author can more effectively identify the model's flaws and utilize them as a guide to develop a new model that will produce more accurate predictions.

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2} 
\]

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i| 
\]

\[
MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{y}_i - y_i}{y_i} \right| 
\]

Step 5: Model Updating. In this step, multiple experiments are conducted on the hyperparameter of the LSTM RNN model to achieve the most precise prediction results.

Experiment first with the time stamp parameter, which is intended to provide additional context for the model to learn from. Second, on the epochs parameter, ensure that the model has sufficient training time to learn patterns to avoid underfitting. Adjusting the dropout values to prevent overfitting is the third step. Fourth, the effect on the number of units in the LSTM layer, which allows the model to learn more complicated data linkages and produce more accurate predictions by capturing long-term dependencies. Fifth, the number of batches necessary to optimize the speed of model training while maintaining the model's capacity for successful learning.

The other parameters to be evaluated are the changes to the model's optimizer based on the dataset and tasks provided. If the model result is not correct when compared to the actual number, it is possible to proceed to step 3 and experiment with other parameters. Over nine months, the factory's electrical energy consumption data was analyzed.

Several trends and patterns in the dataset are identified and will be used to train the model. Based on Fig. 4, it can be seen that Saturday and Sunday have a lower average daily electricity usage than weekdays (Monday through Friday). There is no substantial difference in the average daily electricity use from Monday to Friday.

According to Fig. 5. This visualization is crucial for training the RNN LSTM model since the author will be
able to do data filtering and cleaning, making the data used to train the model more accurate, relevant, and indicative of the problem being solved.

![Average Electricity Consumption by Day of the Week](image1)

**Fig. 4.** Average electricity consumption profile during the week

![Electricity Consumption](image2)

**Fig. 5.** Electricity consumption from January to September

### 3 Result and Discussion

In this study’s simulations, we explore a variety of factors, including Optimizer, Timestamp, Units, Drop out, and Epochs. During trials, the three different optimizers, Rmsprop, Adagrad, and Adam, and other parameters, such as Timestamp with 13 Timestamp, Units with 78 Units, Dropout with 0.1 Dropout, Epochs with 150 Epochs, and Batch size with 16 batch size, are maintained constant. Rmsprop optimizer was selected due to its adaptive learning rate optimization technique, which can alter the learning rate of each parameter based on the most recent gradients; this helps to solve the problem of diminishing learning rate. Adagrad is a second learning rate optimization strategy that may alter the learning rate of each parameter depending on the sum of the squares of gradients. This approach performs well in sparse datasets and can converge more quickly than previous optimization algorithms. Adam, the third optimizer, is a blend of adaptive learning rate optimization algorithms and momentum-based optimization techniques; this achieves a balance between adaptive learning rate methods and momentum approaches. These three Optimizers each have their advantages and weaknesses; the purpose of this study is to determine which one is superior for predicting electricity consumption. Fig. 6 depicts optimizer Adam as the most accurate algorithm when compared to Rmsprop and Adagrad. The Adam optimizer had RMSE values of 13.5499, 7.8026, and 0.0313. This optimizer parameter experiment demonstrates that Adam is the optimal optimizer for estimating electricity use.
The subsequent experiment demonstrates the number of timesteps required for the models to more precisely recognize and capture patterns using Fig. 7. The author conducts experiments on timesteps ranging from one day (13 timesteps) to one week (91 timesteps) while holding all other parameters constant, including Units with 78 Units, Dropout with 0.1 Dropout, Epochs with 150 Epochs, Batch size with 16 batch size, and Optimizer Adam. Using 91 timesteps rather than 13 timesteps to predict future electricity consumption yields more accurate results. Model 5 has lower RMSE, MAE, and MAPE values than model 4, which has RMSE values of 11.1578, MAE values of 7.6233, and MAPE values of 0.0304. In the objective of predicting electricity consumption, a parameter with 91 Timesteps can demonstrate that the model can learn to identify and capture patterns better than 13 timesteps.

In Fig. 8, the author conducts experiments on the unit's parameter to demonstrate the number of units required for the model to learn complex patterns in the data without overfitting. The author maintains other parameters, including Timestamp with 13 Timestamp, Dropout with 0.1 Dropout, Epochs with 150 Epochs, Batch size with 16 batch size, and Optimizer Adam. Model 8 with 78 units performs better than models 6 and 7, which have fewer units than model 8. Model 8’s RMSE is 13.5499, its MAE is 7.8026, and its MAPE is 0.0313. This demonstrates how Model 8 with 78 units can learn complex data patterns without risking overfitting.

Fig. 9 analyzed the experiments with the dropout parameter to demonstrate how large a dropout value must be for the model to be more resistant to overfitting while avoiding underfitting, while holding the other parameters constant, including Units with 78 Units, Dropout with 0.1 Dropout, Epochs with 150 Epochs, Batch size with 16 batch size, and Optimizer Adam.
parameters constant, including Units with 78 Units, Timestamp with 13 Timestamp, Epochs with 150 Epochs, Batch size with 16 batch size, and Optimizer Adam. In this experiment, the value of models 9, 10, and 11 increased the dropout layer by 0.1. In addition, the outcome of the experiment reveals that Model 9 with a 0.1 dropout made the model resistant to underfitting without producing underfitting. With an RMSE value of 13,5499, an MAE value of 7.8026, and a MAPE value of 0.0313, it can be shown that model 9 had a better outcome in evaluation metrics.

The remaining experiments examine the epochs parameter to determine if more epochs are required for more accurate predictions, given that the larger the epochs, the longer it will take to train the model. Other parameters, such as Timestamp with 13 Timestamp, Units with 78 Units, Dropout with 0.1 Dropout, Batch size with 16 batch size, and Optimizer Adam, remain unchanged. The addition of epochs improves prediction accuracy without producing overfitting, as shown in Fig. 10.

For example, model 12 that have 50 epochs an RMSE value of 18.0328, an MAE value of 10.3774, and a MAPE value of 0.0392. As the number of epochs climbed to 150, the metrics values became more precise, with RMSE values of 13,5499, MAE values of 7,8026, and MAPE values of 0.0313.

Fig. 9. Model results with different Drop-out

Fig. 10. Model results with different Epoch

Fig. 11 is a visual representation of a comparison between the Electricity Consumption per hour data predicted by model 5 and the actual Electricity Consumption per hour data in September, where the predicted data is represented with the color blue and the actual data is represented with the color red.

Fig. 11. Predicted electricity consumption vs real data
Based on the visualization, there are several days where the predicted Electricity Consumption per hour is still inaccurate in comparison to the actual data for Electricity. This disparity in consumption per hour is due to the RMSE value of 11.1578, the MAE value of 7.6233, and the MAPE value of 0.0304.

When compared to other papers, this result is reasonably accurate. Kim et al experiments have RMSE values of 0.3085 and MAE values of 31.84. Although it had a lower RMSE value when compared to our RMSE value of 11.1578 and MAPE value of 0.0304, it performed better when compared to the MAPE value of 0.0304, as the RMSE was higher because the author used factory electricity and the author used household electricity consumption [14], which was significantly less than the factories.

4 Conclusion

This research proposes an RNN-LSTM model to estimate manufacturing electricity use based on a nine-month dataset. The suggested model has a total of 5 layers: 4 LSTM layers and 1 Dense layer. The performance of the model was tested using RMSE, MAE, and MAPE to reach high accuracy. The authors picked model 5 utilizing parameters, timestamp 91, units 78, dropout 0.1, 150 epochs, batch size 16, and optimizer Adam as the best accurate model through a series of experiments. In addition, the paper discusses the effect of each parameter on the accuracy of the model.

The suggested RNN-LSTM model demonstrated greater accuracy in terms of MAPE metrics, although having larger RMSE values, in contrast to previous models that use CNN-LSTM to estimate home energy consumption. It is important to note that a comparison was made between the electricity use of a manufacturing industry and that of a household.

This indicates that the suggested RNN-LSTM model is more accurate than the CNN-LSTM model for predicting electricity. The results indicate that the proposed model is good enough for estimating a facility that provides oil and gas equipment service and maintenance on electricity used.

References

1. K. Dong, X. Dong, and Q. Jiang, World Econ 43, 1665 (2020)
5. P. Carter, Climate Action 413 (2020)
7. Enerdata.net, (2023)
10. Pierre Friedlingstein et al., (2023)
17. J. Q. Wang, Y. Du, and J. Wang, Energy 197, 117197 (2020)